

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/30-1-1610>

Lexikos 30

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/30-1-1610>

Lexikos 30

Redakteur

Editor

Hugues Steve Ndinga-Koumba-Binza

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 30:2020

AFRILEX SERIES 30:2020



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/30-1-1610>

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2020 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2020

Tipografie en uitleg deur Tanja Harteveld en Hermien van der Westhuizen
Bandontwerp deur Piet Grobler
Geset in 10 op 12 pt Palatino

ISBN 978-0-9946528-8-1
ISSN 2224-0039

Hierdie werk is gelisensieer ingevolge 'n Creative Commons License CC BY 4.0-lisensie.

Licensed under Creative Commons License CC BY 4.0.

Menings wat in artikels en resensies uitgespreek word, is nie noodwendig dié van AFRILEX of die Buro van die WAT nie.

Opinions expressed in the articles and reviews are not necessarily those of AFRILEX or of the Bureau of the WAT.

Lexikos is elektronies beskikbaar by <http://lexikos.journals.ac.za/>

Lexikos is available online at <http://lexikos.journals.ac.za/>

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest

Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes

Asian Digital Library, Arts and Humanities Citation Index®, Current Contents®/Arts & Humanities, Current Contents®/Social and Behavioral Sciences, ERIH Plus, Index Copernicus Journals Master List, Journal Citation Reports/Social Sciences Edition, Social Sciences Citation Index®, and Social Scisearch®; Linguistic Bibliography Online; Linguistics Abstracts Online; Linguistics and Language Behavior Abstracts; MLA Inter-national Bibliography; R.R.K. Hartmann's Bibliography of Lexicography; SciELO SA; Scopus

Span van Roterende Redakteurs / Team of Rotating Editors

Dr. H.S. Ndinga-Koumba-Binza (RSA en Gaboen/RSA and Gabon)
Prof. Dion Nkomo (RSA en Zimbabwe/RSA and Zimbabwe)
Prof. Elsabé Taljard (RSA)

Adviesraad / Advisory Board

Prof. A. Adamska-Salaciak (Pole/Poland)
Prof. H. Béjoint (Frankryk/France)
Prof. H. Chimhundu (Zimbabwe)
Prof. F. Dolezal (VSA/USA)
Prof. R.H. Gouws (RSA)
Prof. R.R.K. Hartmann (Groot-Brittanje/Great Britain)
Prof. W. Martin (België en Nederland/Belgium and The Netherlands)
Prof. I.A. Mel'čuk (Kanada/Canada)
Prof. A.M.F.J. Moerdijk (Nederland/The Netherlands)
Dr. J. Tent (Australië/Australia)
Prof. J. Van Keymeulen (België/Belgium)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)

Redaksiekomitee / Editorial Committee

Dr. M.M. Bagwasi (Botswana)
Prof. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
Prof. E. Chabata (Zimbabwe)
Prof. C.J. Conradie (RSA)
Prof. A.E. Feinauer (RSA)
Prof. R. Finlayson (RSA)
Dr. S. Hadebe (Zimbabwe)
Prof. I.M. Kosch (RSA)
Dr. P.A. Louw (RSA)
Mnr. K.J. Mashamaite (RSA)
Prof. P.A. Mavoungou (Gaboen/Gabon)
Dr. V.M. Mojela (RSA)
Mnr. M.C. Mphahlele (RSA)
Prof. T.J. Otlogetswe (Botswana)
Prof. A.N. Otto (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	x
Foreword	xii
<i>Hugues Steve Ndinga-Koumba-Binza</i>	
'n Woord van AFRILEX	xiv
A Few Words from AFRILEX	xv
<i>Herman L. Beyer</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii

Artikels / Articles

Exploring the Documentation and Preservation of African Indigenous Knowledge in a Digital Lexical Database	1
<i>Sonja E. Bosch and Marissa Griesel</i>	
e-Dictionaries in a Network of Information Tools in the e-Environment	29
<i>Theo J.D. Bothma and Rufus H. Gouws</i>	
Selecting an Initial Lemma List in Specialized Lexicography: A Case Study in the Field of Graphic Engineering	57
<i>M.A. Esandi-Baztan and Pedro A. Fuertes-Olivera</i>	
Integrating Terminological Resources in Dictionary Portals: The Case of the <i>Diccionarios Valladolid-UVa</i>	90
<i>Pedro A. Fuertes-Olivera and M.A. Esandi-Baztan</i>	
Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners	111
<i>Jun Gao and Haitao Liu</i>	
Special Field and Subject Field Lexicography Contributing to Lexicography	143
<i>Rufus H. Gouws</i>	

An Overview of English Dictionaries of Abbreviations <i>Mojca Kompara Lukančič</i>	171
Der LeGeDe-Prototyp: Zur Erstellung eines korpusbasierten Online-Wörterbuchs zu lexikalischen Besonderheiten im gesprochenen Deutsch. Aktueller Stand und Perspektiven <i>Meike Meliss und Christine Möhrs</i>	197
Users and Contexts of Use of Romanian Multilingual Dictionaries <i>Mihaela Mocanu and Alina-Mihaela Bursuc</i>	238
<i>Sethantšo sa Sesotho and Sesuto–English Dictionary: A Comparative Analysis of their Designs and Entries</i> <i>T.L. Motjope-Mokhali, I.M. Kosch and Munzhedzi James Mafela</i>	258
Performing Non-sexism via Degendering Phoric Forms in English: The Gap between Rules and Practice as Observed in the 9th Edition of <i>Oxford Advanced Learner’s Dictionary of Current English</i> <i>Nshindi-Germain Mulamba and Francis Ngoyi Crequi Tshimanga</i>	275
Teaching Lexicography as a University Course: Theoretical, Practical and Critical Considerations <i>Tortko Prčić</i>	293
Lexicographic Treatment of Negation in Sepedi Paper Dictionaries <i>D.J. Prinsloo</i>	321
Lexicographic Treatment of Zero Equivalence in isiZulu Dictionaries <i>D.J. Prinsloo and Nompumelelo Zondi</i>	346
Which Defining Model Contributes to More Successful Extraction of Syntactic Class Information and Translation Accuracy? <i>Bartosz Ptasznik</i>	363
Towards Accuracy: A Model for the Analysis of Typographical Errors in Specialised Bilingual Dictionaries. Two Case Studies <i>Santiago Rodríguez-Rubio and Nuria Fernández-Quesada</i>	386
Lexicography and Language Planning in 18th Century Sweden <i>Lena Rogström</i>	416

A Critical Evaluation of Three Sesotho Dictionaries <i>Mmasibidi Setaka and D.J. Prinsloo</i>	445
Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture <i>Sven Tarp and Rufus Gouwes</i>	470
'n Model vir 'n aanlyn GIS-vakwoordeboek <i>Lize-Mari Theron</i>	499
On Pronunciation in a Multilingual Dictionary: The Case of Lukumi, Olukumi and Yoruba Dictionary <i>Joy O. Uguru and Chukuma Okeke</i>	519
Création de ressources lexicographiques Français–Slovène d'aide à la traduction spécialisée <i>Sonia Vaupot</i>	540
Normative Labels in Two Lexicographic Traditions: A Slovene– English Case Study <i>Marjeta Vrbinc, Danko Šipka and Alenka Vrbinc</i>	561
Access Routes to BODY PART Multiword Expressions in the 'Big Five' MELDs: Use of Hyperlinks <i>Sylwia Wojciechowska</i>	583

Leksiko-opname / Lexicosurvey

A Survey of Indonesian Students' Use of Dictionaries <i>Fallianda</i>	609
--	-----

Resensies / Reviews

John Considine (Editor). <i>The Cambridge World History of Lexicography</i> <i>Xiangming Zhang and Hai Xu</i>	629
María José Domínguez Vázquez, Mónica Mirazo Balsa and Carlos Valcárcel Riveiro (Eds.). <i>Studies on Multilingual Lexicography</i> <i>Hebing Xu and Wenfei Hu and He Zhao</i>	635

Michael Klotz and Thomas Herbst. <i>English Dictionaries: A Linguistic Introduction</i> Kaihuai Du	642
Marie-Claude L'Homme. <i>Lexical Semantics for Terminology: An Introduction</i> Lingyun Lü and Renqiang Wang	646
Van Wyk, E. (Compiler). <i>The Talking Dictionary</i> Maria Hermione le Roux	654
Publikasieaankondigings / Publication Announcements	664
Voorskrifte aan Skrywers	665
Instructions to Authors	666

Voorwoord

Vanjaar sal sekerlik in die geskiedenis van die mensdom onthou word weens die uitbreek van die Koronaviruspandemie van 2019 aan die einde van verlede jaar. Die pandemie het wêreldwyd alle sferes van die lewe beïnvloed, en *Lexikos* het dit ook nie vrygespring nie. Ons het baie oponthoude in die lewering van keuringsverslae ervaar aangesien baie van ons keurders oor die wêreld heen gepoog het om aan te pas by die werksomstandighede wat op verskillende maniere hanteer is in die onderskeie instansies. Met behulp van baie kollegas wat ek kon kontak, is die keuringsproses van alle artikels uiteindelik suksesvol afgehandel. Om soortgelyke druk in die toekoms te vermy, het ons 'n lys van 91 potensieële keurders vir die vyf tale waarin *Lexikos* publiseer (Afrikaans, Nederlands, Engels, Frans en Duits) saamgestel.

Vanjaar se uitgawe is 'n mylpaal aangesien dit ons 30ste volume is. In die meeste kulture word 'n persoon wat 30 jaar oud word nie meer as 'n jong volwassene beskou nie. *Lexikos* het sodoende ook 'n bewese vlak van volwassenheid bereik wat reeds deur die jare verwerf is deur die gehalte van die onderskeie bydraes. Soos die Franse dramaturg Pierre Corneille dit gestel het: "Aux âmes bien nées, la valeur n'attend point le nombre des années." ("Vir dié wat groots gebore word, volg waagmoed nie eers ná die verloop van jare nie.") *Lexikos* is sedert sy ontstaan 'n gesogte publikasie, enersyds as hulpmiddel vir die verspreiding van kennis en andersyds as 'n bron van kennis in die velde van die leksikografie, leksikologie, terminologie en korpuslinguistiek.

Een van die slagspreuke van die Suid-Afrikaanse Lugdiens maak die aanspraak dat hulle Afrika na die wêreld neem en die wêreld na Afrika. Dit is presies wat *Lexikos* die afgelope drie dekades vermag het met die wêreld se beste kundiges in bogenoemde velde wat gereeld en konsekwent die beste navorsingsuitsette in Afrika in hierdie velde gepubliseer het. Trouens, *Lexikos* is een van baie min joernale wat bydraes van die uithoeke van die wêreld ontvang. Met artikels en resensies wat ontvang is vanuit Afrika, Asië en Europa, in totaal vyftien lande, is hierdie uitgawe geen uitsondering nie. Hierdie diversiteit het *Lexikos* as een van die wêrelderkende joernale wat in Afrika gepubliseer word, gevestig.

Lexikos het van veelsydigheid 'n tradisie gemaak. Dit is nie net verkry deur die verskillende geografiese oorspronge van die bydraes nie, maar hoofsaaklik deur die verskillende teoretiese en praktiese leksikografiese kwessies van 'n groot verskeidenheid tale. Die vernuwing vanjaar is egter dat *Lexikos* in 'n deurlopende aanlyn publikasie verander het.

Ek bedank graag ons outeurs en keurders vir hul voortdurende bydraes wat die kwaliteit en standaard van ons joernaal handhaaf. Weereens het Me. Tanja Harteveld en Me. Hermien van der Westhuizen van die Buro van die Woordeboek van die Afrikaanse Taal (WAT) hulp verleen met die uitleg van

die artikels en die kontroliering van finale taalkundige en strukturele aspekte. Ek is besonder dankbaar vir hul uitstekende werk.

Ons verwelkom prof. Dion Nkomo ook hierdie jaar as lid van die Span van Roterende Redakteurs. Hy vervang prof. Danie Prinsloo wat afgetree het as redakteur. Ons is dankbaar teenoor prof. Prinsloo vir sy aandeel in die redigering van drie *Lexikos*-volumes. Prof. Nkomo beskik oor sowel waardevolle kennis van Afrikataalleksikografie as gesaghebbende ervaring in die redigering van navorsingsjoernale. Met sy aanstelling het die AFRILEX-raad 'n baie goeie keuse uitgeoefen. Wamkelekile Dion!

Ek moet hierdie voorwoord ongelukkig op 'n hartseer noot eindig. Ons het vroeër vanjaar met leedwese verneem van die heengaan van mnr. Godfrey Baile Mareme, die hoofredakteur van die Setswana Nasionale Leksikografie-eenheid (NLE). Sy werk het 'n merkwaardige impak op die ontwikkeling van die Setswana-leksikografie gehad. Tydens die bekendmaking van sy afsterwe het die tesourier van AFRILEX aan hom erkenning gegee as "een van die lojaalste AFRILEX-lede". As huldeblyk het een van sy voormalige dosente in 'n Motswedding FM-radioprogram gemeld dat mnr. Mareme een van die eerste studente was wat ingeskryf het toe die Universiteit van Bophuthatswana (tans bekend as die Mafikengkampus van die Noordwes-Universiteit) tot stand gekom het. Nadat hy sy Baccalaureus Artium in Opvoedkunde en sy Honneursgraad voltooi het, het hy in die leksikografie en vertaling begin belangstel nadat hy, op aanbeveling van sy dosent, in 1989 'n werkswinkel in vertaling in Port Elizabeth bygewoon het. Vervolgens het hy gevorder tot een van die fakulteitsleiers in Setswana-vertaling voordat hy as Hoofredakteur van die Setswana NLE aangestel is. Die meeste kollegas met wie hy saamgewerk het, sal hom onthou vir sy diepgaande kennis van die Mahikeng-geskiedenis. By sy gedenkdienis wat by die Noordwes-Universiteit gehou is, het ons in 'n huldigingswoord van die Premier van die Noordwes Provinsie te wete gekom dat, buiten vir sy merkwaardige akademiese vaardighede, mnr. G.B. Mareme, soos hy aan die meeste van ons by AFRILEX bekend was, ook 'n legendariese motorfietsryer, 'n veteraanlid van die Ride-With MCC in Mafikeng was waar hy bekend gestaan het as 'Bra Lebs' Mareme. By toekomstige AFRILEX-konferensies sal ons nie net die passievolle en vasberade kampvegter vir Setswana mis nie, maar ook 'n blymoedige en onderhoudende kollega. Dit berus nou by die Setwana-leksikografe om sy nalatenskap te bevorder deur by te dra tot die uitbreiding van kennis in ons navorsingsveld.

Kontak ons vir kommentaar en voorstelle deur 'n e-pos te stuur aan lexikos@sun.ac.za.

Hugues Steve Ndinga-Koumba-Binza
Redakteur

Foreword

This year will certainly be remembered in the history of humankind because of the COVID-19 pandemic outbreak at the end of last year. The pandemic affected all spheres of life globally, and *Lexikos* has not been spared. We experienced many delays in the submission of review reports, as many of our adjudicators around the world were trying to adapt to the working conditions that were differently implemented in various institutions. With the help of many colleagues I could contact, the adjudication process of all articles was finally achieved successfully. To avoid pressure of this kind in future, we compiled a list of 91 potential adjudicators for the five languages in which *Lexikos* publishes (Afrikaans, Dutch, English, French and German).

This year's edition is a landmark as it is our 30th volume. In most cultures, a person is no longer a young adult when turning 30. Thus, *Lexikos* has reached an affirmed level of maturity, which was already attained through the quality and the calibre of various contributions all these years. As the French playwright Pierre Corneille put it: "Aux âmes bien nées, la valeur n'attend point le nombre des années." ("For souls nobly born valour does not await the passing of years.") *Lexikos* has been a valuable publication since its inception in 1991, being both a tool for knowledge dissemination and a source for knowledge in the fields of lexicography, lexicology, terminology and corpus linguistics.

One of the catch-phrases of the South African Airways claims to take Africa to the world and the world to Africa. This is exactly what *Lexikos* has been doing for the past three decades, having the world's best experts in the above-mentioned fields contributing on a regular basis and consistently publishing the best of African research outputs in these same fields. In fact, *Lexikos* is one of the very few journals that receive contributions from various corners of the world. The current issue is no exception, with articles and reviews received from Africa, Asia and Europe, a total of fifteen countries. This diversity has been keeping *Lexikos* as one of the worldwide recognized journals published in Africa.

Lexikos has made diversity its tradition. This is not only through the various geographical origins of the contributions but mainly through the different theoretical and practical lexicographic issues from a great variety of languages. The innovation of this year, however, is that *Lexikos* started to do continuous publication online.

I wish to thank our authors and adjudicators for their unceasing contribution to maintain the quality and standard of our journal. Once again, Ms Tanja Harteveld and Ms Hermien van der Westhuizen at the Bureau of the Woordeboek van die Afrikaanse Taal (WAT) assisted with typesetting the articles,

making final language and structural checks. I am particularly thankful for their excellent work.

This year, we also welcome Prof. Dion Nkomo as a member of the Team of Rotating Editors. He is replacing Prof. Danie Prinsloo who has retired as editor. We are grateful to Prof. Prinsloo for his contribution in editing three volumes of *Lexikos*. Prof. Nkomo brings a valuable knowledge of African language lexicography as well as a reliable experience in editing research journals. The Board of AFRILEX made a very good choice in appointing him. Wamkelekile Dion!

I unfortunately have to end this foreword with a sad note. Earlier this year, we learnt with great sadness of the passing of Mr Godfrey Baile Mareme, the Editor-in-Chief of the Setswana National Lexicography Unit (NLU). His work has had a remarkable impact on the development of Setswana lexicography. In the announcement of his passing, the treasurer of AFRILEX recognized him as "one of the most loyal AFRILEX members". As a tribute, one of his former lecturers mentioned in a Motsweding FM radio programme that Mr Mareme was one of the first students to enrol when the then University of Bophuthatswana was established (now known as the Mafikeng Campus of the North West University). After completing his Bachelor of Arts in Education and his Honours degree, he became interested in lexicography and translation after attending, on recommendation by his lecturer, a workshop on translation in Port Elizabeth in 1989. Subsequently, he became one of the faculty leaders in Setswana translation before being appointed Editor-in-Chief of the Setswana NLU. Most colleagues he worked with would remember his extensive knowledge of Mahikeng history. In a eulogy by the Premier of the North West Province at his memorial service held at the North-West University, we learnt that in addition to having remarkable academic skills, Mr G.B. Mareme, as he was known to most of us at AFRILEX, was also a legend biker, a veteran member of the Ride-With MCC in Mafikeng where he was known as Lebolobolo 'Bra Lebs' Mareme. At future AFRILEX conferences, we will miss not only a passionate and uncompromising defender of Setswana, but also a cheerful and amusing colleague. It now remains the duty of Setswana lexicographers to advance his legacy by contributing in broadening the knowledge in our field of research.

For comments and suggestions, please contact us by sending an email to lexikos@sun.ac.za.

Hugues Steve Ndinga-Koumba-Binza
Editor

'n Woord van AFRILEX

Wat 'n jaar was dit nie. 2020 sal ongetwyfeld deurgaans as 'n jaar van diepsnydende globale, maar veral persoonlike ontwrigting vir sekerlik ons almal, en 'n jaar van ondenkbare verlies op soveel terreine vir vele.

Ook die African Association for Lexicography (AFRILEX) het nie ontwrigting vrygespring nie. Vir die eerste maal sedert die Vereniging se bestaan kon sy jaarlikse internasionale kongres nie plaasvind nie, as gevolg van die Covid-19-pandemie. Ons kon nie, soos oorspronklik beplan, ons kwarteeufees op Stellenbosch vier nie. Ons sal dit ook nie volgende jaar persoonlik kan doen nie, want weens die heersende onsekerheid moes die AFRILEX-Raad noodgedwonge uiteindelik besluit om die internasionale jaarlikse kongres van 2021 as 'n ten volle virtuele, aanlyn geleentheid te beplan. Ons sal dus vir 'n ruk nog ons jaarlikse fisiese samesyn as AFRILEX-familie moet ontbeer.

Met die kansellasie van menige vakverenigings se jaarlikse byeenkomste vanjaar het vaktydskrifte byna die enigste kanaal vir die disseminasie van portuurbeoordeelde navorsing geword. *Lexikos* is geen uitsondering nie, soos die groot aantal bydraes deur plaaslike en internasionale outeurs tot vanjaar se nommer getuig. Te midde van die onmiskenbare impak van Covid-19 het die redaksie van *Lexikos* onverpoosd en onder moeilike omstandighede met sy taak voortgegaan om seker te maak dat hierdie belangrike kanaal oop bly. Hiervoor is besondere erkenning en dank verskuldig aan die redakteur, dr. Steve Ndinga-Koumba-Binza, van die Universiteit van die Wes-Kaap, die resensieredakteur, me. Tanja Harteveld, sowel as me. Hermien van der Westhuizen, vir tegniese ondersteuning.

Die Buro van die WAT, as uitgewer van *Lexikos*, en daardeur as onmisbare en gewaardeerde vennoot van AFRILEX, moet weer eens van harte bedank word vir sy voortgesette toewyding tot die metaleksikografiese diskoers.

Dit is my voorreg om namens die Raad en lede van AFRILEX die redaksionele span, die Buro van die WAT en bydraende outeurs van harte te bedank vir nommer 30 van *Lexikos*.

Herman L. Beyer
President: AFRILEX

A Few Words from AFRILEX

What a year this has been. 2020 will undoubtedly go down as a year of brutal global, but especially personal disruption for surely all of us, and a year of unthinkable loss on so many terrains for so many people.

The African Association for Lexicography (AFRILEX) has also not been spared disruption. For the first time in its history, the Association's annual international conference could not take place, as a result of the Covid-19 pandemic. We could not, as originally planned, celebrate our silver jubilee in Stellenbosch. We will also not be able to do it in person next year, because due to current uncertainties the AFRILEX Board was compelled to ultimately decide that the international annual conference of 2021 should be planned as a fully virtual, online event. We will therefore have to do without the physical meeting of the AFRILEX family for another while.

With the cancellation of the annual gatherings of many an association this year, journals have become practically the sole channel for the dissemination of peer-reviewed research. *Lexikos* is no exception, as is evident from the large number of contributions by local and international authors to this year's volume. Amid the undeniable impact of Covid-19, the editorial team of *Lexikos* continued uninterrupted and under difficult conditions with its work to maintain this important channel. For this, special recognition and thanks are due to the editor, Dr Steve Ndinga-Koumba-Binza, of the University of the Western Cape, the reviews editor, Ms Tanja Harteveld, as well as Ms Hermien van der Westhuizen, for technical support.

The Bureau of the WAT, as publisher of *Lexikos*, and therefore as vital and appreciated partner of AFRILEX, should be heartily thanked again for its unwavering dedication to the metalexigraphic discourse.

It is my privilege to, on behalf of the Board and members of AFRILEX, sincerely thank the editorial team, the Bureau of the WAT and contributing authors for volume 30 of *Lexikos*.

Herman L. Beyer
President: AFRILEX

Redaksionele doelstellings

Lexikos is 'n tydskrif vir die leksikografiese vakspesialis en word in die AFRILEX-reeks uitgegee. "AFRILEX" is 'n akroniem vir "leksikografie in en vir Afrika". Van die sesde uitgawe af dien *Lexikos* as die amptelike mondstuk van die *African Association for Lexicography* (AFRILEX), onder meer omdat die Buro van die WAT juis die uitgesproke doel met die uitgee van die AFRILEX-reeks gehad het om die stigting van so 'n leksikografiese vereniging vir Afrika te bevorder.

Die strewe van die AFRILEX-reeks is:

- (1) om 'n kommunikasiekanaal vir die nasionale en internasionale leksikografiese gesprek te skep, en in die besonder die leksikografie in Afrika met sy ryk taleverskeidenheid te dien;
- (2) om die gesprek tussen leksikograwe onderling en tussen leksikograwe en taalkundiges te stimuleer;
- (3) om kontak met plaaslike en buitelandse leksikografiese projekte te bewerkstellig en te bevorder;
- (4) om die interdisiplinêre aard van die leksikografie, wat ook terreine soos die taalkunde, algemene taalwetenskap, leksikologie, rekenaarwetenskap, bestuurskunde, e.d. betrek, onder die algemene aandag te bring;
- (5) om beter samewerking op alle terreine van die leksikografie moontlik te maak en te koördineer, en
- (6) om die doelstellings van die *African Association for Lexicography* (AFRILEX) te bevorder.

Hierdie strewe van die AFRILEX-reeks sal deur die volgende gedien word:

- (1) Bydraes tot die leksikografiese gesprek word in die vaktydskrif *Lexikos* in die AFRILEX-reeks gepubliseer.
- (2) Monografiese en ander studies op hierdie terrein verskyn as afsonderlike publikasies in die AFRILEX-reeks.
- (3) Slegs bydraes wat streng vakgerig is en wat oor die suiwer leksikografie of die raakvlak tussen die leksikografie en ander verwante terreine handel, sal vir opname in die AFRILEX-reeks kwalifiseer.
- (4) Die wetenskaplike standaard van die bydraes sal gewaarborg word deur hulle aan 'n komitee van vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

Lexikos sal jaarliks verskyn, terwyl verdienstelike monografiese studies sporadies en onder hulle eie titels in die AFRILEX-reeks uitgegee sal word.

Editorial Objectives

Lexikos is a journal for the lexicographic specialist and is published in the AFRILEX Series. "AFRILEX" is an acronym for "lexicography in and for Africa". From the sixth issue, *Lexikos* serves as the official mouthpiece of the *African Association for Lexicography* (AFRILEX), amongst other reasons because the Bureau of the WAT had the express aim of promoting the establishment of such a lexicographic association for Africa with the publication of the AFRILEX Series.

The objectives of the AFRILEX Series are:

- (1) to create a vehicle for national and international discussion of lexicography, and in particular to serve lexicography in Africa with its rich variety of languages;
- (2) to stimulate discourse between lexicographers as well as between lexicographers and linguists;
- (3) to establish and promote contact with local and foreign lexicographic projects;
- (4) to focus general attention on the interdisciplinary nature of lexicography, which also involves fields such as linguistics, general linguistics, lexicology, computer science, management, etc.;
- (5) to further and coordinate cooperation in all fields of lexicography; and
- (6) to promote the aims of the *African Association for Lexicography* (AFRILEX).

These objectives of the AFRILEX Series will be served by the following:

- (1) Contributions to the lexicographic discussion will be published in the specialist journal *Lexikos* in the AFRILEX Series.
- (2) Monographic and other studies in this field will appear as separate publications in the AFRILEX Series.
- (3) Only subject-related contributions will qualify for publication in the AFRILEX Series. They can deal with pure lexicography or with the intersection between lexicography and other related fields.
- (4) Contributions are judged anonymously by a panel of highly-rated experts to guarantee their academic standard.

Lexikos will be published annually, but meritorious monographic studies will appear as separate publications in the AFRILEX Series.

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/30-1-1610>

Exploring the Documentation and Preservation of African Indigenous Knowledge in a Digital Lexical Database*

Sonja E. Bosch, *Department of African Languages,
University of South Africa, Pretoria, South Africa (boschse@unisa.ac.za)*
and

Marissa Griesel, *Department of African Languages,
University of South Africa, Pretoria, South Africa (griesm@unisa.ac.za)*

Abstract: Transcending the boundaries of printed lexicographic resources is becoming easier in the digital age, with e-resources facilitating restrictions on the size and type of information that can be included. In this article we explore innovative ways of documenting and preserving African indigenous knowledge, often underrepresented in traditional dictionaries, in an existing digital lexical database. Our approach is based on the extension of the African Wordnet, a lexical database under construction for nine African languages, in this case applied to isiZulu. This article addresses the challenge of consolidating dispersed indigenous knowledge collected from a variety of sources such as conventional dictionaries, interdisciplinary publications and a flat-structured online database, in a digitised hierarchical wordnet structure. A representative sample of traditional domestic utensils in Zulu culture is used to demonstrate the conversion into a set of typical semantic relations in a wordnet structure. By focusing on filling lexical gaps between isiZulu and English as found in the Princeton WordNet, with culturally relevant synsets, the African Wordnet also becomes a useful resource for natural language processing. Finally, it is shown how the hierarchical classification of selected domestic utensils is visually presented in wordnet graphs in the Wordnet-Loom interface.

Keywords: AFRICAN WORDNET, DIGITAL LEXICAL DATABASE, INDIGENOUS KNOWLEDGE, ISIZULU, TRADITIONAL DOMESTIC UTENSILS, SEMANTIC RELATIONS, LEXICAL GAPS, LEXICALISATION

Opsomming: Ondersoek na die dokumentasie en bewaring van inheemse kennis uit Afrika in 'n digitale leksikale databasis. Die digitale era maak dit toenemend moontlik om die grense wat gedrukte leksikografiese hulpbronne stel te oorskry deurdat

* An introduction to this theme was read as a paper at the 24th International Conference of the African Association for Lexicography which was held in Windhoek, Namibia, 26–29 June 2019. A first attempt towards the inclusion of pottery beer vessels in the AfWN was presented in Bosch and Griesel (2019).

e-hulpbronne die beperkings op grootte en tipe inligting wat ingesluit kan word, fasiliteer. In hierdie artikel ondersoek ons innoverende maniere om inheemse kennis uit Afrika, wat dikwels onderverteengewordig is in tradisionele woordeboeke, in 'n bestaande digitale leksikale databasis te dokumenteer en te preserveer. Ons aanslag word op die uitbreiding van die African Wordnet, 'n leksikale databasis wat tans vir nege Afrika-tale ontwikkel word, gebaseer en in hierdie geval op Zoeloe van toepassing gemaak. Die artikel spreek die uitdagings wat die konsolidasie van inheemse kennis uit 'n verskeidenheid bronne soos konvensionele woordeboeke, interdisiplinêre publikasies en 'n aanlyn databasis met 'n plat struktuur in 'n digitale, hiërargiese woordnetstruktuur inhou. 'n Verteengewordigende uittreksel van tradisionele huishoudelike gereedskap in die Zoeloe kultuur word as voorbeeld gebruik om te demonstreer hoe konsepte tot 'n stel semantiese verhoudings in 'n woordnetstruktuur omgeskakel kan word. Deur daarop te fokus om die leksikale gapings tussen Zoeloe en Engels soos dit in die Princeton WordNet vervat is, met relevante kulturele konsepte te vul, word die African Wordnet ook 'n nuttige hulpbron vir natuurliketaalprosessering. Laastens word aangedui hoe die hiërargiese klassifikasie van geselekteerde huishoudelike gereedskap visueel in woordnetgrafieke in die WordnetLoom koppelvlak voorgestel kan word.

Sleutelwoorde: AFRICAN WORDNET, DIGITALE LEKSIKALE DATABASIS, INHEEMSE KENNIS, ZOELOE, TRADISIONELE HUISHOUELIKE GEREEDSKAP, SEMANTIESE VERHOUDINGS, LEKSIKALE GAPINGS, LEKSIKALISERING

1. Introduction

In an article on large-scale lexicography in the digital age, Fellbaum (2014: 378) states:

The Digital Revolution can be fairly said to have shifted the paradigm in lexicology and lexicography. It has opened up new ways of exploring and representing the structure of the lexicon, testing diverse theories of word semantics, and compiling both manually and automatically ever larger and richer resources that reflect multiple dimensions of meaning and lexical organization based on solid empirical data ...

Fellbaum (2014) continues by describing the impact that the digital revolution has had on the construction of lexical resources containing extensive information on aspects of word meaning that are not easily covered in traditional print dictionaries. Gouws et al. (2014: 12) support this statement: "Online accessible corpora and data banks, digitized text editions and electronic editions of older and new dictionaries today offer the lexicographer who depends on sound and comprehensive documentation an ideal working basis".

The most significant limitations that have negatively influenced traditional paper dictionaries are identified as size of databases, access to databases, and methods according to which resources are compiled. When compiling traditional paper dictionaries, lexicographers are often forced to exclude infrequently used concepts and focus on modern language usage, commonly

determined by examining frequency lists extracted from large corpora. This is mostly due to restrictions placed on them by publishers mindful of printing costs and practicality. Since electronic or digital databases are generally not adversely affected by space or size constraints as is the case with printed matter, they lend themselves ideally to the inclusion of additional data, such as indigenous knowledge (IK) that is often underrepresented in conventional dictionaries. Initiatives by the South African Government and the Department of Science and Innovation to bolster the development of digital language resources under less restrictive licenses such as the establishment of the South African Centre for Digital Language Resources (SADiLaR)¹, further promote the accessibility of data. The limitations can therefore be side-lined or overridden in the digital age.

The question now is whether any existing frameworks or digital database structures can be utilised effectively in this move to a more digitally accessible lexicographic working base. Such a framework would need to offer the capacity to include (digitised) data from printed dictionaries and in addition, be easily expandable with less frequently used and underrepresented concepts, bringing the traditional into the digital age. It also needs to be easily accessible for both qualitative and quantitative research or development while being an easy to use reference for language learners or students. Finding such a solution will also address the problem of archiving various forms of missing or dispersed IK in a more sustainable database. Souza et al. (2020: 946) stress that the collection of IK in paper archives and more recently in digital databases is imperative in preserving not only the language, but also traditional customs for posterity.

In this exploratory article we describe a novel application and subsequent expansion of an existing lexical resource for isiZulu which enhances and supplements lexical knowledge from conventional dictionaries and other interdisciplinary sources. We introduce the African Wordnet (AfWN), a prototypical lexical database consisting of words that are grouped into sets of synonyms called synsets, as framework for the digital documentation and preservation of indigenous or cultural knowledge. The opportunities that such a digital lexical database not constrained by size and access can offer with regard to the digital documentation and preservation of indigenous or cultural knowledge, are addressed.

In order to contextualise our study, we start by explaining in general what a wordnet is, including the macrostructure and microstructure of a wordnet such as the English Princeton WordNet (PWN). This is followed by an overview of the African Wordnet (AfWN), currently under active development, and a description of the wordnet editor implemented to provide visualisation features of wordnets. Next, we describe our approach to the design of the AfWN, with a focus on challenges such as lexicalisation and lexical gaps, particularly within the context of the African languages. This is followed by a demonstration of how a representative sample of traditional Zulu domestic utensils, gathered from an interdisciplinary variety of sources, can be transformed into a set

of relations in an electronic lexical resource. We then conclude and provide pointers for future work.

2. Aspects of wordnets

2.1 What is a wordnet?

McCrae et al. (2020: 37) maintain that wordnets have turned out to be one of the most popular types of dictionaries used in natural language processing (NLP) and other areas of language technologies. This can be ascribed mainly to their "structure as a graph of words, that is much easier for computers to understand than the traditional form of a dictionary". Wordnets are primarily built for use by machines in tasks such as automatic text analysis and for artificial intelligence applications. Word sense disambiguation (WSD) and information retrieval (IR), for instance, are performed much more effectively when the semantic relations in a wordnet are used to distinguish between the different meanings of a word in context. It is therefore not surprising that according to Calzolari (2018), wordnets for various languages were among the most cited resources during the Language Resource and Evaluation Conference (LREC) in 2018. Wordnets offer a wealth of information, contained in a machine-readable lexical database where words are grouped into synsets and linked by conceptual-semantic and lexical relations (Miller 1995). An online lexical database for a specific language is, furthermore, an invaluable reference resource for many research and application projects in the linguistics and lexicography domains. Kotzé (2008: 20) states: "[a wordnet is] extremely useful for its accessibility, quick reference and potential for serving as a base or support for other language technological and lexicographical applications".

Each synset in a wordnet is enhanced with lexical information such as the part of speech of the lemmas, a definition and usage example(s) of the concept. Furthermore, in addition to the synonymy relation linking different senses in a synset, other semantic relations between synsets are indicated as well. These relations include the super-subordinate or the hyperonymy, hyponymy relation, the part-whole or meronymy relation, as well as antonymy. Fellbaum (1998: 7) explains: "WordNet is a semantic dictionary that was designed as a network, partly because representing words and concepts as an interrelated system seems to be consistent with evidence for the way speakers organise their mental lexicons".

The first large-scale project to develop a monolingual wordnet was started in the 1990s with the Princeton University WordNet (PWN) for (American) English (Fellbaum 1998) and contains roughly 250 000 synsets for nouns, verbs, adjectives and adverbs. An example of a synset from the PWN can be seen in Figure 1. Note the hierarchical organisation of the synsets using SUMO/MILO classification (cf. SUMO 2002; as well as Niles and Pease 2001) and how this is

easily visualised in the WordnetLoom (Naskret et al. 2018) interface. This makes the PWN and any wordnet developed according to the same ontological structure easy to navigate, both manually and automatically in different digital humanities (DH) applications.

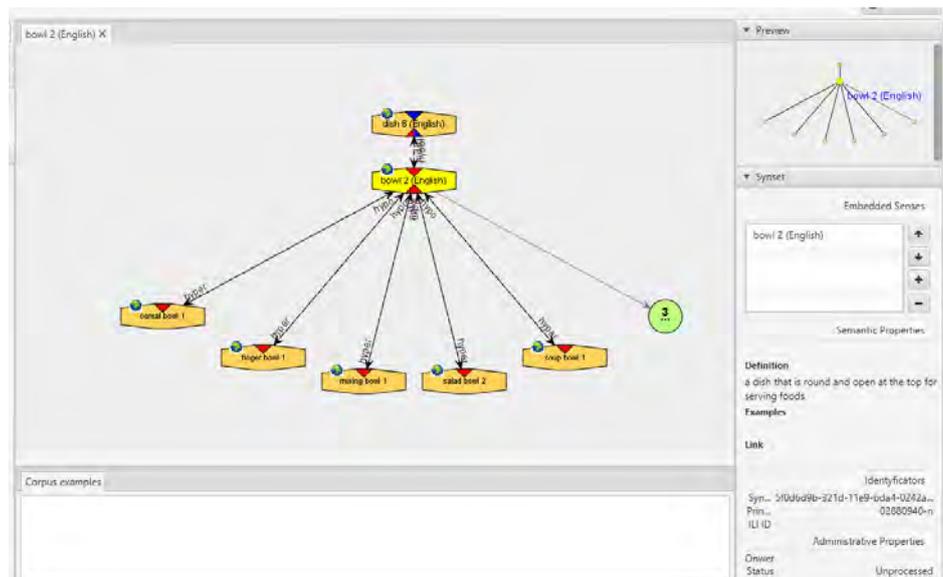


Figure 1: The noun synset 'bowl' in PWN² as visualised in WordnetLoom

The PWN serves as template for many development projects, such as the Hungarian wordnet (Vincze and Almási 2014), the Japanese wordnet (Bond et al. 2008) and the BalkaNet wordnets (Tufis et al. 2004). The hierarchical structure and semantic relations are kept largely unchanged and only the content of each synset, i.e. the lemmas, usage example and definition, is translated into the target language. This, of course, assumes that the target language shares an underlying semantic structure with English as it is captured in the PWN. Ordan and Wintner (2007) and Vossen et al. (2016) refer to this method as the *expand approach* and recommend it for projects with limited lexical resources. If these resources do exist, the so-called *merge approach* is often followed. The PolNet wordnet (Vetulani et al. 2010) was for instance derived from a high quality, monolingual Polish lexicon as base and subsequently aligned with the PWN. Similarly, wordnets are also aligned with the Multilingual Open Wordnet (Bond and Paik 2012) or Global WordNet Grid (Vossen et al. 2016).

2.2 Base concepts

The method by which to develop a wordnet for a new language does not only entail choosing between the expand and the merge approach as described above. Another very important consideration is also which concepts to include, especially in the initial stages of development, to ensure good coverage in the wordnet. To this end, the EuroWordNet and BalkaNet projects created the so called "core base concepts" (CBC) list — a list of seed terms extracted from corpora for various European languages involved in the two projects with which to kickstart wordnet development³. The CBC aims at including those high-level concepts that have many semantic relations with other synsets first, thereby guiding manual inclusion of further synsets in a top down approach. The AfWN also initially followed this internationally accepted method by including most of the CBC in the first version of the wordnets.

However, since the CBC incorporates many concepts that are not lexicalised in the African languages, Anderson et al. (2010: 3763) point out the main disadvantage of the particular approach, namely that "the fundamental WordNet base will be biased to those concepts that are not necessarily core in the new target language". They propose a hybrid approach to building wordnets for African languages, in keeping with the global focus that has "always been on concept hyponymy based on mother tongue speaker understanding" (Anderson et al. 2010: 3763).

Accordingly, the AfWN resorted to incorporating synsets from a more localised seed list — the SIL Comparative African Wordlist (SILCAWL), which was compiled in 2006 by Keith Snider (SIL International and Canada Institute of Linguistics) and James Roberts (SIL Chad and Université de N'Djaména). This list, covering 12 semantic categories, is an English–French bilingual list of lexical data consisting of 1 700 words with glosses, resulting from linguistic research in Africa. Inclusion of seed terms from this list has already resulted in the inclusion of numerous lexicalised concepts such as the elaborate kinship terms in isiZulu and Sesotho sa Leboa in the AfWN in an organised manner (cf. Griesel et al. 2019).

2.3 African Wordnet

The African Wordnet (AfWN) which is still under active development (see Bosch and Griesel 2017; and Griesel and Bosch 2020) aims at steadily growing the number of synsets in wordnets for nine indigenous South African languages. Over the past 10 years of development, the AfWN has grown slowly but consistently with a focus on manually verified, culturally appropriate data that meets international standards. Currently, the AfWN includes roughly 63 000 synsets across the nine languages with 27 000 definitions and 37 000 usage examples added to these synsets. The languages currently included in the project

are isiZulu (ZUL), isiXhosa (XHO), Setswana (TSN), Sesotho sa Leboa (NSO), Tshivenda (VEN), Siswati (SSW), Sesotho (SOT), isiNdebele (NDE) and Xitsonga (TSO). The multilingual wordnet project is a first for these South African languages.

Although there has been an increase in the number and quality of text resources freely available for the South African languages, a recent audit still showed large areas where little to no development has been done and the languages involved in the AfWN can still all be considered resource-scarce. As Moors et al. (2018: 2) conclude in their report on the most recent audit of available resources: "While significant progress has been made since 2009 to develop additional resources across more languages, to develop cutting edge resources, and to develop language independent resources, the more marginalised indigenous languages (particularly Xitsonga, Tshivenda, Sesotho, siSwati, and isiNdebele), remain severely under resourced". This resource scarceness has been one of the main challenges in the development of wordnets and the reason why the *expand approach* was followed, as described in the previous section. Neale (2018) outlines the expand approach as starting point for wordnet construction when few additional resources exist and describes various methods by which automatic extraction of the basic information needed to form a synset can speed up development. However, the basic linguistic and lexicographic resources he describes such as an electronic bilingual dictionary, or even a monolingual lexicon in the target language are not yet freely available for most of the South African languages in the AfWN. This not only makes automatic extraction of synsets impossible, but necessitates a more labor intensive manual process (see Griesel and Bosch (2020) for a discussion of the various methods employed to make the most of the limited resources that do exist for some languages).

The AfWN development team will continue to add more synsets to this base, but we are now ready to delve deeper into language specific questions relating to smaller groups of synsets in order to broaden the coverage and make the AfWN useful not only for language processing applications, but also for human interpretation in the lexicographic process, for instance. An initial experiment into employing the AfWN data in further application saw data from the AfWN used to start populating Kamusi GOLD⁴, a multilingual online dictionary. Furthermore, a free and open mobile dictionary app with an underlying data structure that is not only open for dictionaries, but supports the connection to external resources like the AfWN as well, is also under development (Eckart et al. 2020).

2.4 Editing tool

A wordnet by nature depends on the interconnected relations between concepts as formalised in synsets. Visualising the relations and effortlessly adding

to the network of connections is a crucial aspect of manual development, even more so when working on a multilingual project. At the onset of the current development phase the AfWN was ported to WordnetLoom (Naskret et al. 2018), a freely available, customisable wordnet editor with advanced wordnet visualisation features. The editor has the capability of organising large networks of semantic relations and serves as a browser and development interface. The visual nature of the interface is easy to work in and facilitates visualisation of the connections between concepts. It is envisioned that this tool will also form the basis for a browsing platform to make the AfWN openly accessible and easily searchable as a web service, thereby eliminating the need for specialised software or installation by an expert.

3. Design strategy

Our approach to the design of the AfWN takes place against the background of the wordnet being a source of reference that takes the traditional dictionary to a whole new level, as described by Abubakar et al. (2019). While a dictionary organises words in alphabetical order and can offer information such as meaning, synonyms, parts of speech (POS) and so forth, a wordnet offers the additional feature of synsets, a set of synonyms (synsets) for open word classes, that is nouns, verbs, adjectives and adverbs. Linkage to synsets is provided by various semantic relations such as hypernyms, hyponyms, meronyms, troponyms and antonyms. In the description of our design in this article, we focus on nouns in particular.

In the PWN, which is the point of referral in the expand method described above, the super-subordinate relation (also known as hyperonymy, hyponymy) is the most frequently encoded relation among synsets. This relation links universal synsets like {utensil} to gradually more specific ones like {ceramic ware; pottery; clayware; funnel; server, kitchen utensil etc.}, as can be gleaned from Figure 2. This is an advantage of the expand method since upper levels provide general guidelines, so that the set of relations (hyponyms) can be increased to fill the lexical gaps with indigenous or cultural knowledge concepts of the target language. The vocabulary of a language can be divided into two categories according to Batibo (2016: 135), namely basic vocabulary and cultural vocabulary. He describes basic vocabulary as "the lexical stock which is basic in all human languages. It denotes objects and phenomena that are found universally" while cultural vocabulary is "the lexical stock that a linguistic community develops or adopts through its many cultural experiences after interacting with its physical environment, social milieu and the supernatural world". An aspect of the AfWN that needs to be explored further is the addition of such "cultural" word senses that cannot be linked directly to the PWN. African languages and cultures include many unique word senses that are not easily matched to the core set of meanings in the PWN or for that matter in other wordnets. Such

concepts are not only those that appear in paper dictionaries but also those that may be documented elsewhere and that would be lost to future generations if not preserved digitally and made accessible, ideally in an organised manner.

The mere translation of English concepts contained in the PWN into isiZulu, is not a complete reflection or representation of Zulu cultural knowledge and concepts, especially not on the lower levels of the hierarchy. Therefore, in this study, we take the first steps in creating an isiZulu lexical database that addresses lexicalisation differences. "Lexicalisation differences" are here defined as a) those instances where the source and target languages, here English and isiZulu respectively, lexicalise the same concept with a different kind of lexical unit, be it a word, compound or collocation; and b) those instances where one of the two languages has no lexicalisation for a concept at all and results in a lexical unit in either the source or target language being translated with a description of the concept as a phrase. In the latter case, we therefore have a so-called lexical gap, which Bentivogli and Pianta (2000: 664) define as follows:

A lexical gap occurs whenever a language expresses a concept with a lexical unit whereas the other language expresses the same concept with a free combination of words.

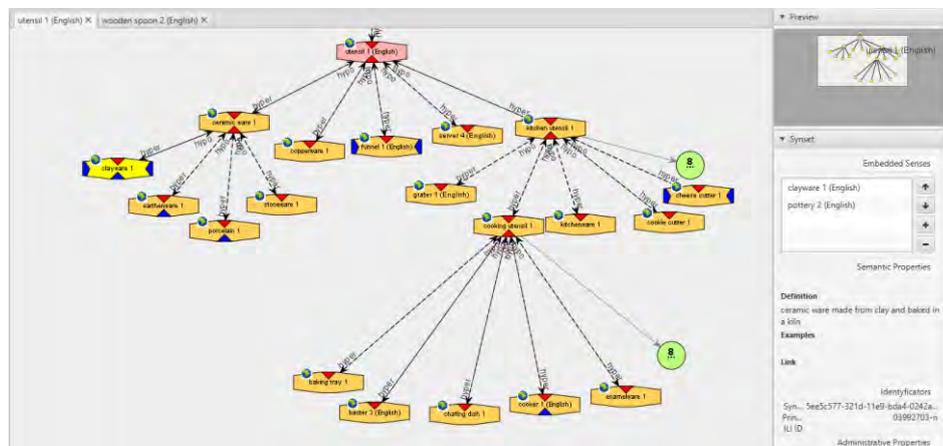


Figure 2: The noun synset 'utensil' in PWN as visualised in WordnetLoom

To fill such gaps, the SIL Comparative African Wordlist (SILCAWL) (Snider and Roberts 2006) is used as benchmark. The items are organised semantically on a continuum, from items relating to human domains at the one extreme, via animate domains, to items relating to non-human domains at the other extreme, and then from concrete items to more abstract items. The twelve main headings are listed in Table 1.

1.	Man's physical being
2.	Man's non-physical being
3.	Persons
4.	Personal interaction
5.	Human civilisation
6.	Animals
7.	Plants
8.	Environment
9.	Events and actions
10.	Quality
11.	Quantity
12.	Grammatical items

Table 1: Headings in the SILCAWL

The headings in Table 1 are subsequently sub-divided into second and third level headings. For example, in the case of *Human civilisation*, the following first level headings are distinguished: SETTLEMENT, CLOTHING AND ADORNMENT OF BODY, FOOD AND DRINK, FOOD PREPARATION, DOMESTIC UTENSILS AND CONTAINMENT, HABITATION, PROFESSIONS AND WORK, and so on. A third level, for instance, in the case of DOMESTIC UTENSILS AND CONTAINMENT includes divisions such as: kitchen utensils, eating utensils and containers, and containment. The parts of speech covered in the SILCAWL are nouns, verbs, adjectives, adverbs, pronouns, interrogatives and conjunctions. Although Snider and Roberts (2006: 4), the compilers of the SILCAWL, concede that they still notice "imperfections and room for improvement (e.g. words that could be deleted, words that could be added, words that could be moved to different semantic domains etc.)", the SILCAWL has proven to be an opportune progression from the CBC list used in the past in the development of the AfWN. The most significant improvement is observed against the background of localisation where the content (of the entries) would be lexicalised within an African environment.

Although the so-called SILCAWL is used as a guide to fill lexical gaps in the African Wordnet, it becomes clear when looking at a fragment of DOMESTIC UTENSILS AND CONTAINMENT in Table 2, that language specific detail needs to be addressed. For instance, in Zulu culture a lexical and conceptual distinction is made between a three-legged cooking pot and a small flat-bottomed cooking pot. Similarly, a lexical and conceptual distinction is made between a wooden spoon for eating or one for serving and stirring, or between

a calabash milk vessel and a calabash beer vessel. These concepts are all lexicalised in isiZulu and will be dealt with in more detail in the next section.

5.5 DOMESTIC UTENSILS AND CONTAINMENT

5.5.1 Kitchen utensils

0610 cooking pot (earthenware)

...

0613 ladle

...

0618 pestle, pounding stick pylon

0619 mortar, pounding pot

5.5.2 Eating utensils

0620 plate

0621 bowl

0622 cup (French: gobelet, coupe, tasse > goblet, cup)

0623 spoon (traditional)

5.5.3 Containers and containment

...

0626 basket

0627 bucket, pail

0628 calabash

Table 2: Fragment of domestic utensils and containment in the SILCAWL

A typical lexical gap that emerges in the wordnet of the source language (PWN), and which can be filled by indigenous knowledge information, is illustrated in Figure 3. Apart from the six hyponyms for "plate", viz. dessert plate, dinner plate, paper plate, salad plate/bowl, soup plate and steel plate, the isiZulu wordnet requires an additional hyponym, namely "earthenware children's plate", translated as *isikhangezo*.

The image shows a screenshot of a WordNet search interface. At the top, it says "WordNet Search - 3.1" with links to "WordNet home page", "Glossary", and "Help". Below this is a search bar with "plate" entered and a "Search WordNet" button. The results are under the heading "Noun". A list of definitions for "plate" is shown, including "home plate", "plate (a sheet of metal or wood)", "plate (a full page illustration)", and "plate (dish on which food is served)". A green box highlights the definition "plate (dish on which food is served...)" and its hyponyms: "dessert plate", "dinner plate", "paper plate", "salad plate", "salad bowl", "soup plate", and "steel plate". A separate green box points to the search results with the text: "Lexical gap: No earthenware plate or children's plate in the Princeton Wordnet".

Figure 3: Example of a lexical gap in the PWN⁵

4. Data and presentation

Cosijn et al. (2002) describe IK as local knowledge that is unique to every culture or society, and that is rooted in community traditions, relationships and rituals. The African continent is rich with sources of IK and Ossai (2010: 2) summarises it as "an embodiment of different modes of thought and epistemology". He also notes the importance of preservation of African IK and points out that it forms part of the decision-making process at the local level for rural communities, involving all aspects of human life — agriculture, health care, food preparation, education, natural resource management, and others. It is therefore important that IK be preserved in such a way as to make it easily accessible to all, including taking advantage of new developments in the digital age, and easily expandable so as to ensure clusters of IK practitioners can add valuable details regarding the intricacies that make African IK so unique and expansive.

According to Cosijn et al. (2002: 94):

Using databases for the representation of IK may offer several advantages. Most importantly, access from a retrieval point of view is much easier in electronic database format than in paper or linear electronic text formats. Secondly, IK can be stored and delivered in multiple copies for those that need it. Furthermore, in

database format, it is possible to annotate IK in various ways from multiple viewpoints to facilitate its analysis. However, in order to realise these advantages, IK in databases must be made accessible.

In this section it will be shown by means of isiZulu examples, how IK concepts gathered from an interdisciplinary variety of sources can be transformed, in some cases from alphabetically ordered entries and in other cases from categorized lists, into a set of relations within the context of a hierarchical wordnet structure. Such a digital knowledge database has almost no physical restrictions and may incorporate, in addition to the conventional wordnet relations, namely synonyms, hypernyms, meronyms and so forth, typical wordnet features such as definitions, usage examples and even dialect information. The data was collected from a variety of sources. Given the fact that dictionaries are the most common resource used for building wordnets, we first of all consulted three authoritative monolingual and bilingual dictionaries (ISZ 2006; ISN 1992 and ZED 1964). To complement the data collected from the dictionaries, two anthropology publications (Krige 1965; and Fowler 2015), diverse cultural publications (Nyembezi and Nxumalo 1966; and Grossert 1985) and an online database, Comparative Bantu Pottery Vocabulary (CBPV) were studied as well.

For the purposes of this study, we will focus on the following aspects of the taxonomy of traditional Zulu domestic utensils, namely function, material (from which the utensil is crafted), size and shape. Krige (1965), Nyembezi and Nxumalo (1966) and Grossert (1985) itemise domestic utensils according to the material from which the utensils are crafted, while Fowler (2015) classifies earthenware or pottery utensils according to their function. Size and shape of the utensils are described arbitrarily in the various sources. In the following discussion we compare descriptions of selected domestic utensils in dictionaries and other resources, point out missing entries in dictionaries — also entries that are not main entries (and would therefore be difficult for users to find), and identify synonyms and meronyms where applicable.

4.1 Kitchen utensils

- **cooking pot** (earthenware)

ikhanzi (plural: *amakhanzi*)

Function	-	cooking
Material	-	pottery
Size	-	large
Shape	-	three-legged, spherical pot with a wide mouth, about 23 centimetres in diameter

In the dictionaries ZED (1964: 381) and ISN (1992: 224), as well as in Nyembezi and Nxumalo (1966: 18) it is stated that *ikhanzi* is a cooking pot. Only ZED (1964: 381) indicates the material as earthenware. ISN (1992: 224) and ISZ (2006: 554) add the additional information that *ikhanzi* is large and three-legged. The shape is

gleaned from other resources, viz. Fowler (2015: 97) who describes it as oval-shaped (accompanied by an illustration), and CBPV as a pot with a wide mouth. CBPV also provides detailed information about size — about 9 inches (23 centimetres) in diameter. This utensil does not feature in Grossert (1985).

isoco (plural: *izoco*)

Function	-	cooking (cereals or vegetables)
Material	-	pottery
Size	-	small
Shape	-	bowl

The dictionary sources ISZ (2006: 936), ISN (1992: 388) and ZED (1964: 634) clearly state that *isoco* functions as an earthenware cooking vessel, while ISZ (2006: 936) and ZED (1964: 634) also comment on the size, characterising it as "small". This is confirmed by Fowler (2015: 98) who adds that it is used specifically for cooking cereals or vegetables, and provides an illustration. This utensil does not feature in CBPV.

isiyoco (plural: *iziyoco*)

Function	-	cooking (cereals or vegetables), frying
Material	-	pottery
Size	-	small
Shape	-	saucepan

The kitchen utensil *isiyoco* does not feature in any of the dictionaries consulted (i.e. ZED, ISN and ISZ), however, it appears in Nyembezi and Nxumalo (1966: 20) where it is described as a frying pan; while Krige (1957: 397) adds the details of size "little saucepan used for cooking, smaller than the ikhanzi". The illustration in Fowler (2015: 98) shows the comparative sizes of *isiyoco* and *isoco*, the former being the smaller of the two cooking pots. This utensil features neither in Grossert (1985) nor in CBPV.

- **ladle**

isikhetho (plural: *izikhetho*)

Function	-	beer-skimmer
Material	-	ilala palm leaves
Size	-	small, flat, curved handle
Shape	-	spoon-like
Meronym:	<i>isibambo</i>	flat and curved handle

Both ZED (1964: 391) and ISN (1992: 527) describe *isikhetho* as a plaited spoon-like utensil used for skimming beer. The former dictionary adds that the utensil

is plaited from palm leaves. It is noteworthy that *isikhetho* does not feature in ISZ (2006).

Krige (1965: 396) provides information on the size of the spoon, namely small, and also specifies the function as that of skimming the beer before it is drunk. Grossert (1985: 23) confirms the function as "When beer is served the *isikhetho* is taken to skim the scum from surface", and she also identifies the palm leaves as those of the ilala palm. The shape of the handle is described by Grossert (1985: 41) as flat and curved.

A meronym or part of the whole is observed in this kitchen utensil, namely a handle *isibambo*. The handle is also made of ilala palm leaves.

imvokoqo (plural: *izimvokoqo*)

Function	-	serving
Material	-	wood
Size	-	large
Shape	-	hollowed out, deep bowl

This utensil is unanimously described by ZED (1964: 837), ISZ (2006: 726) and ISN (1992: 527) as a ladle with a deep bowl. Nyembezi and Nxumalo (1966: 17) concur and add that it is made of wood and used for serving. The serving function is confirmed by ISN (1992: 527). Although the description by Krige (1965: 398) seems to be similar to the ones above, namely a deep-ladled wooden spoon, the term used – *umvokoqa* – is not the same and will therefore not be taken into consideration. This utensil does not appear in Grossert (1985).

4.2 Eating utensils

- bowl

umcakulo (plural: *imicakulo*)

Function	-	eating
Material	-	pottery
Size	-	small (15-20 cm in diameter)
Shape	-	small pot with a wide mouth, shaped like a pudding bowl/basin

Synonym: *umshengele*

All dictionaries consulted, as well as the additional resources, describe *umcakulo* as a small earthenware bowl used for eating. ZED (1964: 100) adds details of the shape, namely "shaped like pudding basin" and "wide mouthed". Grossert (1985: 36) provides details on the size, namely a diameter of 15 to 20 cm. Fowler (2015: 101) adds the specific dishes that are served in these bowls, which include *uphuthu* (mealie meal porridge), *amahewu* (fermented maize porridge drink) and *umdokwe* (porridge).

Nyembezi and Nxumalo (1966: 19); ZED (1964: 737); ISZ (2006: 1102) and ISN (1992: 463) identify *umshengele* as synonym for *umcakulo*. This utensil does not feature in CBPV.

imbenge (plural: *izimbenge*)

- | | | |
|----------|---|------------------------------|
| Function | - | eating |
| Material | - | grass |
| Size | - | small (15-20 cm in diameter) |
| Shape | - | saucer-shaped bowl |

Synonyms: *umhelo*
unyazi

According to Krige (1965: 395) *imbenge* is a basket of woven grass used for serving food, specified by Grossert (1985: 17) as cooked maize, millet or pulses, and by ISN (1992: 295) as boiled maize and sorghum. In ZED (1964: 73) only the size (small) and the material of *imbenge* are described, while Nyembezi and Nxumalo (1966: 21) also describe the shape of the bowl, namely broad and shallow with a wide mouth. Grossert (1985: 17) pays more attention to exact shape and size, namely saucer-shaped and approximately 15 to 20 centimetres in diameter.

It is interesting to note that ISZ (2006: 686) does not include the function of using *imbenge* to serve food, but rather the alternative function of covering food or beer — "*isitsha esakhiwe ngotshani sokwemboza ukhamba noma ukudla*" (a grass bowl for covering food or beer).

Nyembezi and Nxumalo (1966: 21) and Krige (1965: 395) indicate *ingcazi* as synonym, although three of the dictionaries describe *ingcazi* as a different type of utensil. ZED (1964: 550) defines it as a narrow-necked water-pot and not as an eating bowl, while ISZ (2006: 1102) and ISN (1992: 330) define it as a large earthenware pot with a long small mouth, also known as *uphiso*. Until further clarity has been gained, *ingcazi* will not be included as a synonym in the isiZulu wordnet. However, ZED (1964: 73) offers *umhelo* as synonym, even if not as a main entry. ISZ (2006: 919) and ISN (1992: 382) indicate *unyazi* as synonym of *imbenge*.

- **plate**

isikhangezo (plural: *izikhangezo*)

- | | | |
|----------|---|--------------------------------------|
| Function | - | feeding children soft foods, liquids |
| Material | - | pottery |
| Size | - | small (size of a saucer) |
| Shape | - | shallow, shape of a saucer |

Synonym: *umgenqele*

Most sources agree that *isikhangezo* is a small earthenware plate used for feeding children soft foods such as sour milk (ISZ 2006: 552; ZED 1964: 380; ISN 1992: 223; Nyembezi and Nxumalo 1966: 20; and Krige 1965: 397). The shape of the plate is specified by Krige (1965: 397) and ZED (1964:380) as being that of a saucer.

Fowler (2015: 100) is the only source offering a synonym *umgenqele*. This utensil does not feature in CBPV.

- **spoon (traditional)**

ukhezo (plural: *izinkezo*)

Function	-	serving
Material	-	wood
Size	-	average length 30 cm
Shape	-	spoon

Meronym: *isibambo*
long handle

It is stated by all sources, except ISZ (2006), that *ukhezo* is a wooden spoon (Krige 1965: 398; Nyembezi and Nxumalo 1966: 17; ZED 1964: 392; and ISN 1992: 231). Grossert (1985: 40-41) gives an average length of 30 cm, and emphasizes the balance between the length and thickness of the handle compared to the size and shape of the bowl of the spoon. Nyembezi and Nxumalo (1966: 17) add additional information to the function, namely a wooden spoon for eating sour milk or other liquid food such as pumpkin porridge.

A meronym or part of the whole is noted in this utensil, namely a handle *isibambo*. It is a long wooden handle as described by Grossert (1985: 40-41).

isixembe (plural: *izixembe*)

Function	-	serving and stirring
Material	-	wood
Size	-	large
Shape	-	ladle (broad and flat)

Synonyms: *ingxwembe*
isixwembe
ingxembe
isixembu

The dictionaries ZED (1964: 860) and ISN (1992: 231) indicate the material from which *isixembe* is crafted as wood, while ISN (1992: 231) and ISZ (2006: 1285) indicate the size as large. ISN (1992: 231) adds the function as that of serving. Nyembezi and Nxumalo (1966: 18) supply a more complete description, namely that of a wooden stirring spoon with a broad flat shape in front.

A number of synonyms are put forward by the sources consulted. *ingxwembe* appears in Nyembezi and Nxumalo (1966: 18) and ZED (1964: 860) although not as a main entry in the latter dictionary; *isixwembe* appears in Krige (1965: 398), ISZ (2006: 1285) and ISN (1992: 545); *ingxembe* features in ISN (1992: 545) and *isixembu* appears in ISZ (2006: 1285). Even though Krige (1965: 398) records only *isixwembe*, she describes it as a "large, broad, wooden spoon for serving food or eating *ububendi* (blood and mince)".

4.3 Containers and containment

- calabash

igula (plural: *amagula*)

Function	-	container for sour milk
Material	-	calabash
Size	-	not specified
Shape	-	short with wide body and narrow mouth

All sources consulted (Nyembezi and Nxumalo 1966: 21; Krige 1965: 397; ZED 1964: 274; ISZ 2006: 404; and ISN 1992: 156) agree that *igula* is a calabash vessel used for sour or curdled milk. The dictionaries ZED (1964: 274); ISZ (2006: 404) and ISN (1992: 156) provide additional information on function, namely a container for fermenting milk into sour milk; and on shape, that is, short with a wide body and a narrow mouth.

iphaphasi (plural: *amaphaphasi*)

Function	-	container for food
Material	-	calabash
Size	-	not specified
Shape	-	flat sides, wide, open-mouthed

Two of the dictionaries consulted (ISZ 2006: 957 and ISN 1992: 398), as well as Nyembezi and Nxumalo (1966: 21), all describe *iphaphasi* as a calabash container for food. Concerning the shape, ZED (1964: 648) describes it as a wide, open-mouthed calabash, ISN (1992: 398) describes it as a calabash with dented sides while Nyembezi and Nxumalo (1966: 21) point out that it is dented on the sides to form a flat object. Krige (1965: 397) confirms that *iphaphasi* is a flat calabash with flat sides and top is cut off.

This utensil does not feature in Grossert (1985).

isigubhu (plural: *izigubhu*)

Function	-	container for water, beer or fermented maize porridge
Material	-	calabash
Size	-	not specified
Shape	-	not specified

Although ZED (1964: 274) only describes *isigubhu* as a calabash or gourd, the other two dictionaries (ISZ 2006: 398 and ISN 1992: 60) as well as Nyembezi and Nxumalo (1966: 21), add the function, namely that of a container for water, beer or *amahewu* (fermented maize porridge). The latter specifically point out that it is not used as container for sour milk or food. In Krige (1965: 397) the orthography differs slightly (*isigubu*) but refers to the same type of vessel, namely a gourd or calabash used as water or beer container. Grossert (1985: 37) provides two illustrations of calabash containers which she incorrectly terms *isigubo*. There is no accompanying description.

It is interesting to note that the size and shape of the various calabashes are not specified, possibly because calabashes are natural objects that have different sizes.

- **basket**

isichumo (plural: *izichumo*)

- | | | |
|----------|---|---|
| Function | - | container for carrying beer |
| Material | - | ilala palm leaves |
| Size | - | 25-40 cm in diameter with cylindrical neck, about 7 cm across and 5 cm high |
| Shape | - | globular |

Meronym: *isimbozo*

lid or covering for beer container (*isichumo*) made of ilala palm leaves, fits over the neck like a cap.

The dictionary sources consulted all concur that *isichumo* depicts a container for carrying beer, and that it is crafted from ilala palm leaves (ZED 1964: 116; ISZ 2006: 151 and ISN 1992: 60). ZED (1964: 116) describes it as being bottle-shaped, while ISZ (2006: 151) and ISN (1992: 60) describe *isichumo* as having a long narrow mouth ("esinomlomo omude omncane"). The shape is expressed more accurately by Krige (1965: 395), namely globular. Grossert (1985: 19-20) adds the specific details that the spherical part of *isichumo* is 25 to 40 cm in diameter while the cylindrical neck is about 7 cm across and 5 cm high. ZED (1964: 116) as well as ISZ (2006: 151) add that the material of which this container is made, could also be calabash.

A part of the whole or a meronym is observed in this utensil, namely a lid *isimbozo* for covering the mouth of the container (Krige 1965: 395) and fitting over the neck like a cap Grossert (1985: 21). The lid is also made of ilala palm leaves.

isilulu (plural: *izilulu*)

- | | | |
|----------|---|--|
| Function | - | container for grain storage |
| Material | - | (plaited <i>imbubu</i>) grass |
| Size | - | large, 1 to 1,5 metres in diameter |
| Shape | - | pear shaped, broad base with small opening |

The three dictionaries consulted (ZED 1964: 467; ISN 1992: 398 and ISZ 2006: 670) agree that *isilulu* is a container for storing grain. ZED (1964: 467) and ISN (1992: 398) add that it is made of plaited grass and is large, while (ZED 1964: 467) also gives information about the shape of the container namely "the opening of which is small". Nyembezi and Nxumalo (1966: 22) concur that *isilulu* is a large container for storage of the harvest, made of grass. Krige (1965: 395) provides more detail regarding size, namely very large, with a diameter that can be up to 3 to 4 feet. The most detailed description is given by Grossert (1985: 21), who depicts *isilulu* as the largest of the Zulu baskets with a diameter of 1 to 1,5 metres. She describes it as "pear shaped rather than spherical, with a broad base ... woven with bunches of soft *imbubu* grass".

4.4 Proposed synsets

Proposed synsets to be included in the isiZulu wordnet, based on information for selected domestic utensils as extracted from various sources and analysed in detail above, are summarised in Table 3.

Type of domestic utensil	Definition	Usage example
<i>ikhanzi</i>	a large earthenware three-legged, spherical cooking pot with a wide mouth, about 23 cm in diameter <i>imbiza enkulu yokuphekela yebumba enemilenze emithathu, eyindilinga enomlomo obanzi, engamasentimitha angama-23 ububanzi</i>	<i>sebenzisa ikhanzi ukupheka isibindi nosu eziko</i>
<i>isoco</i>	a small earthenware cooking vessel used specifically for cooking cereals or vegetables <i>imbiza encane yokuphekela esetshenziselwa ukupheka ukudla okuyizinhlamvu noma imifino</i>	<i>letha isoco bakuphekele iklabishi</i>
<i>isiyoco</i>	a small earthenware saucepan, smaller than <i>isoco</i> , used for cooking or frying cereals or vegetables <i>isosipani lebumba elincane kunesoco, elisetshenziselwa ukupheka noma ukuthosa ukudla okuyizinhlamvu noma imifino</i>	<i>ugogo ucela isiyoco azopheka ngaso iklabishi nezindlubu</i>
<i>umcakulo</i> Synonym: <i>umshengele</i>	a small earthenware bowl with a diameter of 15-20 cm, used for serving <i>isitsha sebumba esincane esinedayametha eyi-15 kuye ku-20 cm, esisetshenziselwa ukuphakela</i>	<i>ugogo ucela umcakulo azobanikeza ngawo amahewu, umdokwe noma uphuthu</i>

<i>imbenge</i>	a small saucer-shaped bowl of woven grass, used for serving food	
Synonyms: <i>ingcazi</i> <i>umhelo</i> <i>unyazi</i> <i>isikhangezo</i>	<i>isitsha esincane esakheke okwesosi esakhiwe</i> <i>ngotshani obelukiwe esisetshenziselwa ukuphakela ukudla</i> a small, shallow earthenware plate used for feeding children soft food <i>ipuleti elincane, elibunjiwe elisetshenziselwa ukufunza izingane ukudla okuthambile</i>	<i>letha imbenge sizobekela obaba ukudla</i> <i>unina kaSipho ugeza isikhangezo sosana lwakhe</i>
<i>ukhezo</i>	a wooden serving spoon for eating sour milk or other liquid food <i>isipuni esibaziwe sokuphaka amasi noma okunye ukudla okuwuketshezi</i>	<i>ngiphakele amasi ngokhezo olusha</i>
Meronym: <i>isibambo</i>	the part of an object that is designed to be held in order to use it <i>ingxenye yento enzelwe ukubanjwa ukuze isetshenziswe</i>	<i>isibambo sokhezo sephukile</i>
<i>isixembe</i>	a large wooden serving and stirring spoon shaped broad and flat in front <i>isipuni esikhulu esibaziwe sokuphaka noma sokugovuza esivulekile futhi saba yisibaca phambili</i>	<i>udadewethu ugovuza isijingi ngesixembe</i>
Meronym: <i>isibambo</i>	the part of an object that is designed to be held in order to use it <i>ingxenye yento enzelwe ukubanjwa ukuze isetshenziswe</i>	<i>isibambo sesixembe side</i>
<i>igula</i>	a calabash container for sour milk <i>isitsha soselwa sokufaka amasi</i>	<i>igula likagogo elidala liphukile</i>
<i>isigubhu</i>	a calabash container for water, beer or amahewu (fermented maize porridge) <i>isiqukathi soselwa sokufaka amanzi, utshwala noma amahewu</i>	<i>uthenge isigubhu emakethe yezitsha zesintu</i>
<i>iphaphasi</i>	a flat, wide, open-mouthed calabash container for food <i>isiqukathi soselwa esiyisicaba, esibanzi nesinomlomo ovulekile sokufaka ukudla</i>	<i>bamlethele iphaphasi eligcwele izinkobe</i>

<i>isichumo</i>	a spherical beer container with a narrow mouth made of lala palm leaves, and provided with a lid of the same material <i>isitsha sotshwala esisandingilizi esinomlomo omncane esakhiwe ngelala futhi esinesivalo esakhiwe ngelala</i>	<i>thathela umkhulu isichumo sakhe ekhishini</i>
Meronym: <i>isibo</i>	a lid or covering for a beer container made of the same material and which fits over the neck like a cap <i>isivalo noma isembozo sesitsha sotshwala esakhiwe ngelala futhi esingena khaxa esichumweni njengesivalo</i>	<i>vala isichumo ngesimbozo kungaze kungene impukane</i>
<i>isilulu</i>	a large container with a broad base and small opening at the top, plaited from grass for grain storage <i>isitsha esikhulu esivulekile ngaphansi bese sivuleka kancane phezulu, esilukwe ngotshani, sokugcina ukudla okuyizinhlamvu</i>	<i>isilulu esidala siyavuzo</i>
Meronym: <i>imvokoqo/ imvongoqo</i>	a large wooden serving spoon with a deep or hollow bowl <i>isipuni sokuphaka esikhulu esibaziwe esinesitsha esinomgodi oshonayo noma esinomgodi ovulekile</i>	<i>ubaba ubaza imvokoqo</i>
Meronym: <i>isibambo</i>	the part of an object that is designed to be held in order to use it <i>ingxenye yento enzelwe ukubanjwa ukuze isetshenziswe</i>	<i>isibambo semvokoqo side</i>
<i>isikhetho</i>	a spoon made of plaited palm leaves to skim beer before it is drunk <i>isipuni esakhiwe ngosundu sokuhlaza utshwala ngaphambi kokuba buphuzwe</i>	<i>ukhumbule ukuhlaza utshwala ngesikhetho ngaphambi kokuba uthelele abantu abadala</i>
Meronym: <i>isibambo</i>	the part of an object that is designed to be held in order to use it <i>ingxenye yento enzelwe ukubanjwa ukuze isetshenziswe</i>	<i>isibambo sesikhetho siyisicaba futhi sigobile</i>

Table 3: Proposed isiZulu wordnet entries for a representative sample of traditional domestic utensils

When added to WordnetLoom, the hierarchical classification of kitchen utensils such as traditional (cooking) pots and ladles is visually presented in the wordnet graphs as shown in Figure 4. Attributes of the particular synset are displayed in the sub-panels on the right of the graph.

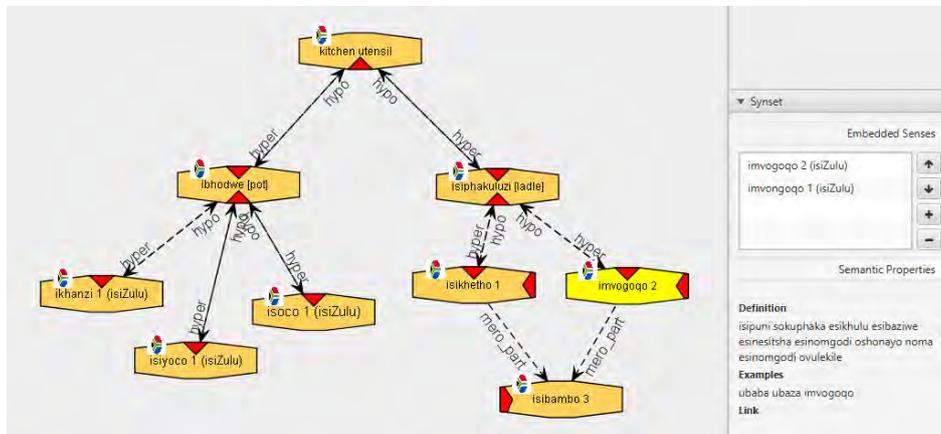


Figure 4: Kitchen utensils represented in WordnetLoom

The arrows in Figure 4 indicate the relevant semantic relations, namely that *ibhodwe* (pot) is in a hypernym relation to *ikhanzi*, *isiyoco* and *isoco* which are all types of cooking pots. *Imvongo* and *isikhetho* are types of ladles and are therefore hyponyms of *isiphakuluzi* (ladle). In the latter examples, a part-whole or meronymy relation exists with regard to *isibambo* (handle). A synonym of *imvongo* is indicated as *imvongo* in the sub-panel of attributes.

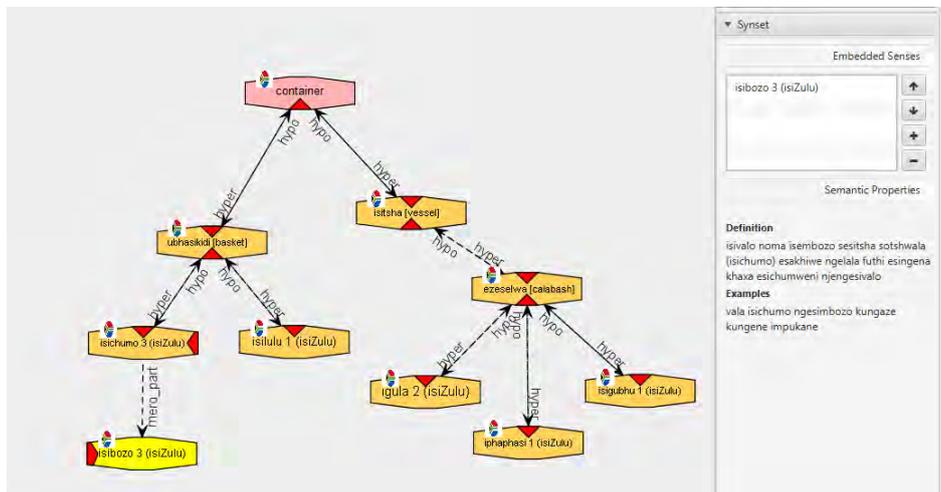


Figure 5: Containers represented in WordnetLoom

The semantic relations in the synsets in Figure 5 indicate for example that *ibhasikidi* (basket) is in a hypernym relation to *isichumo* and *isilulu*. In the case of *isichumo*, a part-whole or meronymy relation exists with *isibozo* (lid for beer container). With regard to the containers resorting under *isitsha* (vessel), it is clear that *igula*, *iphaphasi* and *isigubhu* are types of calabash vessels and are therefore hyponyms of *ezeselwa* (calabash). It is noteworthy that the lexicalized term of a calabash container is determined by the function of the vessel, that is, *igula* is a container for sour milk, *iphaphasi* is a container for water, beer or *amahewu* (fermented maize porridge drink) and *isigubhu* is a container for food.

5. Conclusion and future work

Our core contribution is twofold and ties in with envisioned future work:

Firstly, we describe a new lexical resource for isiZulu which enhances and supplements lexical knowledge from conventional dictionaries and other interdisciplinary sources, thereby addressing the problem of missing or dispersed indigenous knowledge that might be lost for posterity. The nature of the chosen lexical resource, that is the wordnet, furthermore lends itself perfectly to relating the IK to senses used more frequently in modern society, thereby bringing the traditional into the digital age. By focusing on filling lexical gaps between the African language isiZulu, and the English template as found in the PWN with culturally relevant synsets, the AfWN, in turn, becomes an even more useful resource for natural language processing. Digitisation of older manuscripts containing IK is gaining momentum and by including traditional senses in a modern resource, these texts can be automatically unlocked via information retrieval, machine translation and document categorisation. Regarding future work, we envisage expanding the AfWN into a multimodal database including images and sound clips with synsets. Declerck et al. (2020) describe a method by which to add speech data to the Open English Wordnet so that pronunciation information can be added to different senses in each synset. By doing so, they show improvements in disambiguation between different word meanings. This same method could, for instance, be used to add tonal information to the AfWN or, as an additional level, to preserve and document spoken IK such as stories or songs along with the related lexical entries.

Secondly, we lay the foundation for further discussion and development of a common scheme for storing lexical data not only for the South African Bantu languages, but for the Bantu language family as a whole. Schofield (1943), for instance gives a detailed description of the pottery types (with the relevant terminology) of the Nguni and Sotho language groups as well as Tshivenḁa, which could serve as a starting point. Furthermore, the style guides and protocols created for the development of the AfWN according to international best practice, can serve as example for other African languages.

Endnotes

1. See <https://www.sadilar.org>
2. Also available at <http://wordnetweb.princeton.edu/perl/webwn>
3. See a description of the core base concepts list at <http://globalwordnet.org/gwa-base-concepts/>
4. See <https://kamusi.org>
5. Original search results available at:
<http://wordnetweb.princeton.edu/perl/webwn?s=search>

Acknowledgements

The African Wordnet project (AfWN) was made possible with support from the South African Centre for Digital Language Resources (SADiLaR). SADiLaR (<https://www.sadilar.org>) is a research infrastructure established by the Department of Science and Innovation of the South African Government as part of the South African Research Infrastructure Roadmap (SARIR).

Contributions by Ms Lindelwa Mahonga and Ms Celimpilo Dladla, members of the isiZulu AfWN development team, are thankfully acknowledged, while the first author acknowledges the University of South Africa for granting her leave for research purposes.

References

A. Dictionaries

- ISN = Nyembezi, S. 1992. *Isichazamazwi Sanamuhla Nangomuso*. Pietermaritzburg: Reach Out.
- ISZ = Mbatha, M.O. (Ed.). 2006. *Isichazamazwi SesiZulu*. Pietermaritzburg: New Dawn.
- ZED = Doke, C.M. and B.W. Vilakazi. 1964. *Zulu-English Dictionary*. Second Edition. Johannesburg: Witwatersrand University Press.

B. On-line databases

- CBPV = Africa Museum. 2007. Comparative Bantu Pottery Vocabulary. Available:
https://www.africamuseum.be/en/research/discover/human_sciences/culture_society/cb-pottery/browse_zone
- PWN = Princeton University. 2020. WordNet — A Lexical Database for English. Available:
<https://wordnet.princeton.edu/>

C. Other literature

- Abubakar, A.I., A. Roko, A.B. Muhammad and I. Saidu. 2019. Hausa WordNet: An Electronic Lexical Resource. *Saudi Journal of Engineering and Technology* 4(8): 279-285.
- Anderson, W., L. Pretorius and A. Kotzé. 2010. Base Concepts in the African Languages Compared to Upper Ontologies and the WordNet Top Ontology. Calzolari, N. et al. (Eds.). 2010. *Pro-*

- ceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 17–23 May 2010, Valletta, Malta: 3757-3764. Paris: European Language Resources Association (ELRA).*
- Batibo, H.M.** 2016. The Origin and Evolution of Setswana Culture: A Linguistic Account. *Botswana Notes and Records* 48: 134-149.
- Bentivogli, L. and E. Pianta.** 2000. Looking for Lexical Gaps. Heid, U. et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, 8–12 August 2000: 8-12. Stuttgart: Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.*
- Bond, F., H. Isahara, K. Kanzaki and K. Uchimoto.** 2008. Boot-strapping a WordNet Using Multiple Existing WordNets. Calzolari, N. et al. (Eds.). 2008. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), 28–30 May 2008, Marrakech, Morocco: 1619-1624. Paris: European Language Resources Association (ELRA).*
- Bond, F. and K. Paik.** 2012. A Survey of Wordnets and their Licenses. Fellbaum, C. and P.T.J.M. Vossen (Eds.). 2012. *Proceedings of the Sixth Global WordNet Conference 2012 (GWC2012), 9–13 January 2012, Matsue, Japan: 64-71. Brno: Tribun EU.*
- Bosch, S.E. and M. Griesel.** 2017. Strategies for Building Wordnets for Under-resourced Languages: The Case of African Languages. *Literator* 38(1): 1351-1376. Available: <https://doi.org/10.4102/lit.v38i1.1351>
- Bosch, S.E. and M. Griesel.** 2019. Towards the Digital Documentation and Preservation of Indigenous Knowledge in the African Wordnet. Horak, A., K. Osolsobe, A. Rambousek and P. Rychly (Eds.). 2019. *Slavonic Natural Language Processing in the 21st Century: 17-27. Brno, Czech Republic: Tribun EU.*
- Calzolari, N.** 2018. Introduction to LREC 2018. Calzolari, N. et al. (Eds.). 2018. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 7–12 May 2018, Miyazaki, Japan: 1-6. Paris: European Language Resources Association (ELRA). Available: <https://www.aclweb.org/anthology/L18-1088.pdf>*
- Cosijn, E., A. Pirkola, T.J.D. Bothma and K. Järvelin.** 2002. Information Access in Indigenous Languages: A Case Study in Zulu. *South African Journal of Libraries and Information Science* 68(2): 94-103.
- Declerck, T., L. Bajcetic and M. Siegel.** 2020. Adding Pronunciation Information to Wordnets. Declerck, T. et al. (Eds.). 2020. *Proceedings of the Workshop on Multimodal Wordnets (MMWN-2020), 11–16 May 2020, Marseille, France: 39-44. Paris: European Language Resources Association (ELRA).*
- Eckart, T., S. Bosch, U. Quasthoff, E. Körner, D. Goldhahn and S. Kaleschke.** 2020. Usability and Accessibility of Bantu Language Dictionaries in the Digital Age: Mobile Access in an Open Environment. Mabuya, R. et al. (Eds.). 2020. *Proceedings of the First Workshop on Resources for African Indigenous Languages 2020 (RAIL2020), 16 May 2020, Marseilles, France: 9-14. Marseilles, France: European Language Resources Association (ELRA). Available: <https://www.aclweb.org/anthology/2020.rail-1.8/>*
- Fellbaum, C. (Ed.).** 1998. *Wordnet: An Electronic Lexical Database.* Cambridge, MA: MIT Press.
- Fellbaum, C.** 2014. Large-scale Lexicography in the Digital Age. *International Journal of Lexicography* 27(4): 378-395.
- Fowler, K.D.** 2015. Zulu Pottery Technology and Group Identity in the Phongolo Basin, South Africa. *Southern African Humanities* 27:81-111.

- Gouws, R., U. Heid, W. Schweickard and H. Wiegand (Eds.).** 2014. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, Boston: De Gruyter Mouton. Available: <https://doi.org/10.1515/9783110238136>
- Griesel, M. and S.E. Bosch.** 2020. Navigating Challenges of Multilingual Resource Development for Under-Resourced Languages: The Case of the African Wordnet Project. Mabuya, R. et al. (Eds.). 2020. *Proceedings of the First Workshop on Resources for African Indigenous Languages 2020 (RAIL2020), 16 May 2020, Marseilles, France*: 45-50. Marseilles, France: European Language Resources Association (ELRA). Available: <https://www.aclweb.org/anthology/2020.rail-1.8/>
- Griesel, M., S. Bosch and M.L. Mojapelo.** 2019. Thinking Globally, Acting Locally — Progress in the African Wordnet Project. Fellbaum, C. et al. (Eds.). 2019. *Proceedings of the Tenth Global WordNet Conference 2019 (GWC2019), 23–27 July 2019, Wrocław University of Science and Technology, Wrocław, Poland*: 191-196. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
- Grossert, J.W.** 1985. *Zulu Crafts*. Pietermaritzburg: Shuter & Shooter.
- Kotzé, G.** 2008. Ontwikkeling van 'n Afrikaanse woordnet: Metodologie en integrasie [Development of an Afrikaans wordnet: Methodology and Integration]. *Literator* 29(1): 163-184. Available: <http://dx.doi.org/10.4102/lit.v29i1.105>
- Krige, E.J.** 1965. *The Social System of the Zulus*. Pietermaritzburg: Shuter & Shooter.
- McCrae, J.P., E. Rudnicka and F. Bond.** 2020. English WordNet: A New Open-source Wordnet for English. *K Lexical News* 28: 37-44. Available: <https://kln.lexicala.com/kln28/mccrae-rudnicka-bond-english-wordnet/>
- Moors, C., I. Wilken, K. Calteaux and T. Gumede.** 2018. Human Language Technology Audit 2018: Analysing the Development Trends in Resource Availability in all South African Languages. *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists SAICSIT '18: Technology for Change, 26–28 September 2018, Port Elizabeth, South Africa*: 296-304. New York: The Association for Computing Machinery (ACM). Available: <https://dl.acm.org/doi/10.1145/3278681.3278716>
- Miller, G.A.** 1995. WordNet: A Lexical Database for English. *Communications of the ACM* 38(11): 39-41.
- Naskręć, T., A. Dziob, M. Piasecki, C. Saedi and A. Branco.** 2018. WordnetLoom — A Multilingual Wordnet Editing System Focused on Graph-based Presentation. Bond, F. et al. (Eds.). 2018. *Proceedings of Ninth Global WordNet Conference 2018 (GWC2018), 8–12 January 2018, Nanyang Technological University (NTU), Singapore*: 191-200. Singapore: Nanyang Technological University (NTU). Available: <http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/gwc-2018-proceedings.pdf>
- Neale, S.** 2018. A Survey on Automatically-Constructed WordNets and their Evaluation: Lexical and Word Embedding-based Approaches. Calzolari, N. et al. (Eds.). 2018. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 7–12 May 2018, Miyazaki, Japan*: 1705-1710. Paris: European Language Resources Association (ELRA). Available: <https://www.aclweb.org/anthology/L18-1270.pdf>
- Niles, I., and A. Pease.** 2001. Toward a Standard Upper Ontology. Guarino, N. et al. (Eds.). 2001. *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), 17–19 October 2001, Ogunquit, Maine, USA*: 2-9. New York: Association for Computing Machinery.
- Nyembezi, S. and O.E.H. Nxumalo.** 1989. *Inqolobane Yesizwe (The Storehouse of the Nation)*. Pietermaritzburg: Shuter & Shooter.

- Ordan, N. and S. Wintner.** 2007. Hebrew Wordnet: A Test Case of Aligning Lexical Databases across Languages. *International Journal of Translation* 19(1): 39-58.
- Ossai, N.G.** 2010. African Indigenous Knowledge Systems (AIKS). *Symbiosis* 7(2): 1-13.
- Schofield, J.F.** 1943. A Preliminary Study of the Pottery of the Bantu Tribes of the Union of South Africa. *South African Journal of Science* 39: 256-281.
- Snider, K. and J. Roberts.** 2006. *SIL Comparative African Wordlist (SILCAWL)*. Available: https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider_silewp2006-005.pdf
- SUMO.** 2002. *Suggested Upper Merged Ontology*. Available: <http://www.ontologyportal.org>
- Souza, R.R., A. Dorn, B. Piringer and E. Wandl-Vogt.** 2020. Identification of Indigenous Knowledge Concepts through Semantic Networks, Spelling Tools and Word Embeddings. Calzolari, N. et al. (Eds.). 2020. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), 11–16 May 2020, Marseille, France: 943-947*. Paris: European Language Resources Association (ELRA).
- Tufis, D., D. Cristea and S. Stamou.** 2004. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *Romanian Journal of Information Science and Technology* 7(1–2): 9-43.
- Vetulani, Z., M. Kubis and T. Obrębski.** 2010. PolNet — Polish WordNet: Data and Tools. Calzolari, N. et al. (Eds.). 2010. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 17–23 May 2010, Valletta, Malta: 3793-3797*. Paris: European Language Resources Association (ELRA).
- Vincze, V. and A. Almási.** 2014. Non-lexicalized Concepts in Wordnets: A Case Study of English and Hungarian. Orav, H., C. Fellbaum and P. Vossen (Eds.). 2014. *Proceedings of the 7th Global WordNet Conference 2014 (GWC2014), Demonstration Session, Tartu, Estonia, 25–29 January 2014: 118-126*. Tartu: Global WordNet Association.
- Vossen, P., F. Bond and J. McCrae.** 2016. Toward a Truly Multilingual GlobalWordnet Grid. Mititelu, V., C. Forăscu, C. Fellbaum and P. Vossen (Eds.). 2016. *Proceedings of the Eighth Global WordNet Conference 2016 (GWC2016), Bucharest, Romania, 27–30 January 2016: 419-426*. Bucharest: University of Iasi.

e-Dictionaries in a Network of Information Tools in the e-Environment

Theo J.D. Bothma, *Department of Information Science, University of
Pretoria, South Africa (theo.bothma@up.ac.za)*

and

Rufus H. Gouws, *Department of Afrikaans and Dutch, Stellenbosch
University, South Africa (rhg@sun.ac.za)*

Abstract: Traditional dictionaries offer curated data to users. Users should therefore be able to find the correct data to solve their information need. However, users don't necessarily know the exact scope of lexicographic information. Dictionary articles can still demand considerable interpretation by the user to select the appropriate meaning or equivalent.

In the e-environment, users can easily navigate between different e-sources. This is especially evident on various e-book platforms, where one can link multiple dictionaries and other sources to a text or search of the internet. Internet content is obviously not curated, and providing access to such data is therefore anathema to the traditional lexicographer. A traditional dictionary is the result of an application of data pushing procedures. The online environment enables the use of data pulling procedures that give users access to both curated and non-curated data.

These issues are illustrated by means of a number of examples that show that a large number of different and disparate information sources are easily available to the user to satisfy any specific information need, and that the dictionary is one of a plethora of information sources. The information is therefore available on demand, without risking information overload.

It is argued that, when optimising a network of information tools that constitutes a comprehensive search universe, the information retrieval structure should preferably have a circular network as application domain, rather than a linear continuum.

Keywords: CONTEXTUALIZATION, DATA DISTRIBUTION STRUCTURES, INFORMATION TOOLS, SEARCH VENUES, USER INFORMATION NEEDS, NETWORK OF INFORMATION TOOLS, DATA PULLING, DATA PUSHING, INFORMATION SOURCES, SMART E-DICTIONARIES

Opsomming: e-Woordeboeke in 'n netwerk van inligtingswerktuie in die e-omgewing. Tradisionele woordeboeke bied gekeurde data aan gebruikers. Daarom behoort gebruikers die korrekte data te kan vind om 'n inligtingsbehoefte te bevredig. Gebruikers weet egter nie altyd wat die presiese bestek van die leksikografiese inligting is nie. Woordeboekartikels stel steeds eise aan die gebruiker om die gepaste betekenis of ekwivalent te kies.

In die e-omgewing kan gebruikers maklik tussen verskillende e-bronne rondbeweeg. Dit geld veral op verskillende e-boekplatforms waar baie woordeboeke en ander bronne aan 'n teks of internetsoektoeg gekoppel kan word. Internetinhoud is vanselfsprekend nie gekeur nie en om toe-

gang tot sulke data te gee, is bykans 'n gruwel vir tradisionele leksikograwe. 'n Tradisionele woordeboek is die produk van 'n toepassing van datatrekprosedures. Die aanlyn omgewing maak die gebruik van datastootprosedures moontlik wat aan gebruikers toegang tot sowel gekeurde as ongekeurde data gee.

Hierdie kwessies word geïllustreer aan die hand van 'n aantal voorbeelde wat wys hoe talle verskillende en uiteenlopende inligtingsbronne maklik beskikbaar is vir gebruikers om enige spesifieke inligtingsbehoefte te bevredig en dat 'n woordeboek een van baie inligtingsbronne is. Die inligting is op aanvraag beskikbaar sonder die gevaar van inligtingsoorlading.

Daar word aangevoer dat wanneer 'n netwerk van inligtingswerktuie optimaal saamgestel word wat 'n omvattende soekuniversum bied, moet die inligtingsonttrekkingstruktuur verkieslik 'n sirkel-netwerk eerder as 'n lineêre kontinuum as toepassingsgebied hê.

Sleutelwoorde: DATA STOOT, DATA TREK, DATAVERSPREIDINGSTRUKTURE, GEBRUIKERSINLIGTINGSBEHOEFTE, INLIGTINGSBRONNE, INLIGTINGSWERKTUIE, KONTEKSTUALISERING, NETWERK VAN INLIGTINGSWERKTUIE, SLIM e-WOORDEBOEKE, SOEKPLEKKE

1. Introduction

Some dictionaries, especially printed dictionaries, can still be regarded as stand-alone reference products. The emergence and establishment of online lexicography did not only change the medium of dictionaries, their structures, contents and presentation of data. It has had a far-reaching influence on the nature and extent of lexicography as a field of research and on the position of dictionaries as reference sources. Successful utilisation of possibilities available in the online environment demanded an increase in interdisciplinary collaboration. In this regard the importance of mutual research efforts in the fields of lexicography and information science has been apparent for some time, as is evident from a number of articles in Fuertes-Olivera and Bergenholtz (2011) and Granger and Paquot (2012). More recently, the results of these joint efforts can be seen in various publications, e.g. Ball and Bothma (2018), Bergenholtz and Agerbo (2017), Bergenholtz, Bothma and Gouws (2015), Bothma (2018), Bothma and Bergenholtz (2013), Bothma, Gouws and Prinsloo (2016), Bothma, Prinsloo and Heid (2018), Bothma and Tarp (2014), Tarp and Gouws (2019).

One of the realities of online referencing is that dictionaries function along with numerous other reference sources in a network of information tools. This has been stated succinctly by Varantola already in 2002:

Dictionaries need not be regarded as stand-alone lexical tools that should provide all the answers that the users need about language in use. In my vision, the future dictionary is rather an integrated tool or a number of tools in a professional user's toolbox where it coexists with other language technology products such as encyclopedic sources of reference, different types of corpora, corpus analysis tools (Varantola 2002: 34-35).

Successfully accessing and using such sources pose definite challenges to users.

Lexicographers and metalexigraphers need to respond to these challenges in order to ensure the continued use of dictionaries. This paper looks at some of the issues relevant to lexicography when negotiating the position of smart e-dictionaries in a network of information tools.

2. Finding the desired item

Traditional dictionaries, both printed and online variants, typically offer curated data to users. Users should therefore be able to find the correct data to solve their information need. Lexicographers add a caveat — the information need should be a lexicographic information need¹. However, all users don't necessarily know the exact scope of lexicographic information presented in a given dictionary and whether this information will assist them to execute a successful dictionary consultation procedure. A further complication arises where words are not included in the dictionary. Even if a required lemma is found, the dictionary does not necessarily provide the user with only the correct information for the context or does not necessarily contain the required information to satisfy a specific need. To illustrate this: a user reads the phrase "like burrs in a rough cloak" and is unfamiliar with the word *burrs*. When consulting the first monolingual English dictionary available to that user, in this case *Dictionary.com*, the user finds the article that has the lemma *burr* as guiding element, as seen in Figure 1. The occurrence of *burr* as a noun is treated as a polysemous lexical item. Although none of the given paraphrases of meaning reflects the meaning of the word *burrs* with which the user was confronted, the user does not know that and might opt for an inappropriate paraphrase of meaning, e.g. that given as sense number two. This will lead to incorrect text reception.

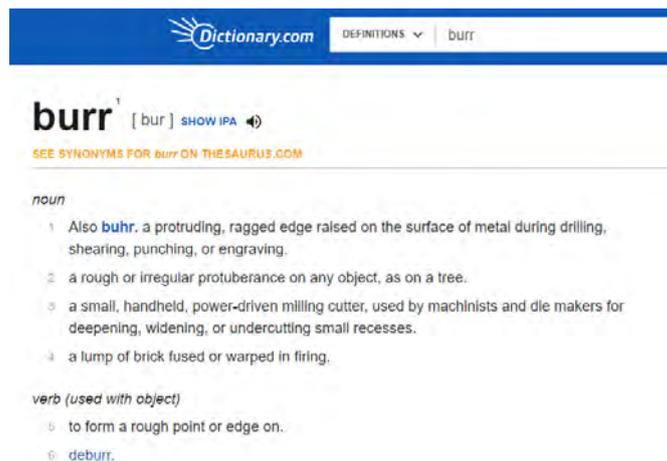


Figure 1: Partial article of the lemma sign *burr* from *Dictionary.com*

When a dictionary user has a text reception problem with the word *burrs* and consults a dictionary article with the required word as lemma it can still demand considerable interpretation by that user to realise that the items included in the article of the specific dictionary article do not provide an answer to the question that prompted the specific dictionary consultation. If the user is able to correctly interpret the information presented by the items in the dictionary article, they will realise that they have to consult another dictionary in order to find a solution to the text reception problem they have encountered. Such a second dictionary may also fail to provide the necessary assistance and even when a dictionary that does offer the required information is found, the user often still faces a real challenge to select the appropriate paraphrase of meaning. The complexity of such a search procedure is aggravated when the selected dictionary lacks sufficient cotextual items in the subcomments on semantics to ensure an optimal and unambiguous retrieval of information from the items given as paraphrases of meaning. This can be seen in Figure 2 where the user needs to read through the full dictionary article (or at least up to item 5) to come to the conclusion that the item giving the paraphrase of meaning for sense 5 of the lemma *burr* is the appropriate one:

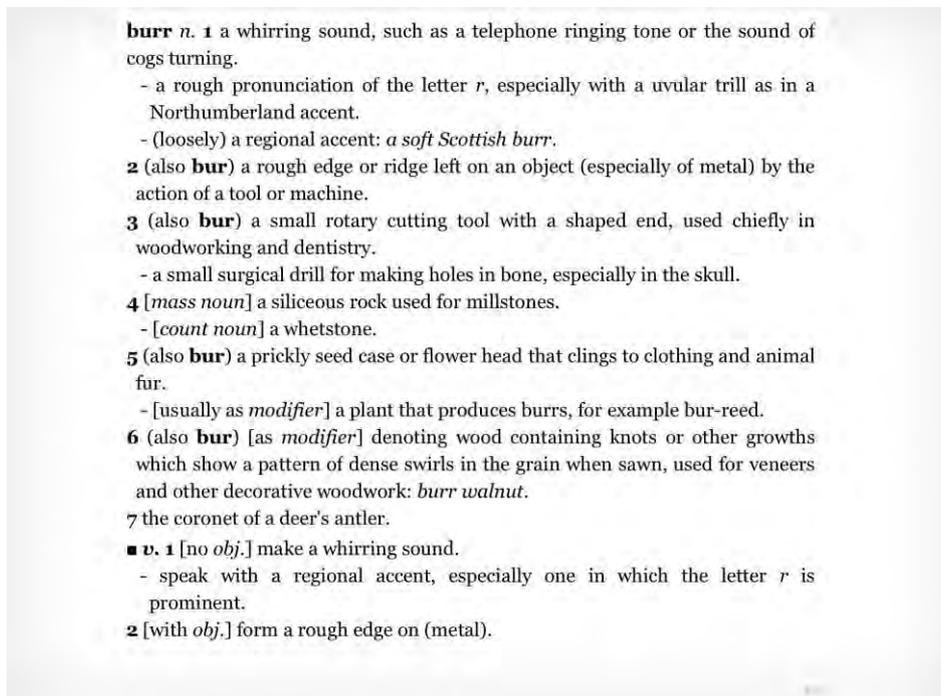


Figure 2: Partial article of the lemma sign *burr* in the *Oxford Dictionary of English* (linked Kindle edition)

Especially in a paper-based environment, users often have to consult multiple sources, which could be time intensive, and could often still lead to users' information needs not adequately satisfied.

The situation is different in the e-environment, where the user can navigate between sources — both e-dictionaries and other e-sources. This is especially evident on various e-book platforms. One can link multiple dictionaries to a text. By clicking on a word, the user has access to them but when that word is not a lemma in a specific dictionary the user can obtain misleading guidance, as in a case where a user needed an explanation for *Tamburlaine* as found in the phrase *I loved both Tamburlaines*. Clicking on that word in the text the user is directed to a lemma in the linked Kindle dictionary that closely resembles the form of the required word, i.e. the lemma *Tamerlane*, as seen in Figure 3, but such a dictionary could be completely misleading because the user has been directed to an incorrect interpretation of the word:

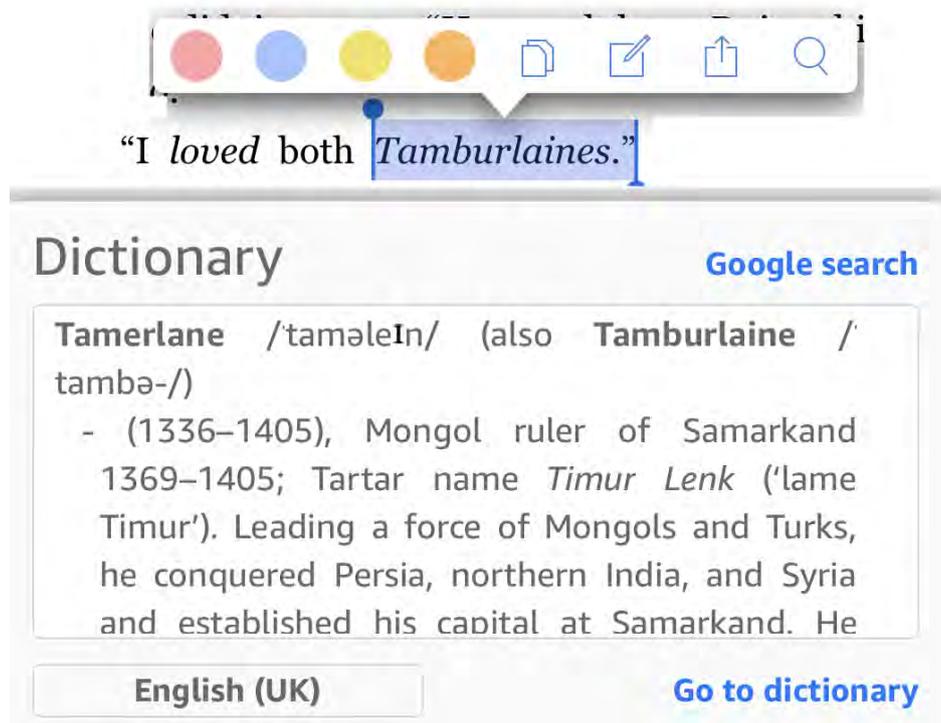


Figure 3: A word, *Tamburlaines*, in the Kindle app on an iPad, linked to the dictionary

Yet again, it is the responsibility of the user looking for help to realise that the consultation has not been successful, and the search has to continue. The user could also go directly to Google, to search the internet, resulting e.g. in a hit like the one presented in Figure 4:

Tamburlaine
Play by Christopher Marlowe

Book preview
Full book available

[READ NOW](#)

3,6/5 · [Goodreads](#)

82% liked this book
Google users

Tamburlaine the Great is a play in two parts by Christopher Marlowe. It is loosely based on the life of the Central Asian emperor, Timur. [Wikipedia](#)

First performance: 1587
Playwright: [Christopher Marlowe](#)
Characters: [Zenocrate](#), [Bajazeth](#), [Mycetes](#), [Cosroe](#), [Techelles](#),
[MORE](#)

Figure 4: Results of a search for *Tamburlaines* on Google

In the consultations of which figures 3 and 4 reflect the results, the specific dictionary is one of a range of information tools available to the user — but not necessarily the source offering the required assistance. It is expected from users to apply their minds in order to ensure a successful information retrieval. This could imply that the user has to move between different information tools constituting a network of reference sources, that is a representation with different reference sources occurring in different positions on this representation.

3. Moving between different information tools

When reading a text on a Kindle, a reader can be compelled to employ a variety of procedures in order to obtain the meaning of a word found in a text on the Kindle. If a user e.g. clicks on the compound *coney-catcher* in a text situated in 16th century London in the UK, the user gets access to the Kindle-linked dictionary and is taken to the lemma sign *coney* — as seen in Figure 5:



Figure 5: The word *coney* in the Kindle app on an iPad, linked to the dictionary

This dictionary does not include the compound *coney-catcher* and the user has to negotiate the treatment allocated to the simplex *coney* in order to try and solve the text reception problem. According to the Kindle-linked dictionary the lemma *coney* has two polysemous senses, referring to a "rabbit" or a "small grouper (fish) found on the coasts of the tropical western Atlantic." Within the

text in which the compound was found, situated in London, the user can rightfully deduce that the second sense does not apply. The question remains whether the first sense is the appropriate one. If it is the appropriate sense the reader therefore

- has to surmise that the compound would most probably not be the literal sense
- has to find the figurative sense, which fits perfectly to the context.

The cotextual environment of *coney-catcher* in the source text combines this compound with the lexical items *pick-pockets* and *masterless men*. How does a literal *rabbit catcher* fit into this environment? The user has to consult other sources and a Google search leads the user to various possibilities, e.g. illustrations like that given in Figure 6:

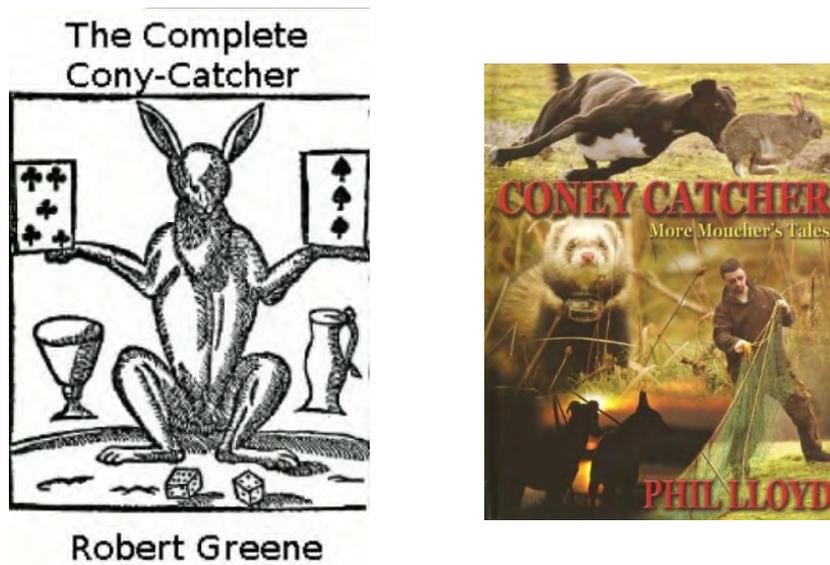


Figure 6: Coney-catcher

There are also occurrences of this word in numerous texts. From these occurrences the user might perhaps be able to deduce that a coney-catcher is a kind of swindler but it is not given. Fortunately for the traditional dictionary user, the Google search also offers reference works like *Lexico*, where the article of the lemma *coney-catcher* can be accessed, and *Wikipedia*, where the article of the lemma *coney-catching* can be accessed, as seen in Figure 7 and Figure 8 respectively:



Figure 7: Partial article of the lemma sign *coney-catcher* in *Lexico*



Figure 8: Partial article of the lemma sign *coney-catching* in *Wikipedia*

The treatment in these articles shows that a *coney-catcher* is a thief, a con man, a swindler or a trickster. This can also be confirmed by the treatment found in the *Oxford English Dictionary* (<https://www.oed.com/view/Entry/40917>).

Finding content in a source that is not a traditional dictionary leads to another challenge for the user. Such content is not curated, and providing access to such data is therefore anathema to the traditional lexicographer, because they have no control over the data presented to the user. Venturing into domains where both curated and non-curated content can be found, demands a different approach to the information quest by the user, i.e. to complement the search results provided by a traditional data pushing model with that of a data pulling model.

4. Data pushing and pulling

Research in various fields, including information science and lexicography, has been directed at aspects of data pulling and data pushing, cf. among other Duan, Gopalan and Dong (2005), Deolasee et al. (2010), Müller-Spitzer (2013) and Gouws (2018). Duan, Gopalan and Dong (2005: 2-3) formulate the difference between a push and a pull medium as follows:

In the sender-push model, the sender knows the identity of a receiver in advance and pushes the message in an asynchronous manner to the receiver. The receiver accepts the entire message, may choose to optionally examine the message, and then accept or discard it. An important aspect of the sender-push model is that the entire message is received before any receiver-side processing is performed.

In the receiver-pull model it is the receiver who initiates the message transfer by explicitly contacting the sender. The sender passively waits for the receiver and delivers the entire content upon receiving a request. Since it is the receiver who initiates the message transfer, the receiver would have explicit greater control over the message transfer and implicit greater trust in the received content, than in the sender-push model.

With regard to this distinction in the field of lexicography Müller-Spitzer (2013: 369) argues:

Generally, the Internet is considered to be a 'pull-medium' rather than a 'push-medium' like television, radio, or books ... Therefore, users are both sender and receiver. They are active in 'pulling' data from the website, saving relevant parts, etc. Thus, the Internet provides a very new form of communication in general.

And further:

It is communication in an innovative combination with new media.... This general property of the Internet as communication medium obviously has consequences for the property of online dictionaries as one type of text on the Internet. The process of pulling and, thus, representing lexicographic data according to a

user request is essential for EDs and must be considered when the textual structures of EDs are being looked at. (ED = electronic dictionary)

Both pushing and pulling approaches are relevant in lexicography. The need for the traditional dictionary still remains, but dictionaries do not necessarily have to be the final destination of an information search. Lexicographers need to plan and structure their e-dictionaries in such a way that they could be either a final destination or a transit area for information seekers. With dictionaries as a point of departure users should be able to obtain data from other sources. This will demand the use of both pushing and pulling approaches. For lexicographic applications of a pulling approach, Gouws (2018: 13) has introduced the notion of a lexicographic data pulling structure. He defines it as follows:

A lexicographic data pulling structure can be regarded as a structure consisting of a number of ordered elements that establishes the steps a dictionary user can follow in order to access from a given position in an existing online dictionary (where this position could be an item or search zone in a dictionary article or an article-external position, e.g. an entry in an outer feature) to dictionary-external sources from which the user can retrieve information to satisfy a specific lexicographic need.

Where traditional printed and online dictionaries result from the application of data pushing procedures, the online environment in addition enables the use of data pulling procedures that give users access to both curated and non-curated data. Deolasee et al. (2010) discuss a different kind of environment but their argument in favour of the need for combining push and pull to disseminate dynamic data also applies when utilising a network of information tools. Although a dictionary can still be used as an isolated stand-alone reference source, it should also be seen as integrated into a range of other sources where it populates a position in a network of sources. Yet again statements by Deolasee et al. (2010), albeit with regard to a different environment, also apply to this environment when they say:

A pull-based approach does not offer high fidelity ... A push-based algorithm can offer high fidelity ...

The occurrence of both dictionaries and other reference sources in a network of information tools demands that users have to be educated to distinguish between different data acquiring approaches and to be fully aware of the risks and opportunities when accessing non-curated data.

5. Reaching data

Regarding a dictionary as one of the sources in a network of information tools implies that users looking for information do not necessarily have to consult the dictionary, but they could also go directly to one or more of the other

sources on this network, albeit that these sources may contain non-curated data. When it comes to the use of dictionaries the approach in this paper supports the idea of a dictionary as a source in the information network that either supplies the user with the required information or allows access to other reference sources in a network where information retrieval results from either data pushing or data pulling procedures. This arrival-and-departure-halls status of a dictionary will not be discussed any further in this paper. The focus will rather be on various aspects of the network of information tools and the kind of reference skills and assessment the users need to ensure an optimal retrieval of information.

When employing data pushing and pulling procedures to access data in any of the sources on the information network, various methods could be used. These include direct access, moving through a portal, data on demand or bi-directional procedures between different sources. Figure 9 shows methods of accessing data — either in a dictionary (curated data) or on the web (non-curated data) from an Android mobile phone portal, whereas an iPad portal enables the result seen in Figure 10:

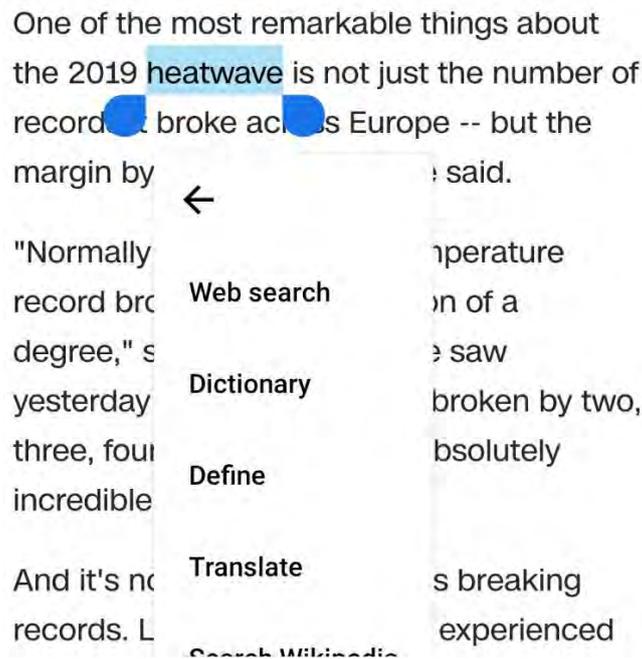


Figure 9: Pop-up portal on an Android mobile phone when the word *heatwave* is selected

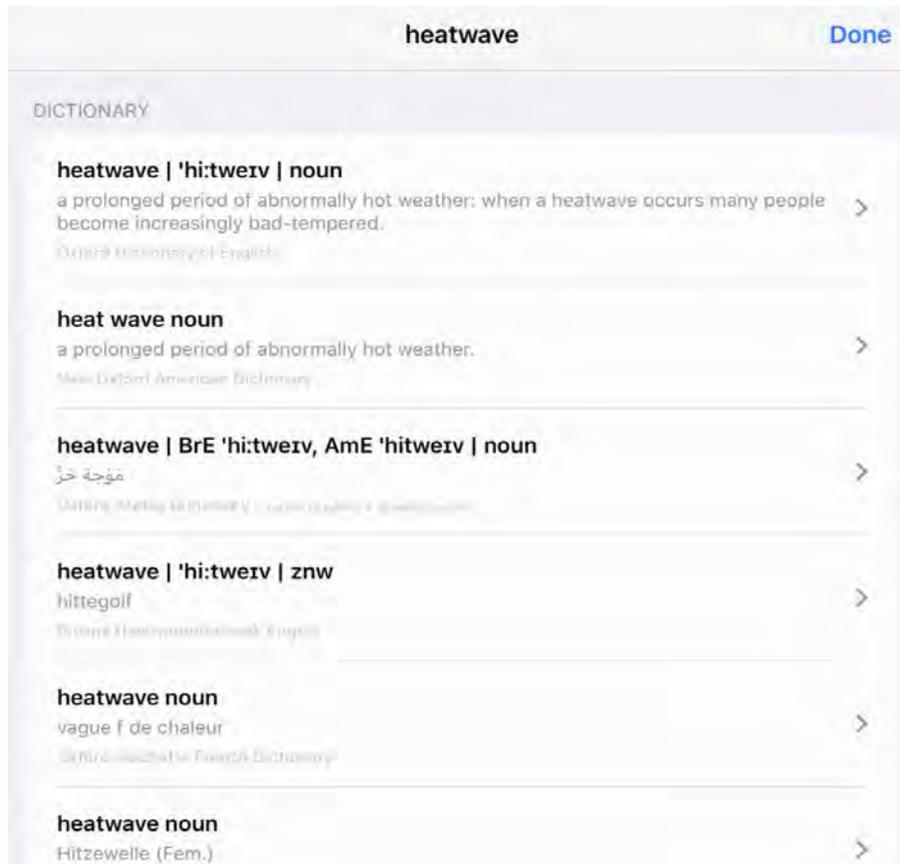


Figure 10: Pop-up portal of dictionaries on an iPad when the word *heatwave* is selected in the same text as in Figure 9

Figures 11 and 12 illustrate the information sources for searches performed in a Kindle general portal and a Kindle dictionary portal respectively.

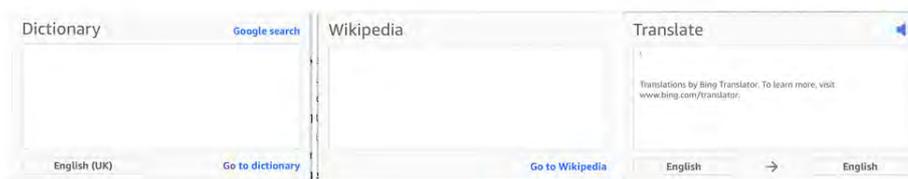


Figure 11: General pop-up portal on a Kindle, which provides access to a dictionary, Google, Wikipedia and a translation tool

In Figure 11, intentionally no content is displayed in the different results blocks.

For the translation tool, the system usually recognises the foreign language of the text, and then provides a translation. Both the source and target languages can be modified to any of the languages supported by the tool; also see Figure 17. English (UK) is the dictionary selected in Figure 11; this can be changed, as illustrated in Figure 12.

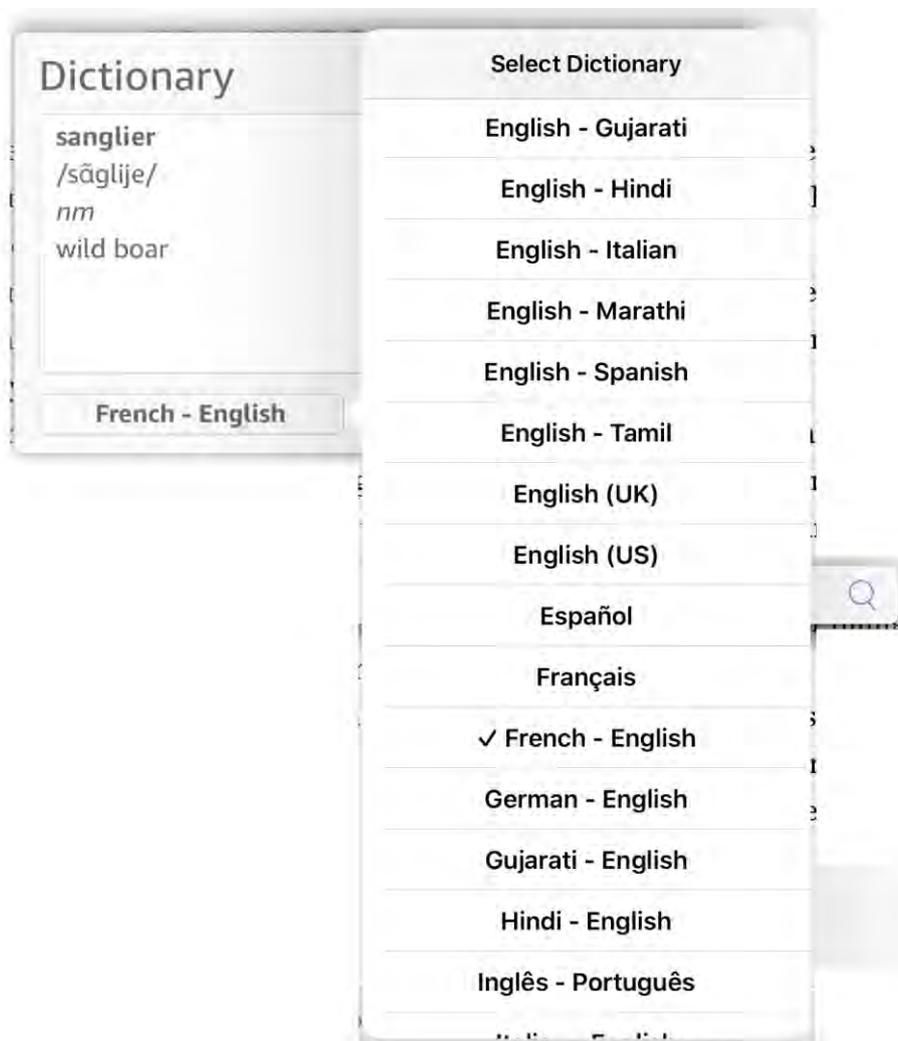


Figure 12: Kindle dictionary portal — the word *sanglier* was selected in the text (not visible on the screen capture) and *inter alia* the listed dictionaries are available

In Figure 12, a foreign language word was selected in the Kindle text. The system recognised *sanglier* as French, and used the French-English translation dictionary to provide the translation equivalent, *wild boar*. The dictionary selection can be changed to any of the free dictionaries available on the Kindle; the dictionaries need to be downloaded for use. Monolingual as well as bilingual translation dictionaries are available. A selection of these dictionaries is listed in Figure 12.

Directly accessing a dictionary article ensures curated data even though the user is not always sure whether this article delivers the solution to satisfy the lexicographic need; cf. also Tarp and Gouws (2019). The search ending in the article presented in Figure 13 might have found a satisfactory solution.



Figure 13: The word *dream* in the Kindle app on an iPad, linked to the dictionary

Figure 14 presents an article with a homonym of the required word as guiding element.



Figure 14: The word *tear* in the Kindle app on an iPad, linked to the dictionary

The homonym marker will make the knowledgeable user aware of the fact that they might have to move on to a subsequent article that also has the word *tear* as lemma. The user not equipped with these dictionary-using skills will not obtain the required assistance, although they have consulted a source that contains curated data. It is important that users should be familiar with the system of presentation in the given dictionary and be capable of assessing the information on offer.

Figure 15 shows that a search in a dictionary and *Wikipedia* for the meaning of the word *rowan* presents the user with appropriate help seeing that the word is treated as a monosemous lexical item.

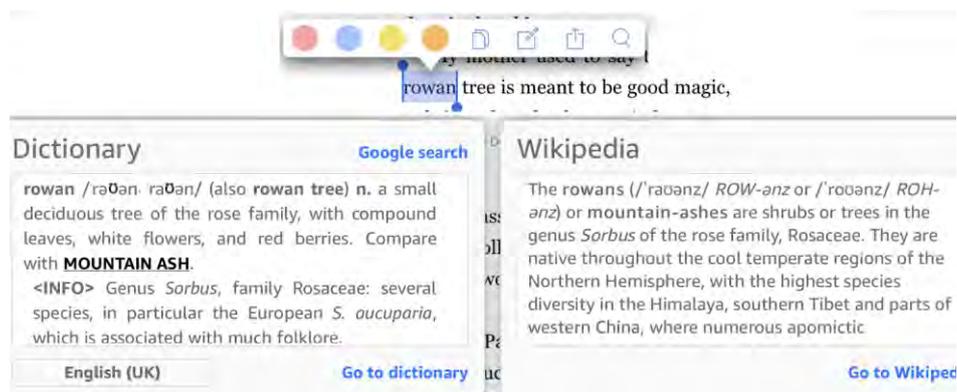


Figure 15: Clicking on *rowan* in the Kindle text provides information from the dictionary and *Wikipedia*

From the text and the two sources shown in the blocks in Figure 16, it is evident that *rowan* refers to a type of tree. The reader may want to see images of the tree, clicks on the Google link, and once the results are displayed, clicks on Images in the navigation bar. The first image is obviously incorrect, but the other images provide data that enhance the textual descriptions.

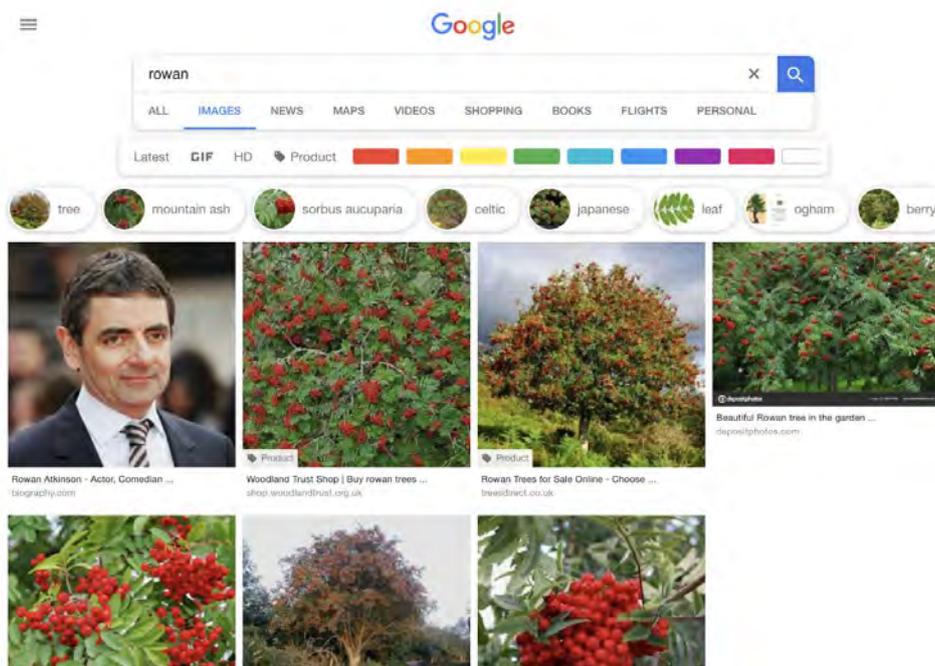


Figure 16: Image results of a search for *rowan* on Google

The value of the non-curated data on offer in some of the information tools should not be underestimated, especially if searches in these sources result from utilising data approaches — either directly or by means of a data pulling structure of a dictionary. Once the paraphrases of meaning presented in Figure 15 have convinced the user that a rowan is a tree or shrub the Google search could enhance the reference endeavour because the images give additional text reception guidance and increase the satisfaction of a cognitive need. For this success to be achieved users need to be able to apply their minds and assess the data on offer.

6. Making choices

In an online environment, users can easily move between a large number of different and disparate information sources to satisfy any specific information need. A dictionary is one of a plethora of information sources. Users have to make informed choices, based on context, prior and general world knowledge. Everywhere an information need exists, they have to make such choices. They have the option to ignore the information need by not consulting any available sources, or to explore the concept in more detail, either by consulting a dictionary article, or by delving deeper into other information sources with the dictionary as point of departure or with a dictionary-external point of departure. The information is therefore available on demand, without risking information overload.

A reader of a text on Kindle needs to obtain information regarding the form *deux chevaux*. Figure 17 shows that the linked dictionary does not provide any help but when the users opt for the translation tool, they do get some assistance. But does it solve the problem that prompted the consultation?



Figure 17: The Kindle app offers a translation of the selected words which do not occur in the linked English dictionary

To establish the meaning of *deux chevaux* in this specific context, the user needs to consult other information tools, e.g. by performing a Google search; cf. Figure 18.

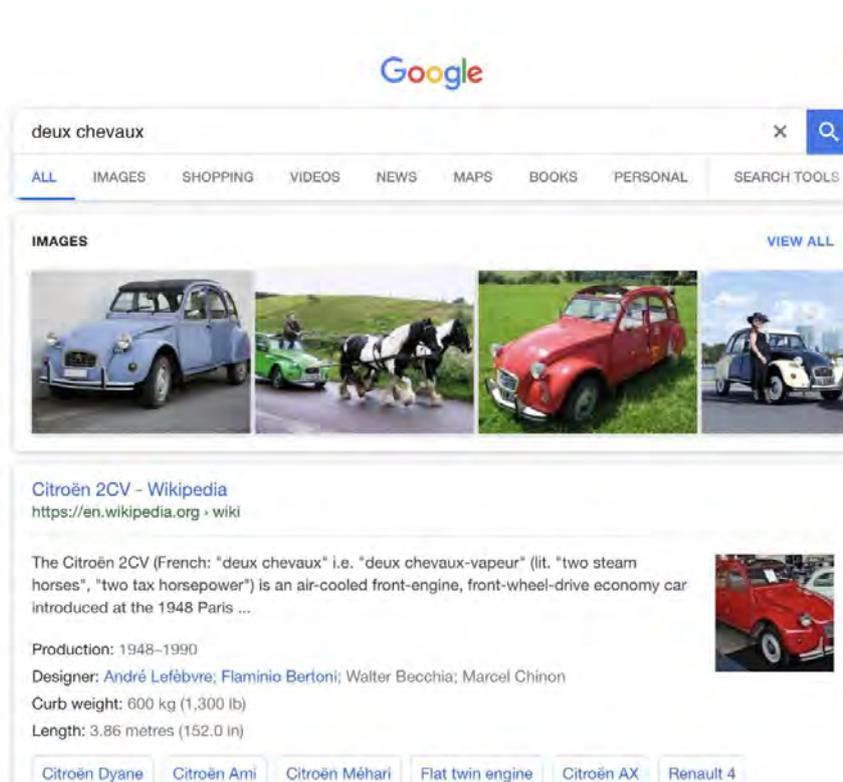


Figure 18: Images of the Citroën 2CV, commonly known as a *deux chevaux*

From the image of the Citroën in Figure 18 the required solution can be found but, yet again, the reader has to interpret the results in context.

Figures 19–21 show the additional information that can be obtained when moving between different information tools in order to find as much assistance as possible regarding a given expression.



Figure 19: The linked dictionary provides no help, but *Wikipedia* provides the required information

Neither the dictionary nor *Wikipedia* provide solutions for *Theses Martinianae*. However, a Google search provides the required information (Figure 20), with an option to drill down to the annotated full text of the book (Figure 21).

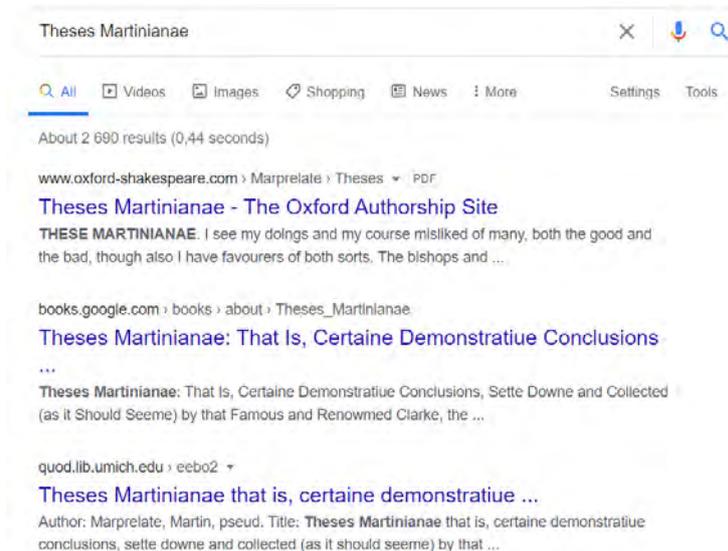


Figure 20: The linked dictionary provides no help for *Theses Martinianae*, but a search on Google provides a number of relevant links (the first few of which are shown here)

MARPRELATE TRACTS: THESES MARTINIANAE

1

THESES¹ MARTINIANAE²

That is,

Certain demonstrative³ conclusions set down and collected⁴ (as it should seem) by that famous and renowned clerk,⁵ the reverend Martin Marprelate the Great,⁶ serving as a manifest and sufficient confutation of all that ever the College of Catercaps - with their whole band of clergy priests - have, or can, bring for the defence of their ambitious and antichristian prelacy.⁷

Published and set forth as an afterbirth⁸ of the noble gentleman himself by a pretty⁹ stripling¹⁰ of his, Martin Junior,¹¹ and dedicated by him to his good neame¹² and nuncka,¹³ Master John Kankerbury.¹⁴

How the young man came by them, the reader shall understand sufficiently in the epilogue. In the meantime, whosoever can bring me acquainted with my father, I'll be bound he shall not lose his labour.¹⁵

¹ Thesis: In logic and rhetoric, a proposition laid down or stated, especially as a theme to be discussed and proved, or to be maintained against attack (in logic, sometimes as distinct from hypothesis; in rhetoric from antithesis). (OED, p.2284)

² Martin perhaps intended an allusion to Luther's 95 Theses, particularly since he calls himself 'Martin'.

³ Demonstrative: Provable by demonstration. (OED, p.517)

⁴ Although the *Theses* were printed as the fifth of the Marprelate tracts, Martin had already dealt with a number of the theses in *Hay Any Work For Cooper*, and it is possible that the entire 110 theses were 'collected' by Martin before any of the tracts were printed.

⁵ Clerk: A man of book learning, one able to read and writer, a scholar. (OED, p.348)

Figure 21: Following one of the links of Figure 20 provides an annotated full-text version of the *Theses Martinianae*

In the network of information tools, information is available on demand, and the reader can drill down to more information, if required. Information overload is therefore not a problem, as discussed in the next section.

7. Information on demand

7.1 Digital natives

Members of digital societies have become used to different possibilities to obtain information on demand. Their demand often exceeds the extent of information contained in a single source. Lexicographers need to take cognizance of this and also of the fact that a growing number of their potential target users are digital natives of which many belong to Generation Z. This generation has totally different experiences with and expectations of reference sources. They are net citizens, netizens, also known as the "Silent Generation" because of the time they spend online; cf. Parker (2013), Finch (2015). This has implications for the way in which data should be made accessible in the online environment. When looking for solutions to their problems members of the digital society seek immediate accessibility and they require uninterrupted connectivity; cf. Gouws (2017). They need to be in a position where they can pull data from a

variety of sources on the web whenever they need it and they need to be able to move from one source to the next if their need is not satisfied completely. To assist these users, guidance regarding the information on offer in a network of information tools should be accessible to them. This need has implications for the traditional notion of a data distribution structure of dictionaries.

7.2 Data distribution structures and data identification in a search universe

To ensure the best possible presentation of data in dictionaries and an optimal access to this data metalexigraphers have been concerned with an optimising of the data distribution structure of dictionaries, cf. Bergholtz, Tarp and Wiegand (1999), Gouws (2018, 2018a). When looking at a dictionary as one source in a network of information tools lexicographers need to be aware of the data allocated to the other sources. Data distribution structures are traditionally employed in dictionaries to regulate the distribution of data in the different search venues within a single dictionary. Wiegand, Beer and Gouws (2013: 63) identify these search venues as the search field (i.e. the central list of a dictionary), the search area (i.e. each individual dictionary article) and the search zone (i.e. each slot in a dictionary article that accommodates one or more items). Gouws (2018; to appear) argues in favour of comprehensive data distribution structures, characterised by an expansion of search venues to include a search region, i.e. all the components of the textual book structure (cf. Hausmann and Wiegand 1989: 330), which will also include all components of the frame structure of a dictionary; cf. Kammerer and Wiegand (1998). For online dictionaries Gouws (to appear a) further increases the application area of the data distribution structure by identifying a search universe which covers dictionary-external sources like other dictionaries in the same dictionary portal but also other sources, including the internet. This implies that the lexicographer may put some data in dictionary-external venues, e.g. a corpus or other sources on the internet. A dictionary user venturing into a search universe from a position within a specific dictionary as point of departure, may be in a position where the retrieval of information could result from the application of both data-pushing and data-pulling procedures.

For the planning and compilation of dictionaries employed as sources in a network of information tools a comprehensive data distribution structure and extended search venues are needed. However, although a comprehensive data distribution structure can allocate data to many sources, a search universe will typically also include data of which the lexicographer of any dictionary in a network of information tools is not aware. This implies that a dictionary user accessing the search universe that constitutes the application domain of the comprehensive data distribution structure of that dictionary is likely to encounter reference sources with which they have not been familiar prior to the search. Although these data occur in the search universe these data are not positioned there by the data distribution structure of any dictionary. Where a

dictionary user may use the data distribution structure of the dictionary to identify all the available search venues and data types linked to that dictionary, a search universe allows access to additional information. Where a user executes a search for information not allocated to its position in the search universe by the data distribution structure of a dictionary occurring in a network of information tools, such a user needs guidance with regard to the available information possibilities. Knowledgeable users of such a network should be aware of what they can find where. Within a search universe of a given network of information tools it would be optimal to have a menu at the disposal of the users of the information tools to help these users with the identification of the data available in that search universe.

7.3 Satisfying information needs

Providing assistance to users and answers to relevant questions is at the heart of putting any information tool to use. In a search universe, users will have access to different information tools in a network and they will have different options to ensure that their information needs can be satisfied. To satisfy their information needs users will have the option to access additional support tools, drill-down to more detail, filter information or link to additional sources. Expanding their access to different tools does not only serve to satisfy the original information need but it often leads to the retrieval of additional information that can add the satisfaction of a cognitive function to the search. Because the additional information is available on demand and users are not automatically immersed into such an overwhelming information pool the user is never exposed to the danger of an overload of information.

Within such a search universe an e-dictionary becomes an information tool amongst a plethora of information tools. Varantola (2002: 38) describes this as a "modular network with seamless connections between the modules". In practice, this network allows a portal integration that allows movement from dictionary to dictionary, from dictionary to other sources and, very important, from other sources to a dictionary. In a modular network of information tools that constitutes a mutual search universe, a given dictionary is not only a source from which users move on to other sources. It can be a first stop, a transit stop, but also a non-initial and final stop for users executing search procedures to satisfy their information needs. In this regard an e-dictionary should also be a source that can be targeted by users of any other tool in the network to provide additional information on demand.

As already indicated, the demand for additional information to satisfy a specific need may lead to the satisfaction of needs not previously expressed. When obtaining information to satisfy a text reception need the extent of the accessed information may be of such a nature that it satisfies an existing albeit not previously expressed cognitive need. In this regard a second phase cognitive function is satisfied by the specific information search.

To ensure an optimal use of a network of information tools the search universe needs to be planned in such way that a user can access any tool from any other tool in the network. It should not only be mono- or even bi-directional access but rather multi-directional access.

A linear continuum presupposes a linear movement between information sources:

- From A to B to C
- From Z to Y to X

This implies that any single source can only be directly linked with its linear neighbouring sources. As one of the tools on a linear continuum a given dictionary would not allow its user an optimal access to the search universe but only to two sources. This would seriously hamper the application of the data-pulling structure of such a dictionary and would neutralise any quest for information on demand. A linear continuum is therefore not an efficient way of visualising such tools.

When visualising the network of information tools that constitutes a comprehensive search universe the information retrieval structure should preferably have a circular representation as application domain, as seen in Figure 22:

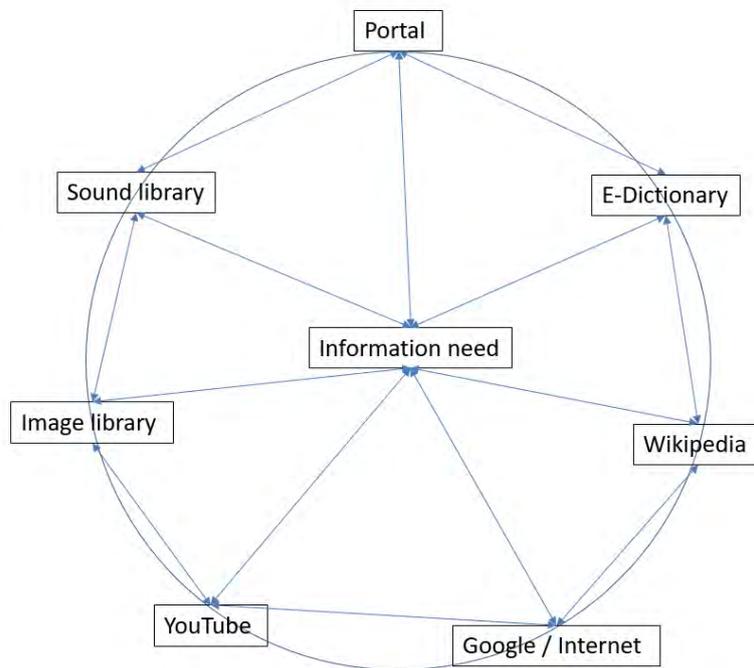


Figure 22: A modular network

Such a modular network puts the needs of the user at the centre, and the user equidistant from all information tools. From this hub the user can move along any spoke to a specific source. Users can also move between any two sources although the complexity of all possible linkages has not been indicated in Figure 22. Moving between sources can either be done directly or via the hub where the specific need is again established. The network includes a slot for e-dictionaries, some sources (like *Wikipedia*) that are not dictionaries but are "... utility tools with formal properties of a lexicographic nature" (Wiegand 2013: 285). Like dictionaries these sources could also contain curated data. This network can also include devices for user support, e.g. writing assistants. In the network one also finds sources of a non-lexicographic nature, including sources that contain non-curated data. The information need of a user will determine the source or sources they consult. Such a network is an ideal search universe for an information-on-demand approach where a user can move between sources.

When retrieving information from such a collection of information tools both pushing and pulling methods can be employed and the user should know what to use when and where. This demands not only dictionary-using skills but also a more general set of information literacy skills. These skills fall within a new kind of literacy that needs to be required, i.e. dictionary and information literacy, that will take user requirements into account.

8. In conclusion

A modular network of information tools in the e-environment has implications for lexicography. However, such a network and its different information tools should not only be discussed and analysed from a lexicographic perspective. Related fields like information science also need to contribute to a comprehensive discussion of these networks, the tools allocated to such a network, the information types populating these tools, the potential target users of one or more of these tools and the needs and information literacy skills of these users. This still needs to be done.

From a lexicographic perspective, it is clear that the dynamics of the e-environment compels metalexigraphers to carefully consider the nature and extent of the relevant lexicographic processes. It is also important to realise that lexicographers need to collaborate with experts from other fields and that the traditional notion of a dictionary culture has to be substituted by a more general information literacy culture which includes dictionary literacy.

A new assessment of users and their needs is necessary. Within the current model of dictionary research, research into dictionary use is regarded as one of the four central sections; cf. Wiegand (1998: 114). This component will have to be expanded to account for dictionary users having to use other information tools functioning along with a given dictionary in the search universe of

a network of information tools. This should also lead to the emergence of a higher level of both dictionary literacy and information literacy.

A modular network of information tools requires the user to apply their mind in all cases. This is evident from the examples discussed earlier: the user often has multiple opportunities to access or select information based on a single choice or consecutive choices they have made. An incorrect choice of meaning from the choices at hand will lead to an incorrect solution to the information need of the user, even in the case of curated information. The set of tools is not aware of the context of the information need, and only the user is aware of the context. Such a network of information tools is therefore not intelligent or "smart", but requires the user to analyse the context and make an informed choice. To be truly "smart", any set of tools should be able to provide the user with only the information that is required, as already formulated by Haas in 1962 in the context of dictionaries: "The perfect dictionary is one in which you can find the thing you are looking for preferably in the very first place you look" (Haas, 1962: 48). This level of sophisticated retrieval is currently not feasible and requires further technological developments, such as context-aware search engines and the implementation of artificial intelligence, neural networks and machine learning. Collaborative inter- and multi-disciplinary research is required to develop such systems. As Tarp and Gouws formulated this in their 2019 article: "Modern-day lexicographers are in a position to make some of the unfulfilled dreams of the past a reality. The challenge of the future is to make the impossible possible. We have work to do" (Tarp and Gouws 2019: 266). Only then will genuinely smart dictionaries and smart information tools become available.

Endnote

1. Where the term *information* is used in this paper with regard to dictionaries it covers the meaning of both *data* (= what is put into the dictionary by the lexicographer) and *information* (= what the user retrieves from a dictionary).

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference numbers (UID) 85434 and 95925). The grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

References

- Ball, Liezl H. and Theo J.D. Bothma.** 2018. Establishing Evaluation Criteria for e-Dictionaries. *Library Hi Tech* 36(1): 152-166.
- Bergenholtz, Henning and Heidi Agerbo.** 2017. Types of Lexicographical Information Needs and their Relevance for Information Science. *Journal of Information Science Theory and Practice* 5(3): 6-16.
- Bergenholtz, Henning, Theo J.D. Bothma and Rufus H. Gouws.** 2015. Phases and Steps in the Access to Data in Information Tools. *Lexikos* 25: 1-30.
- Bergenholtz, Henning, Sven Tarp and Herbert Ernst Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, Lothar et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2: 1762-1832.* Berlin: De Gruyter.
- Bothma, Theo J.D.** 2018. Lexicography and Information Science. Fuertes-Olivera, P.A. (Ed.). 2018: 197-216.
- Bothma, Theo J.D. and Henning Bergenholtz.** 2013. Information Needs Changing over Time: A Critical Discussion. *South African Journal of Libraries and Information Science* 79(1): 22-34.
- Bothma, Theo J.D., Rufus H. Gouws and Danie J. Prinsloo.** 2016. The Role of e-Lexicography in the Confirmation of Lexicography as an Independent and Multidisciplinary Field. Margalitadze, T. and G. Meladze. 2016. *Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity, Tbilisi, Georgia, 6-10 September, 2016:* 109-116. Tbilisi: Ivane Javakhishvili Tbilisi University Press. http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_008_p109.pdf (Last accessed on 20 April 2020).
- Bothma, Theo J.D., Danie J. Prinsloo and Ulrich Heid.** 2018. A Taxonomy of User Guidance Devices for e-Lexicography. *Lexicographica* 33: 391-422.
- Bothma, Theo J.D. and Sven Tarp.** 2014. Why Relevance Theory is Relevant for Lexicographers. *Lexicographica* 30: 350-378.
- Deolasee, Pavan, Amol Katkar, Ankur Panchbudhe, Krithi Ramamritham and Prashant Shenoy.** 2010. *Adaptive Push-Pull: Disseminating Dynamic Web Data.* <http://www-ccs.cs.umass.edu/~krithi/web/WWW10/www10/> (Last accessed on 2 April 2020).
- Duan, Zhenhai, Kartik Gopalan and Yingfei Dong.** 2005. *Push vs. Pull: Implications of Protocol Design on Controlling Unwanted Traffic.* <https://pdfs.semanticscholar.org/6f63/d37b4f8dd655e3594185e74daf4689f55aa1.pdf> (Last accessed on 2 April 2020).
- Finch, Jeremy.** 2015. *What Is Generation Z, And What Does It Want?* <http://www.fastcoexist.com/3045317/what-is-generation-z-and-what-does-it-want> (Last accessed on 28 May 2015).
- Fuertes-Olivera, Pedro A. (Ed.).** 2018. *The Routledge Handbook of Lexicography.* London/New York: Routledge.
- Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.).** 2011 *e-Lexicography: The Internet, Digital Initiatives and Lexicography.* London/New York: Continuum.
- Gouws, Rufus H.** 2017. La sociedad digital y los diccionarios. Domínguez Vázquez, María José and Maria Teresa Sanmarco Bande. 2017. *Lexicografía y didáctica:* 17-34. Frankfurt: Peter Lang.
- Gouws, Rufus H.** 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, Rufus H.** 2018a. Expanding the Data Distribution Structure. *Lexicographica* 34: 225-237.

- Gouws, Rufus H.** (to appear) Expanding the Use of Corpora in the Lexicographic Process of Online Dictionaries. Piosik, M., J. Taborek and M. Woźnicka (Eds.). 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven, 19.–20. Oktober 2018, Adam-Mickiewicz-Universität in Poznań, Polen*. Lexicographica: Series Maior. Berlin/New York: De Gruyter.
- Gouws, Rufus H.** (to appear a) 'n Verbeterde leksikografiese dataverspreiding- en inligtingsonttrekkingstruktuur. *SPIL-Plus* 2020.
- Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.)**. 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin: De Gruyter.
- Granger, Sylviane and Magali Paquot (Eds.)**. 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Haas, M.R.** 1962. What Belongs in a Bilingual Dictionary? Householder, Fred W. and Sol Saporta (Eds.). 1962. *Problems in Lexicography. Report of the Conference on Lexicography held at Indiana University, November 11–12, 1960*: 45-50. Bloomington: Indiana University.
- Hausmann, Franz J. and Herbert Ernst Wiegand**. 1989. Component Parts and Structures of General Monolingual Dictionaries. Hausmann, Franz J. et al. (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 328-360. Berlin: Walter de Gruyter.
- Kammerer, Matthias and Herbert Ernst Wiegand**. 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Müller-Spitzer, Carolin**. 2013. Textual Structures in Electronic Dictionaries. Gouws, Rufus H. et al. (Eds.). 2013: 367-381.
- Parker, Phil**. 2013. *Do You Know How Generation Z Pupils Learn?* <https://www.sec-ed.co.uk/blog/do-you-know-how-generation-z-pupils-learn/> (Last accessed on 3 April 2020).
- Tarp, Sven and Rufus H. Gouws**. 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
- Varantola, Krista**. 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, Marie-Hélène (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 30-44. s.l.: EURALEX.
- Wiegand, Herbert Ernst**. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin/New York: De Gruyter.
- Wiegand, Herbert Ernst**. 2013. Gedruckte Gebrauchsgegenstände mit lexikographischen Formeigenschaften. *Lexicographica* 29: 285-307.
- Wiegand, Herbert Ernst, Sandra Beer and Rufus H. Gouws**. 2013. Textual Structures in Printed Dictionaries. An overview. Gouws, R.H. et al. (Eds.). 2013: 31-73.

Selecting an Initial Lemma List in Specialized Lexicography: A Case Study in the Field of Graphic Engineering

M.A. Esandi-Baztan, *Department of Materials Science and Metallurgical Engineering, Graphic Expression in Engineering, Cartographic Engineering, Geodesy and Photogrammetry, Mechanical Engineering and Manufacturing Process Engineering, Universidad de Valladolid, Spain*
(mariaangeles.esandi@uva.es)

and

Pedro A. Fuertes-Olivera, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa and International Centre for Lexicography, Universidad de Valladolid, Valladolid, Spain* (pedro@emp.uva.es)

Abstract: Specialized lexicography in graphic engineering has never had a tool in dictionary format that facilitates knowledge acquisition for students and meets their specific lexicographic needs in a technical university context. This paper presents an initial lemma list for a Spanish dictionary in the field of graphic engineering. The Delphi method has been used to assure expert consensus in this area. In addition, two equally professional surveys have been conducted by an expert and endorsed by experts. The fundamental concepts are considered, either because of the difficulty they represent, or due to their importance within our field of knowledge for which engineering documents and technical standards with technical graphic elements have been employed. Students have taken part in the process, thereby allowing identification of convergences and divergences between both groups of participants. The paper concludes with, on the one hand, a final list of accepted concepts and, on the other, of additional uncertain concepts, which could satisfy dictionary user needs and constitute the beginning of a functional online dictionary.

Keywords: SPECIALIZED LEXICOGRAPHY, FUNCTIONAL ONLINE DICTIONARY, DELPHI METHOD, USER NEEDS, EXPERT CONSENSUS, GRAPHIC ENGINEERING, PERCEIVED CONCEPT DIFFICULTY, PRIOR KNOWLEDGE

Opsomming: Die selektering van 'n aanvanklike lemmalys in gespesialiseerde leksikografie: 'n Gevallestudie in die veld van die grafiese ingenieurswese. Gespesialiseerde leksikografie in die grafiese ingenieurswese het nog nooit oor 'n hulpmiddel in woordeboekformaat wat kennisverwerwing vir studente vergemaklik asook hul spesifieke leksikografiese behoeftes binne 'n tegniese konteks op universiteitsvlak bevredig, beskik nie. Hierdie artikel bied 'n aanvanklike lemmalys vir 'n Spaanse woordeboek in die veld van die grafiese ingenieurswese aan. Die Delphi-metode is gebruik om gesaghebbende konsensus in hierdie area te verseker. Daarbenewens is twee gelykstaande professionele opnames deur 'n kundige uitgevoer en

deur kundiges bekragtig. Die fundamentele konsepte is oorweeg, óf weens die uitdagings wat hulle verteenwoordig, óf weens hul belangrikheid binne ons kennisveld waarvoor ingenieursdokumente en tegniese standaarde met tegniese grafiese elemente aangewend is. Studente het aan die proses deelgeneem om sodoende ooreenkomste en verskille tussen beide groepe deelnemers in aanmerking te neem. Hierdie artikel sluit enersyds af met 'n finale lys aanvaarde konsepte en andersyds met 'n lys addisionele vae konsepte wat woordeboekgebruikers se behoeftes sou kon bevredig en die ontstaan van 'n funksionele aanlyn woordeboek sou kon vorm.

Slutelwoorde: GESPECIALISEERDE LEKSIKOGRAFIE, FUNKSIONELE AANLYN WOORDEBOEK, DELPHI-METODE, GEBRUIKERSBEHOEFTE, GESAGHEBBENDE KONSENSUS, GRAFIESE INGENIEURSWESE, WAARGENOME MOEILIKHEIDSGRAAD VAN KONSEPTE, VOORAFKENNIS

1. Introduction

This article discusses some of the main issues involved in a lexicographic project. Firstly, it aims at conceptualizing an online dictionary targeting Spanish students of graphic expression in engineering. Secondly, it presents the basics of research that has "used" experts to select the main and initial lemma list of the dictionary project and submit their proposal to further scrutiny by the potential beneficiaries, Spanish students of graphic expression in engineering.

The project is based on the tenets of the Function Theory of Lexicography (Bergenholtz and Tarp 2003; Fuertes-Olivera and Tarp 2014; Tarp 2008), adding to its main assumptions a major degree of adaptation to meet potential users' needs. In this way, we will show how the presentation and representation of knowledge in the field of Graphic Engineering can be enhanced with the joint participation of experts and students.

Fuertes-Olivera and Bergenholtz (2011), among others, have analyzed some of the changes and challenges brought about by the internet, especially those concerned with lexicographic tasks in production and reception situations. It is true that the internet facilitates the conception and dynamic development of a dictionary as well as the immediate and simultaneous access of an unlimited number of users to lexicographic and extra-lexicographic data. It is possible to update an online dictionary as there is the advantage of its being flexible, dynamic and extensible (Pinnis et al. 2013: 123), resulting in new possibilities for the representation of specialized lexicography. The social and personal use of information and communication technologies enables and supports the initiatives of this representation on devices such as computers, mobile phones and tablets, indicating new patterns in the use of dictionaries.

The data in an online dictionary can be updated and extended and its articles (dictionary entries) allow a customized access and retrieval of information. It must allow for a specific type of information and present a specific selection of lexicographically treated data conditioned by the user's needs (Fuertes-Olivera and Bergenholtz 2018: 191). An appropriate methodology enables immediate

access to data (adding and storing), the content of which is adapted to use situations. As Tarp states, "The traditional dictionary articles with static data and fixed structures should be replaced by articles containing dynamic data which are, so to say, unique for each search related to a specific type of user in a specific type of user situation" (Tarp 2009: 29). We agree that access to lexicographic and multimedia data presented should satisfy those needs within an acceptably short time and with the required degree of detail and the data must be presented in a form that is maximally comprehensible.

In the domain of graphic expression in engineering, data are either scant or difficult to distinguish and systematize. To the best of our knowledge, there are no reference works in this field of knowledge. Hence, we have resorted to novel methods and methodologies for preparing an initial lemma list representative of this field and allowing for enlargement. As shown below, we have used the Delphi method to start our project. This consists of initiating a consensus process by a panel of experts in this area; these are responsible for developing a *Concept inventory for Engineering Graphics*, e.g. the method used by Sadowski and Sorby (2015), who pioneered the use of this technique to produce such an inventory. The results obtained, in the present study, will be then re-evaluated by experts and students so as to analyze more focused objectives, i.e. the ones with which our students are more familiar and in greater need of assistance. In fields such as business (Czinkota and Ronkainen 1997), engineering (Hallowell and Gambatese 2010) and medicine (Flume et al. 2010) there are numerous studies that use this technique to formalize the consensus process of a panel of experts and to develop explicit recommendations for a majority of people, but adaptable to the specific needs of individuals.

Graphic communication in engineering involves using visual material to relate ideas. A drawing is a form of graphic communication that represents an idea, a concept or an entity which actually or potentially exists in life (Goetsch, Chalk and Nelson 2000). The domain of graphic engineering requires a complex cognitive process and the conceptual system that assists certain mental operations in this domain is also complex (Kageura 1997: 119). In this paper, we aim to demonstrate that it is possible to make this conceptual system accessible via a lexicographic tool, and to show the characteristics of its lexicographic representation. We must first identify the concepts that best fit our objectives, establish the structure and content of each concept, and verify whether the Delphi method used in its selection is a success. In each case, we make a proposal for a definition that helps to disambiguate the concept, that is, to determine the meaning of the concept in the context. The consensus opinion of a panel of experts minimizes subjectivity, whilst allowing us to draw up a list of concepts avoiding systematic errors, and define the lexicographic functions of the dictionary.

Section 2 deals with lemma selection in dictionaries, the influence of the use of corpora and other recommended methods in specialized lexicography, the modified Delphi method and the influence of the sources used. We show,

in section 3, the results acquired in two rounds with experts (as an example of analysis to support the strategy based on the experience and consensus of a group of experts) and one extra round with students. Section 4 focuses on the discussion and analysis of the data, a proposal for concept definition and data structure of an initial lemma list, together with the convergences and divergences between students and experts. Finally, the conclusion (section 5) attempts to indicate the practical implications of this project and future research.

2. Material and methods

Following the principles of the Functional Theory of Lexicography and the reasoning of the deductive method, the process begins by focusing on the user of the dictionary, their situation and their needs. Once these premises have been identified, the nature of the field of specialized knowledge conditions its lexicographic treatment.

A review of the literature on lemma selection processes indicates that there are four main methods for selecting the lemma list of a new dictionary project: (a) existing dictionaries; (b) grammar and etymology; (c) canonical texts (e.g. lemmas must be selected from literary works); and (d) corpora. Authors such as Bowker and Pearson (2002), Čermák (2003), Pérez Hernández, Moreno Ortiz and Faber Benítez (1999), among others, have analyzed the advantages and disadvantages of the different methods, including the choice of specialized terms. For instance, Verlinde and Selva (2009: 598) have studied corpus-based versus intuition-based lexicography, and have concluded that the former provides the lexicographer's personal intuition with important and necessary empirical evidence; however, intuition is still useful for filling the gaps that may occur in a corpus. In other words, the authors have called our attention to the necessity of submitting any process of lemma selection to a critical analysis, as each way of selecting has its pros and cons (see Fuertes-Olivera and Tarp 2014 for a review of this issue). Following this line of thought, we have used a different method for selecting the initial lemma list of our project, namely, the Delphi method.

2.1 The Delphi method

The Delphi method (Linstone and Turoff 1975) consists of structuring the communication process of a group of experts who function as a set. The group is comprised of a selected panel whose opinion is sought via successive strategically designed surveys. We are interested in the fundamental concepts considered, either because of the difficulty they present, or because of their importance within the field of knowledge and our field of action. The method allows us to identify convergences of opinions and predicts provisional consensus in order to finally determine and quantify the consensus of the group. It is possible to use both quantitative techniques (Rowe and Wright 1999) and

qualitative ones when we are interested in the interpretation, understanding and experimentation of the social context within which the data are collected, resulting in a holistic understanding of the latter (Mason 1996). Qualitative research should be carried out strategically but in a flexible and contextual way (Mason 2002: 7) so as to obtain highly reliable data from, in our case, reputable experts.

There are two distinct methods of the Delphi process: the *paper-and-pencil* version known as "Delphi Exercise" or conventional Delphi and the "Delphi Conference" or the real-lucres Delphi version (Linstone and Turoff 1975: 5). The differences between the two lie in the use of a physical or computer support, and in that of software for data analysis in the second case. The latter allows, in real time, the drawing-up of a new questionnaire addressed to the group, but requires that the characteristics of the communication be defined before the Delphi process begins. However, in a Delphi exercise with "paper and pencil" these characteristics can be adjusted for the next round, depending on the answers obtained in the previous session. In essence, the Delphi technique consists of a series of sequential questionnaires interspersed by controlled feedback used to obtain the consensus of opinion from a panel of experts (Powell 2003: 377).

2.1.1 Process

We focus on the *expert method endorsed by experts*, with, therefore, an expert who designs the questionnaires that are sent to a group of fellow professionals. To carry out this work, we have informants chosen among professional professors, from the graphic expression in engineering (*EGI*) area of knowledge, to identify the fundamental concepts, and student informants of an engineering school (*Escuela de Ingenierías Industriales* at the University of Valladolid). The members of the panel of experts are five full-time professors (from a set of ten) in the field of *EGI* at the University of Valladolid. They were selected on the basis of their long experience in *EGI* (between 26 and 40 years) in subjects such as Metric Geometry, Representation Systems, Technical Drawing and Computer-Aided Design and Technical Drawing in industrial engineering degrees.

The first panel of experts' contact is a semi-structured "brainstorm", in which our research topic is set out whilst considering other topics related to the fields of knowledge of this research, namely, Lexicography (specifically with the Function Theory of Lexicography) and the *EGI*. The contribution to the debate by the members of the group was enriching and comprehensive; it was recognized that an online dictionary would represent a real opportunity for students, due to their needs and their familiarity with digital devices.

Two rounds were carried out with feedback on the results and the paper-and-pencil option was employed. Once the first paper questionnaire had been completed, it was returned and the results were analyzed so that the necessary adjustments could be made depending on the comments received. In accordance with the results and contributions, a second questionnaire was prepared

based on the assessment of the group response, and the panellists had the chance to re-evaluate their own answers in light of group responses. The Delphi method was modified by including, in the second round, not only the concepts identified by our experts but also some of those identified in the study conducted by Sadowski and Sorby (2015) on graphic engineering. We accordingly considered and evaluated the impact of all of these items.

2.1.2 Material

The questionnaires are derived from two compendia of practices used in the student learning process, namely, a theoretical manual on representation systems and the standard document on the general principles of presentation in technical drawings. The participants were familiar with the internship documents, three of them being, in fact, co-authors of the 2018 collection. We considered it necessary to use a standard as a resource since this involves published recommendations which provide communicative facts. The resources for the round I questionnaire statements were as follows:

- Nieto Oñate, M., V. González García and R. López Poza. 1970. *Prácticas programadas de Dibujo Técnico. Delineantes industriales*;
- Álvarez Garrote, S., E. Rebotto Rodríguez, Q. Rodríguez Ovejero and M. Escribano Negueruela. 2018. *Prácticas de Expresión Gráfica III*;
- The standard UNE 1-032-82 (1982) corresponding to ISO 128;
- González García, V., R. López Poza and M. Nieto Oñate. 1982. *Sistemas de representación. Sistema diédrico*.

2.1.3 Delphi process and questionnaire

The Delphi process consisted of two rounds conducted with the panel of experts and a final evaluation. Questionnaire I was not structured, seeking an open answer (Powell 2003: 378) and, consisted of 111 statements (12 + 22 + 64 + 13); respondents might firstly identify the relevant concepts and then assess, on a scale of 1 (little) to 5 (a great deal), the estimated importance and the difficulty foreseen for a semi-experienced student in the *EGI* at the university. In all the statements the participants were given the option to accept (respond), reject (not respond), provide comments and include the concepts they considered of interest and which do not appear in the statements. They were asked to identify, at their discretion, the concepts among the proposed statements, considering that the set consists of two parts, one verbal and one graphic. Regarding the four sources, the first two displayed identical characteristics: statements with a descriptive-informative verbalized statement and graphic support (a sample item is shown in **Figure 1**); for the statements corresponding to the UNE 1-032 (ISO 128) standard, the concepts (verbal and graphic) were previously identified at source and the standardized verbal denomination was used

(see **Figure 2**). The fourth source corresponded to representation systems. In this case, the concepts were previously identified by an expert and were the result of decomposition into simple geometric constructions of complex ones, that is, the decomposition of complex concepts into simpler ones (see **Figure 3**). The constructions were analyzed individually, defined and gave rise to new unified and sometimes reformulated concepts; statements consisted of two parts, one verbal and one graphic. It should be noted that in these items the verbal component and the graphic component were equivalent, highlighting the expressiveness of the graphic component.

• *Croquizar la pieza facilitada al mayor tamaño posible. Las medidas de las vistas serán proporcionales a las dimensiones de la perspectiva.*

Valora la importancia

<input type="radio"/>				
1 poco	2	3	4	5 mucho

Valora la dificultad

<input type="radio"/>				
1 muy fácil	2	3	4	5 muy difícil

Comentarios:

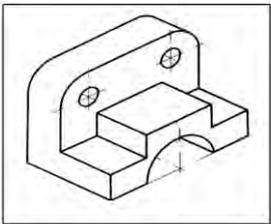


Figure 1: Example of a non-structured item with an open answer, with a verbal and graphic component (Source: Questionnaire I)

• *Vistas particulares*

Valora la importancia

<input type="radio"/>				
1 poco	2	3	4	5 mucho

Valora la dificultad

<input type="radio"/>				
1 muy fácil	2	3	4	5 muy difícil

Comentarios:

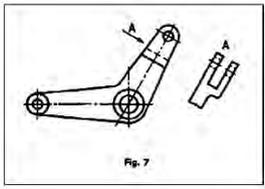


Figure 2: Example of concepts identified thanks to their graphic support (Source: UNE 1-032; ISO 128)

Parte 3. La mínima distancia entre dos rectas que se cruzan, es la distancia de cualquier punto de la recta s al plano π .

Valora la importancia

1 poco 2 3 4 5 mucho

Valora la dificultad

1 muy fácil 2 3 4 5 muy difícil

Comentarios:

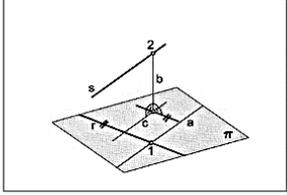


Figure 3: Example of representational systems. Complex concepts are simplified (Own source)

In questionnaire II, items were specific, e.g. **Figure 4**, with the aim of having experts review and either confirm or not confirm the results of the first round. The results of questionnaire I allowed us to identify, relate and reformulate questionnaire II. The format avoided ambiguity in the statements so that the concepts identified by the expert informants were unambiguous and rigorous. There were 54 concepts previously identified and they were ranked in order from most to least important, being accepted by at least three experts; as we have already mentioned, 14 concepts from the aforementioned study by Sadowski and Sorby (2015) were also added.

- **Localización del plano de corte**
Valora la importancia

1 poco 2 3 4 5 mucho

- **Teorema de las tres perpendiculares**
Valora la importancia

1 poco 2 3 4 5 mucho

Figure 4: Example of a specific item without ambiguity (verbal component) (Own source)

2.1.4 Administration of the questionnaire

The questionnaire was delivered by hand at different moments and there was no communication between the members while the test was being carried out,

thereby guaranteeing anonymity. Once the questionnaires had been collected and the answers analyzed, the results of the first round were passed on to the participants, who receive the identified concepts in order of greater or lesser importance according to the group; this constituted the main part of the second questionnaire. It was necessary to make certain adjustments, because in the source literature there were anachronistic denominations of the same concept (and from now on they will appear in the same item), and also because there was ambiguity in several cases. Some concepts were dismissed because experts rejected them, and several concepts from the study *Concept inventory for Engineering Graphics* were incorporated. The second questionnaire was administered and carried out analogously to the first one two months later.

2.1.5 Measuring consensus

The degree of consensus is defined by the number of participants who have identified the concept. We establish four sections, and if at least three participants have accepted the concept we consider it to be "identified" (Table 1).

Table 1: Terms that define the degree of consensus

Term	Participants that accept
Full consensus	All participants (5) identify the concept
Very good consensus	At least 4 participants identify the concept
Good consensus	At least 3 participants identify the concept
Some consensus	1 or 2 participants identify in consensus
Without consensus	No participant identifies the concept

2.1.6 Measuring importance assigned to the concept

We quantify the degree of consensus regarding the importance assigned to the concept among the concepts identified, by at least three experts (Table 2).

Table 2: Terms that define the importance assigned

Importance	Average
Very important	Media 5
Important	Media ≥ 4
Slightly important	Media ≥ 3
Less important	Media ≥ 2
Without importance	$0 \geq \text{Media} \geq 1$

2.2 Student survey

Out of the Delphi process and related with it, this process produces additional information deriving from a comparison of experts' and students' opinions. As the final objective of this process is to design a pedagogical cognitive tool, it must include these identified concepts to successfully fulfil the task of satisfying users' needs, initially in the academic context and probably in the professional one at a later date. The fundamental concepts are considered either because of their difficulty or due to their importance within the field of knowledge; here both experts and students have something to contribute. Currently, the students are doing an industrial engineering degree at the School of Industrial Engineering (Escuela de Ingenierías Industriales (EII)) at the University of Valladolid (UVa). We must remember that they are involved in a broad spectrum of technical studies (not only in the graphic field). They are taught engineering graphic subjects in their first, third and fifth semesters, respectively. They were invited to answer the questionnaire if they wished. We made the second Delphi survey (Figure 5) accessible to these students at the university online campus for three weeks, and we obtained eighty-six answers. A comparative study of, and subsequent reflection on, the difficulty assigned by the experts and students is summarized in the next section.

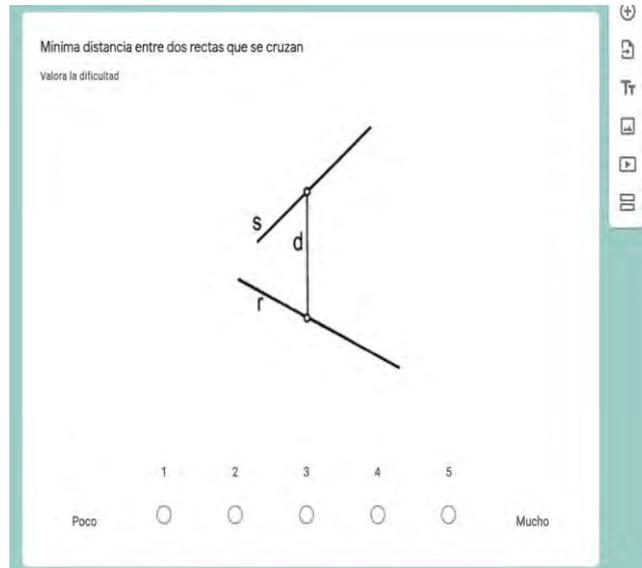


Figure 5: *Minimum distance between two lines that cross (without intersection).* Example of online survey accessible at the moment in (<https://docs.google.com/forms/d/1jj-cYVLXSCgipRPzxShSIAX1QxAURW4vseqfB0Zd5V4/edit?ts=5dd2590a>)

3. Results

We present the concepts identified by the experts (three is the minimum number of experts required to accept a concept); these answers constitute the results of the first round. Experts again give their opinion so that they can reaffirm their initial assessment, but this time on a document in which the concepts appear in order of importance decided on by the group; such responses represent the results of the second round. Finally, we present a comparison of the opinions of experts and students regarding the difficulty of these concepts.

3.1 First round results

Appendix 1 shows the seventy concepts that at least three experts have accepted; they show an average greater than 4 in terms of importance (1 is of little importance and 5 is very important). It also shows their degree of difficulty.

3.2 Second round results

In the second round we present, in **Appendix 2**, the previously identified and ordered sixty-seven concepts of greater or lesser importance in the opinion of the group; the different denominations corresponding to the same concept have been unified, with the incorporation of some of the concepts of graphic engineering that were considered for *Engineering Graphics Concepts: A Delphi Study* (Sadowski and Sorby 2015), and which are suitable for our target user. In order to identify and quantify the level of consensus among experts on a series of concepts belonging to the field of graphic expression in engineering, we compare the results from both stages (**Appendix 3**) and establish a high degree of consensus, namely, a difference of less than or equal to 0.2 points; as a result, we obtain an initial list of concepts obtained by a consensus of experts. **Table 3** summarizes the final statement of 20 concepts agreed on by the experts together with their ratings. Note that in **table 3**, the concepts are ordered from highest to lowest importance in round II. This should be the case since this second questionnaire is definitive because the experts reaffirm and highlight their opinion when necessary.

Table 3: Consent concepts by experts

Concept	Importance	Importance
	Round I	Round II
1. Acotar	5	5
2. Seguir las normas / Normalizado	5	4,8
3. Elección de las vistas	4,8	4,8
4. Rectas horizontales de plano	4,8	4,8
5. Rectas frontales de plano	4,8	4,8
6. Posición relativa de las vistas	4,71	4,6
7. Teorema de las tres perpendiculares	4,6	4,6
8. Método de proyección del primer diedro	4,4	4,6
9. Corte / corte por un plano	4,3	4,4
10. Disposición de los cortes	4,6	4,4
11. Línea fina de trazos y puntos	4,4	4,4
12. Rayado de diferentes partes cortadas de una misma pieza	4,4	4,4
13. Recta perpendicular a un plano	4,4	4,4
14. Corte en ángulo / Corte por dos planos concurrentes	4,4	4,3
15. Disposición normalizada de las vistas / Seis vistas normalizadas	4,25	4,2
16. Corte de nervios, elementos de fijación, árboles, radios de ruedas y otros elementos análogos	4,4	4,2
17. Identificar el plano de corte	4,33	4,2
18. Sección abatida / sección transversal	4,2	4
19. Línea fina de trazos	4,2	4
20. Línea de máxima inclinación de un plano	4,2	4

3.3 Students' assessment

A version of the second survey had been digitally sent and eighty-six students freely chose to fill in the questionnaire. This allowed us to identify the concepts that involved some degree of difficulty and that resulted in a cognitive need that could be satisfied. The minimum difficulty value assigned is 1.65 and the maximum is 3.45 (1 to 5). It should be noted that of the 20 concepts lower level of difficulty, 18 correspond to the same UNE 1032 standard (ISO 128) which is studied in class since the very first day. This reinforces our opinion that the direct teacher-student relationship is positive. Abstract concepts and those requiring a working method are perceived as having a higher degree of difficulty. See **Appendix 4** for the results.

Likewise, being able to compare the (expert) teacher's point of view with that of the student allows us to draw conclusions regarding both convergence and divergence on perceived concept difficulty.

3.4 Experts and the concepts of Sadowski and Sorby's study

The experts acknowledge the concepts proposed deriving from the work of Sadowski and Sorby. These are the last fourteen concepts in **Appendix 2** and their importance is rated between 3.5 and 5.

4. Discussion

As a result of the Delphi process, an initial list of concepts is obtained. This list will be the starting point for further decisions affecting the configuration of the dictionary and to attain our objectives:

- (i) We have an initial lemma list
- (ii) The data structure must be configured
- (iii) We suggest precise definitions adapted to the user
- (iv) We analyze the rating when appreciating the difficulty of some concepts

4.1 The initial lemma list

The initial lemma list includes twenty concepts suitable for intermediate students of industrial engineering, and which can fulfil their information needs. This list initiates a process to visualize graphic engineering concepts in terms, on the one hand, of the verbal part for future engineers and, on the other, of the graphical part for lexicographers. Each concept may increase the initial lemma list with both verbal lemmas and graphic lemmas (see our entry for *acotar* below).

There are twenty concepts agreed on by the experts (listed in the Results section) from a set of seventy-one identified and accepted in the process we have carried out. The initial lemma list consists of twenty concepts and, with the exception of one, all are multi-word lexical items. This contradicts the structure and format of many traditional dictionaries, but is in line with current research on specialized lexicography (Fuertes-Olivera and Tarp 2014).

We firmly believe that we can use the Delphi method to augment an essential list of concepts in any other sub-field of graphic engineering, whilst ensuring that the needs of a particular user of a specialized dictionary are met. In addition, the second questionnaire is available on the university's digital services and can be reused in the future.

4.2 Configuration of the data structure

We determined the data structure for organizing pieces of information in order to satisfy user' needs and to provide an efficient repository with access functions. When determining the functions of the dictionary, it is necessary to con-

sider the visual reference, the structure and functions of the graphic representation and its functional relationship with the lexical part with which is complemented. We must account some considerations regarding the data storage structure and the data presentation structure:

- The data storage structure must be a lexicographic record that has a template format organized in fields that houses lexical and graphic units that must be labelled together (whenever possible).
- The data presentation structure must be appropriate for each level of complexity and present simultaneously a verbal component and an associated graphic component (up to an advanced level). At the initiation level in EGI, both components are simple and direct textually and graphically. At an intermediate level both components are complex, so they include previous reformulations. At an advanced level there is only a graphic component or the textual component would be minimal (it would be very complex, imprecise, and incomplete to lexicalize certain geometric concepts or to lexicalize to precisely define an object to be mass-produced).

Studies on specialized lexicographic need to address the data fields that we propose to be linked to the types of information required (Fathi 2014). We have included them in three blocks for each concept: **grammatical information**, **meaning**, and **graphic representation**. Grammatical information include collocations, lexical relationships (antonyms, synonyms etc.) and a domain label. Meaning describes as accurately as possible the concept given in literature definitions, duly standardized, by authors of accredited prestige, before our proposed definition was given. Graphic representation includes standardized forms and examples of practical application.

The example of the concept "*acotar*" (*dimensioning*) shows how we have organized the data structure in the abovementioned three sections, each of which contains subsections:

Concept: *acotar* (*dimensioning*) (each "collocation" and 'inflexions' can be also included as lexical and/or graphic lemmas in future developments of the online dictionary. Hence, each basic or fundamental concept may result in more than 100 different lemmas in the proposed dictionary)

1. **Grammatical information**

Word class: *Verbo* (verb)

Inflexions: *acotación, cota, acotado*

Collocations grouped by domain label: *Dibujos técnicos. ACOTACIÓN: acotación funcional directa; acotación funcional indirecta; elementos de acotación; método de acotación; acotación de radios; acotación de cotas superpuestas; acotación en serie; acotación a partir de un elemento en común; acotación en paralelo; acotación por coordenadas; acotación combinada; acotación de elementos equidistantes dispuestos linealmente a intervalos; acotación de elementos equidistantes*

dispuestos angularmente a intervalos; acotación simplificada de chaflanes; acotación de conjuntos; acotación de las tolerancias; cota; cota funcional; cota no funcional; cota auxiliar; cota lineal; cota angular; cifra de cota; línea auxiliar de cota; línea de cota; cotas superpuestas; líneas de cota oblicuas; cifras de cotas angulares; cota fuera de escala; símbolo que precede la cifra de cota; disposición de las cotas; cadenas de cotas; línea de cota del radio; cota angular; cota de altitud; cota de nivel
Dibujos técnicos. Tolerancias de cotas lineales y angulares: cota con tolerancia; cota nominal; cota superior; cota inferior; cota con límite superior; cota con límite inferior; cota del agujero; cota del eje

Dibujos técnicos. Tolerancias de cotas lineales y angulares: cota con tolerancia; cota nominal; cota superior; cota inferior; cota con límite superior; cota con límite inferior; cota del agujero; cota del eje

Especificaciones geométricas de producto (GPS). Acotado geométrico: acotado geométrico; acotación geométrica; acotación geométrica de piezas; elemento geométrico acotado; cilindro acotado; superficie media acotada; característica a acotar; figuras acotadas

Dibujos técnicos. Métodos de proyección. Parte 3: Representaciones axonométricas: acotados

Dibujos técnicos. Roscas y piezas roscadas: Acotación de piezas roscadas, acotar la longitud del roscado, acotar la profundidad del agujero ciego

2. Meaning

Definitions by authors of accredited prestige and standardized ones:

- a. *señalar las magnitudes que determinan la geometría o características geométricas de un objeto atendiendo a aspectos tales como funcionamiento, fabricación, montaje o verificación de las magnitudes* (A. Valin)
- b. *el proceso de situar las medidas de la pieza sobre el plano, y los elementos que reflejan las medidas reales de la pieza se denominan cotas* (Félez and Martínez 2008)
- c. *acotar es dimensionar, las dimensiones y notas describen el tamaño, el terminado y otros procesos de fabricación en el dibujo, de tal forma que el objeto quede total y precisamente definido. Las dimensiones describen el tamaño y la ubicación de los elementos de un objeto* (Giesecke et al. 2016)
- d. *(De «cota», número.) Poner números o cotas en un plano* (Moliner and Moliner 1998)

Our proposed definition:

- e. *proceso de consignar sobre la representación de un objeto sus dimensiones reales, definiendo con precisión el tamaño, la ubicación y la geometría (si fuera necesario) teniendo en cuenta el proceso de fabricación y las recomendaciones normalizadas de la representación*

3. Graphic representations

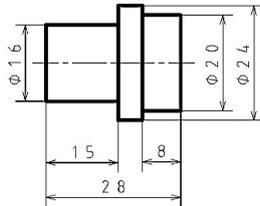


Figure 6: Examples of practical application
The drawing precisely defines the geometry and dimensions of an object.

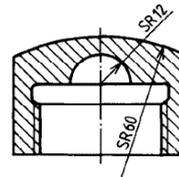


Figure 7: Standard representation UNE 1-039-94

4.3 Deconstructing definitions

We recommend precise definitions adapted to the user. This means adapting the definition to their needs and situation. This online dictionary has a defined data presentation template with the content located in a specific fixed place on the screen. The data access protocol is also defined and is clickable by user situation that corresponds to the function of the dictionary. There are adapted definitions of the same concept because we consider the possibility that the student needs previous knowledge to progress. In some cases, these adapted definitions are interpretations and in another, they are possible implications that a student at a training level is not able to assimilate initially. We start from the most abstract general definition and present definitions of simpler concepts that the student may need to review. All the details of this proposal are part of a doctoral thesis that will be defended in the coming months. The needs mentioned occur in an extra-lexicographic situation involving acquisition of technical principles where prior knowledge is required. According to the Function Theory of Lexicography, the context of acquiring technical knowledge is obviously that of cognitive situations. However, the purpose of the drawings used and the geometric elements involved refer unequivocally to operational (Figure 3) and interpretative (Figures 6, 7) situations. We note, as an example, the concept *acotar* (*dimensioning*) and a proposed deconstructed definition in case some prior knowledge must be clarified or made accessible. The information we consider should be made accessible to the user, e.g. by clarifying its definition verbally and graphically, as shown below:

Acotar: proceso de consignar sobre la representación de un objeto sus dimensiones reales, definiendo con precisión el tamaño, la ubicación y la geometría (si fuera necesario) teniendo en cuenta el proceso de fabricación y las recomendaciones normalizadas para la representación

Dimensioning: the process of adding the true dimensions of an object on

its representation, defining accurately size, location and geometry (if necessary) considering the manufacturing process and standardized recommendations of the representation

- *consignar sobre la representación de un objeto (adding on the representation of an object), hace referencia por una parte a que se debe reservar un espacio para disponer las cotas, por otra, a que son necesarios ciertos elementos que sitúan y definen las cotas con precisión y finalmente a que la representación puede ser sobre una representación plana o una perspectiva*
- *consignar: destinar un lugar o sitio para poner o colocar en él algo (RAE 2014)*
- *dimensiones reales de un objeto, el objeto puede estar representado a escala por lo que los valores de las cotas han de ser los correspondientes al objeto real*
- *definir con precisión, hace referencia por una parte que debemos limitar las líneas de cota al elemento geométrico y por otra parte que debemos acotar sin ambigüedad lo que implica no dar información dimensional que pueda conducir a una indeterminación de la geometría*
- *definir el tamaño implica definir los elementos geométricos que conforman el objeto y sus características y, conocer de antemano las cotas necesarias para definir el objeto*
- *definir la ubicación de los elementos geométricos y de sus relaciones entre sí*
- *puede ser necesario definir, aclarar, precisar la geometría; se refiere a los símbolos que preceden a la cifra de cota y que "permiten la identificación de las formas y mejoran la interpretación del dibujo" (UNE 1-039)*
- *elegimos las dimensiones con aplicación directa en la fabricación del objeto, no las que son consecuencia del proceso de fabricación*
- *las pautas para consignar las dimensiones y los elementos que intervienen en la acotación de un objeto sobre su representación está(n) especificado(s) en distintas Normas*

Figure 8 shows, as an example, the data proposal when starting a query with the 'acotar' key according to the interpretative function. The basic dictionary template model is organized into blocks containing: the lemma (key) or concept and its lexical family; the associated graphic representation; grammatical information (if necessary); function and definitions or descriptions (depending on the function).

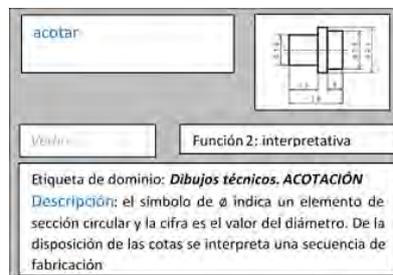


Figure 8: Acotar key, interpretative function

From a quantitative perspective, the election of the concept '*acotar*' is justified as an example. As a result of the Delphi method, this concept was identified by all the experts in questionnaire I, with an average importance of 5 and an average difficulty of 4.5, and the same results were obtained in questionnaire II. This reaffirms the initial opinion and consolidates a full consensus of the group of experts. In addition to our findings we have the results of research by Sadowski and Sorby (2015), the concept '*dimensioning*' occupies position 9 of 39 in the *Concept inventory for Engineering Graphics*.

4.4 Rating of the concepts according to their difficulty

We analyze the rating when appreciating the difficulty of some concepts by contrasting the opinion of six experts and eighty-six students regarding 52 concepts (see **Appendix 5**). In the graph (see **Figure 9**) we can observe the trends, convergences, divergences and rating of both in absolute terms.

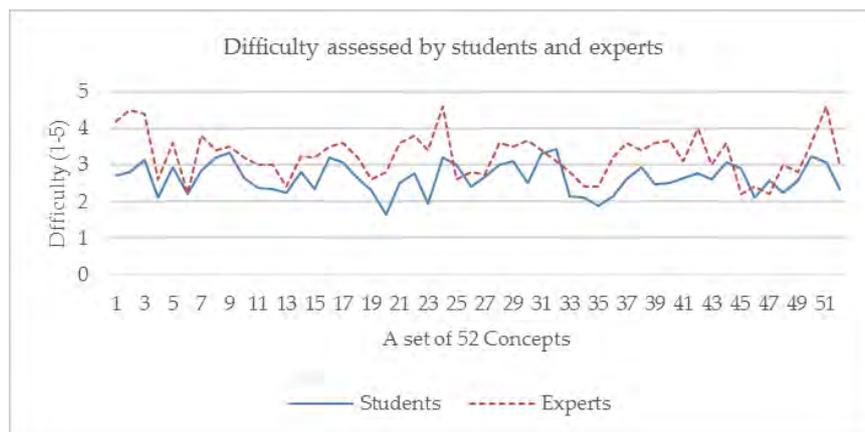


Figure 9: Difficulty assessed by students and experts with regard to 52 concepts. The dotted line represents the opinion of the experts and the solid line the opinion of the students.

A relevant finding when we first look at the graph is that the experts generally assign greater difficulty values to the concepts than the students. However, they concur on six concepts and a certain parallelism is appreciated in both lines. Next, we observe a greater oscillation in the rating of concepts by experts compared to that in the case of students. An explanation of why experts appreciate a greater difficulty regarding concepts is related to their proficient understanding of these and their bearing in mind the implications and the external and internal references. However, in the case of students, there is less oscillation in the rating given to the concepts. In a learning process, with contact, as in

this case, between teachers and students, the difficulties presented by the concepts are exposed and resolved, so the degree of difficulty is lower and the perception is more homogeneous.

Both concur when assessing the following concepts: **6** *línea llena gruesa* (continuous wide line ISO 128-24); **9** *corte longitudinal de una forma de revolución (que contiene detalles regularmente repartidos no situados en el plano de corte y siempre que no se produzca ambigüedad)*; **13** *corte total o corte por un plano*; **27** *Disposición normalizada*; **31** *corte en ángulo o corte por dos planos concurrentes* and **49** *representación de elementos repetitivos*.

The most prominent divergences appear in the following concepts: **25** *segmento paralelo a un plano coordenado*; **32** *Sección múltiple abatida o disposición de las secciones sucesivas*; **45** *indicación del método de proyección* and **47** *Interrupción del rayado*. Earlier, we mentioned that the expert recognizes the external implications of a concept which is more complex. For example, in concept **32**, an expert is aware that an object can also be sectioned by a conical and cylindrical surface, whereas a student's initial approach would probably be to expect a flat surface.

It is worth highlighting the concepts that experts perceive as involving greater difficulty and those which are more difficult for students (four from those of expert assessment and eight from the students'). It should be noted that two concepts appear in both choices:

The concepts which, according to the experts, imply much greater difficulty vis-à-vis the other:

- 2. *Acotar*
- 3. *Seguir las normas/normalizado*
- 24. *Teorema de las tres perpendiculares*
- 51. *Línea de máxima inclinación de un plano*

The concepts which, in the opinion of the students, are considerably more difficult than the others:

- 3. *Seguir las normas/normalizado*
- 9. *Corte longitudinal de una forma de revolución (que contiene detalles regularmente repartidos no situados en el plano de corte y siempre que no se produzca ambigüedad)*
- 16. *Planos no paralelos a los de proyección*
- 24. *Teorema de las tres perpendiculares*
- 29. *Rosca métrica*
- 32. *Sección múltiple abatida o disposición de las secciones sucesivas*
- 44. *Sección abatida o sección transversal*
- 50. *Segmento mínima distancia*

We propose adding these concepts, in a second phase of the project, to the twenty concepts of the initial lemma list. This means adding six concepts to the initial list. These concepts are identified as numbers 9, 16, 29, 32, 34 and 50.

Finally, we must cite the cases, hitherto unmentioned, in which the students assign greater difficulty to a concept than the experts: *25 segment paralelo a un plano coordenado*; *45 Indicación del método de proyección*; *47 interrupción del rayado*. In fact, in these three cases the difficulty assigned by both students and experts is less than three. However, since there is a discrepancy we will ensure the inclusion of these elements in successive phases.

5. Conclusion

Specialized concepts require specialized codes for their definition and representation and, therefore, differentiated criteria are required to compile a specialized dictionary. This study concerns "an expert method endorsed by more experts", using the Delphi technique to identify and agree on an expert list of concepts that are not dealt with from a lexicographic point of view in the literature on *EGI*. We have been able to visualize certain concepts that students find slightly or considerably difficult, a factor which will probably hinder their learning. We believe that the difficult concepts, identified as such by experts and tested by students, must be included in our dictionary proposal in order to fulfil their cognitive needs.

We mainly, but not exclusively, propose, in line with the tenets of the Function Theory, a cognitive function dictionary (Bergenholtz and Nielsen 2006: 287) with online access. This will contain and present multimedia data on a specialized area of knowledge, involving both a verbal and a graphic component in coexistence, depending on operations relating to communication or enquiry. As "functions are the heart and soul of lexicography", it is understandable that lexicographic functions govern decisions about content, form, preparation and access to lexicographic data. Our proposal is for specialized lexicography and an online dictionary which is "multi-functional allowing for individualized data access" (Fuertes-Olivera and Tarp 2014: 62-64), as we have already mentioned (see Section 4.3.)

We believe that communicative, cognitive, operational and interpretive situations take place in a specialized learning context in the field of graphic engineering. All of them do not take place every time and at the same time, but it depends on the characteristics and needs of the users' and their situation. A general notion of graphic expression in engineering indicates that:

- a drawing (a projection of an idea or an object onto a plane) is a "message" that can avoid language barriers
- it deals with a process of acquiring practical knowledge for engineers
- to manufacture any single object, it must first have been graphically defined and drawn
- some of the drawings (iconic representation) shown in this study (**figure 3 and 6**) can and should be interpreted with the appropriate prior knowledge.

In the near future, it may be necessary to expand the literary sources within the field of study (diedric system, axonometric system, *UNE* and ISO standardization), other fields within *EGI* (CAD) and extensively to Industrial Graphic Engineering. The structure defined for the data (**grammatical information, meaning and graphic representation**) in this study is applicable to all the dictionary entries with a verbal and graphic presentation for each of the collocations. It is possible and advisable to repeat the Delphi method with expert engineers and the questionnaire on the online campus with future students, thereby enlarging the results and content of the dictionary.

We believe that it will be necessary to explain the nature of the graphic component and the linguistic-graphic relationship for identifying and representing particular specialized knowledge in *EGI*. It is a fact that the domain of graphic engineering requires a complex cognitive process, and that the conceptual system underlying many of the mental operations in this area is also complex (cf. Kageura 1997: 119), these mental operations require geometric constructions with points, lines and surfaces. It is likely that future consultations on an online specialized dictionary can be made for graphic units that represent graphic concepts, ideas and objects.

Acknowledgements

We would like to express our gratitude to our colleagues, professors S. Alvarez, E. Rebotto, Q. Rodríguez, A. Santos and A. Valin of the University of Valladolid (*UVa*) for their honesty and for finding time between their teaching activity to cooperate in this work; to the students of *EGI* (Graphic Expression in Engineering) *GIE* (Electrical Engineering Degree) and *GIM* (Mechanical Engineering Degree), to the students of *DAO* (Computer Aided Design) at *GIM* and *GIDIyDP* (Industrial Design Engineering and Product Development) at the University of Valladolid, for their prompt and altruistic collaboration. We also thank two anonymous reviewers for the suggestions made to a previous draft of this paper.

References

Dictionaries

- Moliner, M. and M. Moliner. 1998. *Diccionario de uso del español*. Vol. 2. Madrid: Gredos.
- Real Academia Española. 2014. *Diccionario de la lengua española*. 23rd edition, [versión 23.3 en línea]. <<https://dle.rae.es>> [Last accessed: July, 2, 2020].

Other Literature

- Álvarez Garrote, S., E. Rebotto Rodríguez, Q. Rodríguez Ovejero and M. Escribano Negueruela. 2018. *Prácticas de Expresión Gráfica III*. Valladolid: Los autores.

- Arias, A.M., M. de la Torre and M.I. Fijo. 2013. *ICTIOTERM: Base de datos terminológicos y de identificación de especies pesqueras de las costas de Cádiz y Huelva*. Available at: http://www.ictioterm.es/proyecto_ictioterm.php [Last accessed: August, 21, 2020].
- Bergenholtz, H. and S. Nielsen. 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12(2): 281-303.
- Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories. On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *HERMES. Journal of Linguistics* 31: 171-196.
- Bowker, L. and J. Pearson. 2002. *Working with Specialized Language. A Practical Guide to Using Corpora*. London/New York: Routledge.
- Čermák, F. 2003. Source Materials for Dictionaries. Van Sterkenburg, Piet (Ed.). 2003. *A Practical Guide to Lexicography*: 18-25. Amsterdam/Philadelphia: John Benjamins.
- Comité técnico AENOR. 1982. Norma UNE 1-032-82 (ISO 128). *Dibujos técnicos. Principios generales de representación*. Madrid: AENOR.
- Comité técnico AENOR. 1994. Norma UNE 1-039-94. *Dibujos técnicos. Acotación. Principios generales definiciones, métodos de ejecución e indicaciones especiales*. Madrid: AENOR.
- Czinkota, M.R. and I.A. Ronkainen. 1997. International Business and Trade in the Next Decade: Report from a Delphi Study. *Journal of International Business Studies* 28(4): 827-844.
- Fathi, B. 2014. Experts and Specialised Lexicography: Perspectives and Needs. *Termin@lia* 9: 12-21.
- Félez, J. and M.L. Martínez. 2008. *Ingeniería gráfica y diseño*. Madrid: Editorial Síntesis.
- Flume, P.A., P.J. Mogayzel Jr., K.A. Robinson, R. Rosenblatt, L. Quittell and B.C. Marshall. 2010. Cystic Fibrosis Pulmonary Guidelines: Pulmonary Complications: Hemoptysis and Pneumothorax. *American Journal of Respiratory and Critical Care Medicine* 182(3): 298-306.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Fuertes-Olivera, P.A. and H. Bergenholtz. 2018. Diccionario Español de Definiciones. *Círculo de Lingüística Aplicada a la Comunicación* 74: 169-194.
- Fuertes-Olivera, P. A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Giesecke, F.E., A. Mitchell, H. Spencer, I. Hill, J. Dygdon and J. Novak. 2016. *Technical Drawing with Engineering Graphics*: 478. Boston: Prentice Hall.
- Goetsch, D.L., W. Chalk and J.A. Nelson. 2000. *Technical Drawing*. Fourth edition. Albany/Boston: Delmar Publishers.
- González García, V., R. López Poza and M. Nieto Oñate. 1982. *Sistemas de representación. Sistema diédrico*. Valladolid. Texgraf.
- Hallowell, M.R. and J.A. Gambatese. 2010. Qualitative Research: Application of the Delphi Method to CEM Research. *Journal of Construction Engineering and Management* 136(1): 99-107.
- Kageura, K. 1997. Multifaceted/Multidimensional Concept Systems. Wright, S.E. and G. Budin (Eds). 1997. *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*: 119-132. Amsterdam: John Benjamins.
- Linstone, H. and M. Turoff. 1975. *The Delphi Method: Techniques and Applications*. London: Addison-Wesley.
- Mason, J. 1996. *Qualitative Researching*. Thousand Oaks, USA: Sage Publications.

- Mason, J.** 2002. Linking Qualitative and Quantitative Data Analysis. Bryman, A. and R.G. Burgess (Eds.). 2002. *Analyzing Qualitative Data*: 103-124. London: Routledge.
- Nieto Oñate, M., V. González García and R. López Poza.** 1970. *Prácticas programadas de Dibujo Técnico. Delineantes industriales*. Valladolid: DITEC.
- Pérez Hernández, C., A. Moreno Ortiz and P. Faber Benítez.** 1999. Lexicografía computacional y lexicografía de corpus. *Revista Española de Lingüística Aplicada* 1: 175-214.
- Pinnis, M., T. Gornostay, R. Skadiņš and A. Vasiljevs.** 2013. Online Platform for Extracting, Managing, and Utilising Multilingual Terminology. Kosem, I. et al. 2013. *Electronic Lexicography in the 21st Century: Thinking outside the Paper. Proceedings of the Third Biennial Conference on Electronic Lexicography, eLex 2013, Tallin, Estonia, 17–19 October, 2013*: 122-131. Ljubljana/ Tallinn: Institute for Applied Slovene Studies/ Eesti Keele Instituut.
- Powell, C.** 2003. The Delphi Technique: Myths and Realities. *Journal of Advanced Nursing* 41(4): 376-382.
- Rowe, G. and G. Wright.** 1999. The Delphi Technique as a Forecasting Tool: Issues and Analysis. *International Journal of Forecasting* 15(4): 353-375.
- Sadowski, M.A. and S.A. Sorby.** 2015. *Engineering Graphics Concepts: A Delphi Study*. Paper presented at 2015 ASEE Annual Conference & Exposition, Seattle, Washington, 14–17 June 2015. DOI: 10.18260/p.23965.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.
- Tarp, S.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.) 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopaedias Today, Lexicographical Tools Tomorrow*: 17-32. Bern: Peter Lang.
- Verlinde, S. and T. Selva.** 2009. Corpus-based versus Intuition-based Lexicography: Defining a Word List for a French Learners' Dictionary. *Coach* 83: 1424.

Appendix 1: Results of the first round of the Delphi method including the degree of consensus

Round I RESULTS				
Identified Concept	Importance* given (1-5)	Consensus* *	Difficulty*** appreciated (1-5)	
1	<i>Croquizar</i>	4.5	Good	2.8
2	<i>Posición relativa de las vistas</i>	4.71	Good	3.2
3	<i>Mínimas vistas necesarias</i>	4.66	Good	4.2
4	<i>Sección</i>	4.6	Full	3.6
5	<i>Sección total</i>	4.75	Very good	3.25
6	<i>Media sección</i>	4	Good	3.5
	<i>Media sección con giro</i>	4	Very good	4
7	<i>Sección en ángulo</i>	4.4	Full	3.6
8	<i>Sección múltiple abatida</i>	4.4	Full	3.2
9	<i>Secciones auxiliares</i>	4	Good	3.6
10	<i>Planos no paralelos a los de proyección</i>	4.66	Good	3.5
11	<i>Escala</i>	4.42	Good	3.66
12	<i>Seis vistas normalizadas</i>	4.25	Very good	2.75
13	<i>Croquis de la perspectiva</i>	4.33	Good	3.66
14	<i>Medio corte</i>	4,13	Full	3
15	<i>Corte quebrado</i>	4.5	Very good	4
16	<i>Identificar el plano de corte</i>	4.33	Good	3.1
17	<i>Sección abatida</i>	4.2	Full	3.6
18	<i>Acotar</i>	5	Full	4.5
19	<i>Seguir las normas</i>	5	Full	4.4
20	<i>Rosca métrica</i>	4.5	Very good	3.5
21	<i>Coefficiente de reducción</i>	4.33	Good	4
22	<i>Croquis de definición</i>	4.33	Good	3
23	<i>Rosca métrica de paso fino</i>	4	Very good	4
24	<i>Croquis de producto terminado</i>	4.75	Very good	3.25
25	<i>3/8 de pulgada</i>	4	Good	3.6
26	<i>Denominación de las vistas</i>	4.4	Full	2.6
27	<i>Método de proyección del primer diedro</i>	4.8	Full	2.8
28	<i>Indicación del método de proyección</i>	4.2	Full	2.2
29	<i>Elección de las vistas</i>	4.8	Full	3.6
30	<i>Vistas parciales</i>	4	Full	2.4
31	<i>Línea llena gruesa</i>	4.8	Full	2.2
32	<i>Línea llena fina</i>	4.4	Full	2.4
33	<i>Línea llena fina a mano alzada</i>	4	Full	2.2
34	<i>Línea fina de trazos</i>	4.2	Full	2.4
35	<i>Línea fina de trazos y puntos</i>	4.4	Full	2.4
36	<i>Traza del plano de corte</i>	4.6	Full	3.2
37	<i>Anchura de la línea</i>	4.6	Full	2.6
38	<i>Orden de prioridad de líneas coincidentes</i>	4.8	Full	3.8
39	<i>Líneas de referencia</i>	4	Full	2.4
40	<i>Rayado</i>	4.6	Full	2.8
41	<i>Rayado de diferentes partes cortadas de una misma pieza</i>	4.4	Full	3.2

42	<i>Interrupción de rayado</i>	4.2	Full	2.2
43	<i>Corte</i>	4.8	Full	3.4
44	<i>Plano de corte</i>	4.6	Full	3.6
45	<i>Disposición de los cortes</i>	4.6	Full	3.8
46	<i>Localización del plano de corte</i>	4.6	Full	3.4
47	<i>Corte de nervios, elementos de fijación, árboles, radios de ruedas y otros elementos análogos</i>	4.4	Full	3.6
48	<i>Corte por un plano</i>	4.4	Full	2.4
49	<i>Corte por dos planos paralelos</i>	4.2	Full	3
50	<i>Corte por planos sucesivos</i>	4.4	Full	3.2
51	<i>Corte por dos planos concurrentes</i>	4.4	Full	3.4
52	<i>Corte longitudinal de una forma de revolución que contiene detalles regularmente repartidos y no situados en el plano de corte, y siempre que no se produzca ambigüedad</i>	4.8	Full	3.5
53	<i>Secciones abatidas con desplazamiento</i>	4	Full	3.4
54	<i>Secciones abatidas sin desplazamiento</i>	4	Full	2.8
55	<i>Corte local</i>	4	Full	2.6
56	<i>Disposición de las secciones sucesivas</i>	4.4	Full	3
57	<i>Vistas interrumpidas</i>	4	Full	2.6
58	<i>Representación de elementos repetitivos</i>	4.2	Full	2.8
59	<i>Detalles representados a escala mayor</i>	4	Full	2.4
60	<i>Mínima distancia entre dos rectas que se cruzan</i>	4.3	Full	3.4
61	<i>Recta perpendicular a un plano</i>	4.4	Full	3.6
62	<i>Teorema de las tres perpendiculares</i>	4.6	Full	4.6
63	<i>Segmento mínima distancia</i>	4.2	Full	3.6
64	<i>Segmento paralelo a un plano coordenado</i>	4.6	Full	2.6
65	<i>Verdadera magnitud de un segmento a partir de sus proyecciones</i>	4.8	Full	3.2
66	<i>Ángulo recta-plano</i>	4	Full	3
67	<i>Rectas horizontales de plano</i>	4.8	Full	3
68	<i>Rectas frontales de plano</i>	4.8	Full	3
69	<i>Línea de máxima pendiente de un plano</i>	4	Full	4.4
70	<i>Línea de máxima inclinación de un plano</i>	4.2	Full	4.6

* The Importance given (from 1 to 5) is the criterion for belonging to the initial lemma list.

** The Consensus is considered if at least three panelists identify that concept

*** The Difficulty appreciated (from 1 to 5) permits a comparison of two points of view: those of experts and students.

Appendix 2: Results of the second round of the Delphi method

Round II RESULTS		
Concept		Importance given (1-5)
1	<i>Vistas suficientes / Mínimas vistas necesarias / Vistas necesarias para definir ...</i>	4,8
2	<i>Acotar</i>	5
3	<i>Seguir las normas / Normalizado</i>	4,8
4	<i>Mecanizar superficies</i>	3,6
5	<i>Denominación de las vistas</i>	4,4
6	<i>Elección de las vistas</i>	4,8
7	<i>Línea llena gruesa</i>	4,4
8	<i>Orden de prioridad de líneas coincidentes</i>	4
9	<i>Corte / corte por un plano</i>	4,4
10	<i>Corte longitudinal de una forma de revolución que contiene detalles regularmente repartidos y no situados en el plano de corte, y siempre que no se produzca ambigüedad</i>	3,8
11	<i>Verdadera magnitud de un segmento a partir de sus proyecciones</i>	4,4
12	<i>Rectas horizontales de plano</i>	4,8
13	<i>Rectas frontales de plano</i>	4,8
14	<i>Corte total / Corte por un plano</i>	4,4
15	<i>Croquis de producto terminado</i>	4,4
16	<i>Posición relativa de las vistas</i>	4,6
17	<i>Planos no paralelos a los de proyección</i>	4
18	<i>Sección</i>	3,8
19	<i>Traza del plano de corte</i>	4,2
20	<i>Anchura de la línea</i>	3,8
21	<i>Rayado</i>	4
22	<i>Plano de corte</i>	4,2
23	<i>Disposición de los cortes</i>	4,4
24	<i>Localización del plano de corte</i>	4
25	<i>Teorema de las tres perpendiculares</i>	4,6
26	<i>Segmento paralelo a un plano coordenado</i>	4,4
27	<i>Croquizar</i>	4,2
28	<i>Disposición normalizada de las vistas / Seis vistas normalizadas</i>	4,2
29	<i>Corte quebrado / Corte por planos sucesivos</i>	4,2
30	<i>Rosca métrica</i>	4,2
31	<i>Escala</i>	4,8
32	<i>Corte en ángulo / Corte por dos planos concurrentes</i>	4,3
33	<i>Sección múltiple abatida / Disposición de las secciones sucesivas</i>	4
34	<i>Método de proyección del primer diedro</i>	4,6
35	<i>Línea llena fina</i>	3,8
36	<i>Línea fina de trazos y puntos</i>	4,4
37	<i>Rayado de diferentes partes cortadas de una misma pieza</i>	4,4
38	<i>Corte de nervios, elementos de fijación, árboles, radios de ruedas y otros elementos análogos</i>	4,2
39	<i>Mínima distancia entre dos rectas que se cruzan</i>	3
40	<i>Recta perpendicular a un plano</i>	4,4
41	<i>Croquis de la perspectiva</i>	3,8
42	<i>Identificar el plano de corte</i>	4,2
43	<i>Coefficiente de reducción</i>	3,6
44	<i>Croquis de definición</i>	5

45	<i>Sección abatida / sección transversal</i>	4
46	<i>Indicación del método de proyección</i>	3,4
47	<i>Línea fina de trazos</i>	4
48	<i>Interrupción de rayado</i>	3,4
49	<i>Corte por dos planos paralelos</i>	3,6
50	<i>Representación de elementos repetitivos</i>	3,8
51	<i>Segmento mínima distancia</i>	3,8
52	<i>Línea de máxima inclinación de un plano</i>	4
53	<i>Medio corte o corte al cuarto</i>	4,6
54	<i>Plano de proyección</i>	4,4
55	<i>Plano de referencia</i>	3,8
56	<i>Plano Datum</i>	3,5
57	<i>Cota de situación</i>	4,4
58	<i>Ortogonal, perpendicular, normal</i>	4,6
59	<i>Plano representado por una línea, plano proyectante</i>	4
60	<i>Dirección y sentido de la visual / Línea (dirección) de observación</i>	4,8
61	<i>Verdadera forma</i>	5
62	<i>Verdadera dimensión-magnitud</i>	5
63	<i>Elemento de revolución</i>	4
64	<i>Alineamiento de las vistas</i>	5
65	<i>Perspectiva estallada</i>	5
66	<i>Proyección cilíndrica ortogonal / vistas ortogonales</i>	4,8
67	<i>Oblicuo / NO ortogonal</i>	3,6

Appendix 3: Comparative summary of the importance perceived by the experts in round I and round II

Concept	Importance in round I	Importance in round II
1 <i>Vistas suficientes / Mínimas vistas necesarias / Vistas necesarias para definir ...</i>	4,8	4,8
2 <i>Acotar</i>	5	5
3 <i>Seguir las normas / Normalizado</i>	5	4,8
5 <i>Denominación de las vistas</i>	4,8	4,4
6 <i>Elección de las vistas</i>	4,8	4,8
7 <i>Línea llena gruesa</i>	4,8	4,4
8 <i>Orden de prioridad de líneas coincidentes</i>	4,8	4
9 <i>Corte / corte por un plano</i>	4,3	4,4
10 <i>Corte longitudinal de una forma de revolución que contiene detalles regularmente repartidos y no situados en el plano de corte, y siempre que no se produzca ambigüedad</i>	4,8	3,8
11 <i>Verdadera magnitud de un segmento a partir de sus proyecciones</i>	4,8	4,4
12 <i>Rectas horizontales de plano</i>	4,8	4,8
13 <i>Rectas frontales de plano</i>	4,8	4,8
15 <i>Croquis de producto terminado</i>	4,75	4,4
16 <i>Posición relativa de las vistas</i>	4,71	4,6
18 <i>Sección</i>	4,6	3,8
19 <i>Traza del plano de corte</i>	4,6	4,2
20 <i>Anchura de la línea</i>	4,6	3,8
21 <i>Rayado</i>	4,6	4
22 <i>Plano de corte</i>	4,6	4,2
23 <i>Disposición de los cortes</i>	4,6	4,4
24 <i>Localización del plano de corte</i>	4,6	4
25 <i>Teorema de las tres perpendiculares</i>	4,6	4,6
26 <i>Segmento paralelo a un plano coordenado</i>	4,6	4,4
27 <i>Croquizar</i>	4,5	4,2
28 <i>Disposición normalizada de las vistas / Seis vistas normalizadas</i>	4,25	4,2
29 <i>Corte quebrado / Corte por planos sucesivos</i>	4,45	4,2
30 <i>Rosca métrica</i>	4,5	4,2
31 <i>Escala</i>	4,42	4,8
32 <i>Corte en ángulo / Corte por dos planos concurrentes</i>	4,4	4,3
33 <i>Sección múltiple abatida / Disposición de las secciones sucesivas</i>	4,4	4
34 <i>Método de proyección del primer diedro</i>	4,4	4,6
35 <i>Línea llena fina</i>	4,4	3,8
36 <i>Línea fina de trazos y puntos</i>	4,4	4,4
37 <i>Rayado de diferentes partes cortadas de una misma pieza</i>	4,4	4,4
38 <i>Corte de nervios, elementos de fijación, árboles, radios de ruedas y otros elementos análogos</i>	4,4	4,2
39 <i>Mínima distancia entre dos rectas que se cruzan</i>	4,3	3
40 <i>Recta perpendicular a un plano</i>	4,4	4,4

42	<i>Identificar el plano de corte</i>	4,33	4,2
46	<i>Sección abatida / sección transversal</i>	4,2	4
47	<i>Indicación del método de proyección</i>	4,2	3,4
48	<i>Línea fina de trazos</i>	4,2	4
49	<i>Interrupción de rayado</i>	4,2	3,4
50	<i>Corte por dos planos paralelos</i>	4,2	3,6
51	<i>Representación de elementos repetitivos</i>	4,2	3,8
52	<i>Segmento mínima distancia</i>	4,2	3,8
53	<i>Línea de máxima inclinación de un plano</i>	4,2	4

Appendix 4: Results of students' opinions

The students quantify the difficulty of the Delphi second round concepts

	Concepts	Difficulty (1-5)
1	<i>Vistas suficientes / Mínimas vistas necesarias / Vistas necesarias para definir</i>	2,70
2	<i>Acotar</i>	2,80
3	<i>Seguir las normas / Normalizado</i>	3,13
4	<i>Mecanizar superficies</i>	3,22
5	<i>Denominación de las vistas</i>	2,10
6	<i>Elección de las vistas</i>	2,93
7	<i>Línea llena gruesa</i>	2,20
8	<i>Orden de prioridad de líneas coincidentes</i>	2,83
9	<i>Corte</i>	3,20
10	<i>Corte longitudinal de una forma de revolución que contiene detalles regularmente repartidos y no situados en el plano de corte; y siempre que no se produzca ambigüedad</i>	3,34
11	<i>Verdadera magnitud de un segmento a partir de sus proyecciones</i>	2,63
12	<i>Rectas horizontales de plano</i>	2,37
13	<i>Rectas frontales de plano</i>	2,33
14	<i>Corte total / Corte por un plano</i>	2,25
15	<i>Croquis de producto terminado</i>	2,79
16	<i>Posición relativa de las vistas</i>	2,35
17	<i>Planos no paralelos a los de proyección</i>	3,19
18	<i>Sección</i>	3,06
19	<i>Traza del plano de corte</i>	2,64
20	<i>Anchura de la línea</i>	2,32
21	<i>Rayado</i>	1,65
22	<i>Plano de corte</i>	2,49
23	<i>Disposición de los cortes</i>	2,78
24	<i>Localización del plano de corte</i>	1,94
25	<i>Teorema de las tres perpendiculares</i>	3,21
26	<i>Segmento paralelo a un plano coordenado</i>	2,99
27	<i>Croquizar</i>	2,40
28	<i>Disposición normalizada</i>	2,67
29	<i>Corte quebrado / Corte por planos sucesivos</i>	3,02
30	<i>Rosca métrica</i>	3,11
31	<i>Escala</i>	2,51
32	<i>Corte en ángulo / Corte por dos planos concurrentes</i>	3,29
33	<i>Sección múltiple abatida / Disposición de las secciones sucesivas</i>	3,43
34	<i>Método de proyección del primer diedro</i>	2,15
35	<i>Línea llena fina</i>	2,12
36	<i>Línea fina de trazos y puntos</i>	1,86
37	<i>Rayado de diferentes partes cortadas de una misma pieza</i>	2,14
38	<i>Corte de nervios; elementos de fijación; árboles; radios de ruedas y otros elementos análogos</i>	2,60
39	<i>Mínima distancia entre dos rectas que se cruzan</i>	2,93
40	<i>Recta perpendicular a un plano</i>	2,46
41	<i>Croquis de la perspectiva</i>	2,52
42	<i>Identificar el plano de corte</i>	2,65

43	<i>Coefficiente de reducción</i>	2,76
44	<i>Croquis de definición</i>	2,62
45	<i>Sección abatida / sección transversal</i>	3,08
46	<i>Indicación del método de proyección</i>	2,90
47	<i>Línea fina de trazos</i>	2,12
48	<i>Interrupción de rayado</i>	2,57
49	<i>Corte por dos planos paralelos</i>	2,25
50	<i>Representación de elementos repetitivos</i>	2,57
51	<i>Segmento mínima distancia</i>	3,23
52	<i>Línea de máxima inclinación de un plano</i>	3,07
53	<i>Medio corte o corte al cuarto</i>	2,35
54	<i>Plano de proyección</i>	2,45
55	<i>Plano de referencia</i>	2,70
56	<i>Plano Datum (de medidas)</i>	3,45
57	<i>Cota de situación</i>	3,05
58	<i>Ortogonal; perpendicular; normal</i>	2,52
59	<i>Plano representado por una línea; plano proyectante</i>	2,87
60	<i>Dirección y sentido de la visual / Línea (dirección) de observación</i>	2,80
61	<i>Verdadera forma</i>	2,83
62	<i>Verdadera dimensión</i>	3,02
63	<i>Elemento de revolución</i>	2,78
64	<i>Alineamiento de las vistas</i>	2,59
65	<i>Perspectiva estallada</i>	3,36
66	<i>Proyección cilíndrica ortogonal / vistas ortogonales</i>	3,00
67	<i>Oblicuo / NO ortogonal</i>	3,28

Appendix 5: Comparison of the students' perceived difficulty and that of the experts

Concepts	Students difficulty	Experts difficulty
1 <i>Vistas suficientes / Mínimas vistas necesarias / Vistas necesarias para definir</i>	2,70	4,2
2 <i>Acotar</i>	2,80	4,5
3 <i>Seguir las normas / Normalizado</i>	3,13	4,4
4 <i>Denominación de las vistas</i>	2,10	2,6
5 <i>Elección de las vistas</i>	2,93	3,6
6 <i>Línea llena gruesa</i>	2,20	2,2
7 <i>Orden de prioridad de líneas coincidentes</i>	2,83	3,8
8 <i>Corte</i>	3,20	3,4
9 <i>Corte longitudinal de una forma de revolución que contiene detalles regularmente repartidos y no situados en el plano de corte; y siempre que no se produzca ambigüedad</i>	3,34	3,5
10 <i>Verdadera magnitud de un segmento a partir de sus proyecciones</i>	2,63	3,2
11 <i>Rectas horizontales de plano</i>	2,37	3
12 <i>Rectas frontales de plano</i>	2,33	3
13 <i>Corte total / Corte por un plano</i>	2,25	2,4
14 <i>Croquis de producto terminado</i>	2,79	3,25
15 <i>Posición relativa de las vistas</i>	2,35	3,2
16 <i>Planos no paralelos a los de proyección</i>	3,19	3,5
17 <i>Sección</i>	3,06	3,6
18 <i>Traza del plano de corte</i>	2,64	3,2
19 <i>Anchura de la línea</i>	2,32	2,6
20 <i>Rayado</i>	1,65	2,8
21 <i>Plano de corte</i>	2,49	3,6
22 <i>Disposición de los cortes</i>	2,78	3,8
23 <i>Localización del plano de corte</i>	1,94	3,4
24 <i>Teorema de las tres perpendiculares</i>	3,21	4,6
25 <i>Segmento paralelo a un plano coordenado</i>	2,99	2,6
26 <i>Croquizar</i>	2,40	2,8
27 <i>Disposición normalizada</i>	2,67	2,75
28 <i>Corte quebrado / Corte por planos sucesivos</i>	3,02	3,6
29 <i>Rosca métrica</i>	3,11	3,5
30 <i>Escala</i>	2,51	3,66
31 <i>Corte en ángulo / Corte por dos planos concurrentes</i>	3,29	3,4
32 <i>Sección múltiple abatida / Disposición de las secciones sucesivas</i>	3,43	3,1
33 <i>Método de proyección del primer diedro</i>	2,15	2,8
34 <i>Línea llena fina</i>	2,12	2,4
35 <i>Línea fina de trazos y puntos</i>	1,86	2,4
36 <i>Rayado de diferentes partes cortadas de una misma pieza</i>	2,14	3,2
37 <i>Corte de nervios; elementos de fijación; árboles; radios de ruedas y otros elementos análogos</i>	2,60	3,6
38 <i>Mínima distancia entre dos rectas que se cruzan</i>	2,93	3,4
39 <i>Recta perpendicular a un plano</i>	2,46	3,6
40 <i>Croquis de la perspectiva</i>	2,52	3,66
41 <i>Identificar el plano de corte</i>	2,65	3,1
42 <i>Coefficiente de reducción</i>	2,76	4
43 <i>Croquis de definición</i>	2,62	3
44 <i>Sección abatida / sección transversal</i>	3,08	3,6

45 <i>Indicación del método de proyección</i>	2,90	2,2
46 <i>Línea fina de trazos</i>	2,12	2,4
47 <i>Interrupción de rayado</i>	2,57	2,2
48 <i>Corte por dos planos paralelos</i>	2,25	3
49 <i>Representación de elementos repetitivos</i>	2,57	2,8
50 <i>Segmento mínima distancia</i>	3,23	3,6
51 <i>Línea de máxima inclinación de un plano</i>	3,07	4,6
52 <i>Medio corte o corte al cuarto</i>	2,35	3

Integrating Terminological Resources in Dictionary Portals: The Case of the *Diccionarios Valladolid-UVa*

Pedro A. Fuertes-Olivera, *Department of Afrikaans and Dutch, University of Stellenbosch, South Africa and International Centre for Lexicography, Universidad de Valladolid, Valladolid, Spain (pedro@emp.uva.es)*

and

M.A. Esandi-Baztan, *Department of Materials Science and Metallurgical Engineering, Graphic Expression in Engineering, Cartographic Engineering, Geodesy and Photogrammetry, Mechanical Engineering and Manufacturing Process Engineering, Universidad de Valladolid, Spain (mariaangeles.esandi@uva.es)*

Abstract: This paper advocates the convergence of terminology and lexicography, and illustrates this view by presenting some of the steps taken for incorporating terminological resources and ideas in an online dictionary portal that is being constructed at the University of Valladolid (Spain). This dictionary portal contains several dictionary types, was designed by the same team and is being constructed from the same theoretical perspective, regardless of whether some of the lexical items included are judged "lexicographic", i.e. related to *general language expressions*, or "terminological", i.e. connected with *terms*. In addition to dealing with certain basic tenets of dictionary portals, the paper describes an ad-hoc typology of *definitions* that has been created for two main reasons. Firstly, it makes the process of compilation easier, more uniform, and more readily systematised, thus facilitating the efforts of different people in different places at different times. Secondly, these definitions will feed the Spanish-English *Write Assistant*, a commercially driven language tool that uses a language module based on statistics and is in the process of using Artificial Intelligence (AI) technologies, e.g. machine learning and neural networks, for creating patterns. We have found that precise definitions, similar to terminological (i.e. encyclopaedic) definitions, for most lemmas increase the tool's functions. Such definitions offer a very different picture of current monolingual Spanish and bilingual Spanish-English dictionaries.

Keywords: DICTIONARY PORTAL, ONLINE DICTIONARIES, DEFINITION, ENGLISH, SPANISH, CONVERGENCE OF TERMINOLOGY AND LEXICOGRAPHY, AI TECHNOLOGIES

Opsomming: Die integrasie van terminologiese hulpbronne in woordeboekportale: Die geval van die *Diccionarios Valladolid-UVa*. In hierdie artikel word die konvergensie van terminologie en leksikografie bepleit, en hierdie siening word geïllustreer deur sommige van die stappe vir die inorporering van terminologiese hulpbronne en idees in 'n aanlyn woordeboekportaal wat by die Universiteit van Valladolid (Spanje) saamgestel word,

voor te lê. Hierdie woordeboekportaal bevat verskeie woordeboektippe, is deur dieselfde span ontwerp en word vanuit dieselfde teoretiese perspektief saamgestel, ongeag of sommige van die leksikale items wat ingesluit word as "leksikografies" beskou word, m.a.w. verwant aan *algemene taaluitdrukkings*, of as "terminologies", m.a.w. verwant aan *terme*. Benewens die hantering van sekere basiese woordeboekportaalbeginsels, beskryf die artikel ook 'n ad-hoc-tipologie van *definisies* wat hoofsaaklik om twee redes geskep is. Eerstens vergemaklik dit die samestellingsproses en maak dit ook eenvormiger en makliker om te sistematiseer. Sodoende word die pogings van verskillende mense op verskillende plekke en tye gefasiliteer. Tweedens sal hierdie definisies as bron dien vir die Spaans-Engelse *Write Assistant*, 'n kommersieelgedrewe taalhulpmiddel wat gebruik maak van 'n taalmodule wat op statistiek berus en wat Kunsmatige Intelligensie (KI)-tegnologie, bv. masjienleer en neurale netwerke, inspan vir die skep van patrone. Ons het bevind dat presiese definisies, soortgelyk aan terminologiese (m.a.w. ensiklopediese) definisies, vir die meeste lemmas die funksies van die hulpmiddels uitbrei. Sulke definisies bied 'n heeltemal ander blik op bestaande eentalig Spaanse en tweetalig Spaans-Engelse woordeboeke.

Sleutelwoorde: WOORDEBOEKPORTAAL, AANLYN WOORDEBOEKE, DEFINISIE, ENGELS, SPAANS, KONVERGENSIE VAN TERMINOLOGIE EN LEKSIKOGRAFIE, KI-TEGNOLOGIE

1. Introduction

Terminology may be defined as "the study of and field of activity concerned with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialized areas of usage of one or more languages" (Sager: 1990: 2). This view connects terminology with lexicography, a process that the coming of age of the Internet, the preponderance of the user paradigm, and the influence of corpus methodology has accelerated. Bowker (2018: 138) summarises this approach, which has been well attested since the 1990s, by commenting on Sager's suggestion that "in its objectives, terminology is akin to lexicography which combines the double aim of collecting information about the lexicon of a language with providing an information — and sometimes even an advisory — service to language users." She (2018: 147) documents her stance with existing practices in both camps, in which we can observe that dictionaries are incorporating terms and that terminological resources are adding general language expressions:

Both lexicographers and terminologists have indeed taken up this challenge to produce resources that meet user needs. For instance, in response to user requests, the second edition of the *Macmillan English Dictionary* (2007) now incorporates a range of specialist terms along with its general language offering. Meanwhile, term bases for private corporations, as well as those developed for use by translators, now regularly include general language expressions alongside specialized terms.

Bowker (2018: 147) concludes her analysis on the relationship between both disciplines by claiming that "although some differences remain, the two disciplines nonetheless appear to be converging with regard to many aspects of their practice." Table 1 shows how she views the evolution and convergence of characteristics of lexicography and terminology:

	Lexicography	Terminology
Practitioner	mainly lexicographers, but with greater involvement from the general public (via crowdsourcing)	mainly terminologists, but with greater involvement from the general public and subject matter experts (via open and closed crowdsourcing)
Object of study	mainly words, but also some terms	mainly terms, but also some general language words or expressions
Domain	mainly general language, but also some specialised language	mainly the language of a specialised domain, but also some general language
Point of view	mainly descriptive	mainly normative/prescriptive in the public and academic sectors, but incorporating more descriptive elements in commercial settings
Approach	mainly semasiological	increasingly semasiological, but retaining some onomasiological elements where useful
Organisation	mainly alphabetical, but sometimes incorporating thematic elements	mainly thematic, but allowing alphabetic searching
Main information provided	words, meanings, examples, usage information (e.g. collocations, frequency, phraseology), a range of linguistic information (e.g. part of speech, pronunciation)	preferred term, variants, context and usage information (e.g. collocations, frequency, phraseology), meaning, conceptual relations
Intended users	lay people, professional and academic audiences	public sector (for language planning), domain experts, scientific/technical writers, translators (for bi- or multilingual resources), commercial enterprises

Table 1: Evolution and convergence of characteristics associated with lexicography and terminology. Source: Bowker (2018: 148)

This paper elaborates on Bowker's main conclusions and adds several reasons for advocating the integration of lexicography and terminology. Firstly, instead of single dictionaries, we have to design and construct dictionary portals, where traditionally associated differences between lexicography and terminology can be downplayed or even eliminated (Section 2). Secondly, integration will work better if the dictionary portal contains *up-market online resources*, as these will be based on the use of *disruptive technologies* that are necessary for designing and constructing commercially-driven language tools. Thirdly, the nature of these tools is not affected by the data type they contain — i.e. whether we work with *words* or *terms* — but by the philosophy underlying them and the technical and technological resources used for dealing with their computational and linguistic aspects (Section 3). Finally, the integration process is illustrated with an ad-hoc typology of *definitions*, i.e. a collection of different types that was created for several reasons. For instance, this typology makes the process of compilation easier, more uniform, and more readily systematised, thus facilitating the efforts of different people in different places and at different times (Section 4). A final conclusion will summarise our main findings and will enumerate certain topics that may merit more attention in future research.

2. Dictionary portals

Engelberg and Müller-Spitzer (2013: 1023) define a dictionary portal as "a data structure that is presented as a page or set of interlinked pages on a computer screen and provides access to a set of electronic dictionaries, and where these dictionaries can also be consulted as standalone products". With the underpinning of several criteria — type of access provided, cross-referencing structures, ownership relations between the portal and the dictionaries and layout of the portal — Engelberg and Müller-Spitzer (2013) propose a typology of dictionary portals comprising (a) dictionary nets, (b) dictionary search engines, and (c) dictionary collections. Their analysis, as well as that of Boelhouver et al. (2018), concludes that dictionary portals are widespread and that they may illustrate the way ahead for future online dictionaries.

The above definition merits some comments. Firstly, a dictionary portal is a "data structure". This can offer users more than existing corpus-based online dictionaries provided that designers of such portals assume that data structures might be analysed with big data analytics. For instance, designers of the project *Diccionarios Valladolid-UVA* (see Section 3, below) have used big data analytics for analysing around 60 million look-ups in existing online dictionaries; the aim was to discover real searches, for which an initial lemma list of 20,000 English words and 16,000 Spanish words was created. This process has comprised several stages and is based on an analysis of around one million daily searches in several dictionaries, e.g. an English–Spanish/Spanish–English dictionary, an English–German/German–English dictionary, an English monolingual dictionary, a Spanish monolingual dictionary, and so on. Around 80% of the searches

can be matched, i.e. the same search is identified in the logfiles of different dictionaries and can, therefore, be interpreted as an identification of the most popular dictionary articles in the dictionaries under scrutiny. After two months of work with the search logfiles, which amounted to more than 60 million look ups, IT professionals working on this project were able to produce the above-mentioned lemma lists (Fuertes-Olivera 2019).

Secondly, the data structure will be "presented as a page or set of inter-linked pages on a computer screen." It is obvious that mobiles and similar devices, e.g. tablets, are also carriers of data structures and that most people view them on these screens rather than on those of computers. This is influencing the design of online dictionaries, which must make provisions for allowing users to consult data on smaller screens than those originally conceived.

Thirdly, it "provides access to a set of electronic dictionaries". This sentence is clearly confusing, as most existing dictionary portals only contain retro-digitised dictionaries, i.e. printed dictionaries converted to a digital format. An analysis of the 37 dictionary portals identified in Boelhouwer et al. (2018: 765-767) has allowed us to classify them into three main types, the criterion being the presence of real online dictionaries.

The *Free Dictionary.com* illustrates the first type. It *mostly* contains data from printed reference works — typically from dictionaries, glossaries or encyclopaedias — and from free access online ones, e.g. *Wikipedia*. For instance, the entry for *cost accounting* in *The Free Dictionary.com* offers the following data:

- Definitions from the *Collins English Dictionary* 12th edition, and the *Random House Kernerman Webster's College Dictionary*.
- The symbol "n" for noun.
- The related words *cost accountant* with the symbol "n" for "noun" and a definition extracted from the *American Heritage® Dictionary of the English Language, fifth edition*, as well as *accountancy*, *accounting* and *costing*.
- An image that connects *cost accounting* with *costing*, *accountancy*, and *accounting*, without explaining the connection.
- A very long encyclopaedic article of *accounting* from the *West's Encyclopedia of American Law, 2nd edition*, published in 2008. In this long entry we are told that various accounting methods are employed and that one of them is the *cost method of accounting*. This article includes a section on "Further readings" and cross-references to *Accrual Basis*; *Cash Basis*; and *Income Tax*.
- A Definition from the *Farlex Financial Dictionary*.
- A Definition from the glossary *Wall Street Words: An A to Z Guide to Investment Terms for Today's Investor*, authored by David L. Scott.
- An encyclopaedic article from the *Great Soviet Encyclopedia, 3rd Edition*, published between 1970 and 1979.
- An encyclopaedic article from *Wikipedia*.

Users searching for *cost accounting* in *The Free Dictionary.com* will face many

problems, three of which are relevant for this paper: (a) they will not be sure about the meaning and usage of *cost accounting* in today's financial world, as they will come across contradictory information, e.g. *cost accounting* is defined in the context of the Soviet Union of 1970, Wall Street and the United States of 2008 (before the Great Recession and after the Enron Failure), and the changes introduced after the Great Recession so as to make future financial crises more difficult; (b) the consultation process takes a lot of time and energy, as users must decide for themselves which, if any, of the examples of contradictory information they find suitable; and (c) there are no usage contexts, e.g. grammatical information, example sentences, and so on.

Lexico.com is an example of the second type (see at <https://www.lexico.com/>). It includes free and restricted areas, the latter for subscribers, and especially for users of *English* and *Spanish dictionaries*. In this portal, users access printed and online information tools. For reasons of space and convenience, we will only analyse the free part of the portal, in which we can look up in English and Spanish online dictionaries as well as a Thesaurus, an English grammar and a Spanish grammar. The dictionaries included in the portal contain features of both online dictionaries, e.g. they do not use abbreviations, and printed dictionaries, e.g. they continue using *recursive definitions* and definitions by synonyms. Example 1 shows the definition of *accrual accounting* in the English Dictionary. This is an example of *recursive definition*, i.e., users must search "accrual basis" and "accounting" in order to understand its meaning:

accrual accounting

The practice of accounting on an accrual basis.

Example 1: Definitions of *accrual accounting* in *Lexico.com*

English is the node language in this dictionary portal and it seems to have considerable influence on how dictionaries are designed, compiled and updated. Having a node language makes design and construction processes cheaper, as data can easily be made reusable, i.e. transferred from, say, a monolingual English dictionary to a bilingual English–Spanish one. The grammar tabs display grammatical information, e.g. definitions of grammatical concepts such as "personal pronoun" and so on. The thesaurus offers synonyms and antonyms under an example sentence.

Searching for *cost accounting* in the free area has allowed us to extract the following information:

- A definition.
- Part of speech (NOUN).
- Grammar information: it is a mass noun.
- Several example sentences, all of which seem to have been extracted from a specialised corpus.

- Pronunciation. Users can hear the pronunciation of the word and read its phonetics.
- Equivalentents in Spanish and, for some lemmas, translations of example sentences in Spanish. For instance, we have the Spanish equivalentents *contabilidad analítica de costes* and *contabilidad analítica de costos* (the former is used in European Spanish and the latter in American Spanish; although important, this information is missing).

Looking up this dictionary portal, users will also face several difficulties, three of which are no doubt problematic for Spanish users:

- the definition of concepts is poor: it lacks reference to concepts, domains and the like. For instance, we are not informed that *cost accounting* is an accounting method;
- users have to infer the context of usage, as the grammar given is rather weak (for instance, the English–Spanish part does not inform that *cost accounting* is uncountable) and usage expressions are scarce or non-existing for some meanings and usages. By way of example, *accounting board* has two meanings in the *English Dictionary*, but all the English sentences and their Spanish translations are on a different interface and are accompanied by the warning that the portal cannot verify their accuracy; and
- the context sentences work as "proxies", i.e. they can or cannot refer to the concept, which force users to decide for themselves whether or not they are suitable.

OWID is an example of the third type. It is a platform "which integrates multiple dictionaries via a common interface, with a joint lemma list and single search option" (Storjohann 2018: 574). What makes this portal different from the other two types is that it *also* offers external resources so as to integrate different academic lexicographic resources with the focus on contemporary German. In *OWID* there is a kind of macrostructure, i.e. a list of alphabet letters at the top of the page that can be clicked to recover the lemmas initiated with the letter chosen. The lemma list is shown on the left of the page and can be modified by clicking on the up and down arrow-like symbols that accompany the lemma list. Searching *Kostenrechnung* (*cost accounting* in German) retrieves the following data:

- Its orthography, including syllabic divisions, etymology and sub-comments on semantics.
- Its grammar and "context profile", i.e. which words typically go with it. This is achieved by cross-referencing users to *CanooNet*, also a German dictionary portal with precise grammar descriptions of each German word, and to *CCCB*, a lexical database that offers contexts, contrasts, lexicographic profiles, semantic proximity and related collocations of each lemma searched for. In other words, it offers lexical maps, i.e. collocational profiles of lemmas.

- Sentence examples taken from the *lexiko corpus*.
- Automatic selection of texts that aim to offer a kind of definition(s) of the lemma.

For the purpose of this paper, *OWID* is interesting as it includes several types of dictionaries with other reference tools. For instance, users can access an online dictionary of contemporary German, a dictionary of Neologism, a collocations dictionary and a "Discourse Dictionary", i.e. a sort of dictionary containing lexical maps. These resources are integrated, i.e. they are not a random collection of unrelated dictionary tools, but a collection of lexical data organised through an innovative concept of data modelling and structuring (Müller-Spitzer and Möhrs 2008). Following this line of work, we also defend integrating lexicography and terminology in the same dictionary portal, as shown below.

3. Integrating terminological resources in up-market dictionary portals: The *Diccionarios Valladolid-UVa*

The *Diccionarios Valladolid-UVa* were initially designed as independent online dictionaries in 2012. However, in 2017 they were re-designed and integrated in a new type of reference tool termed "integrated dictionary portal" and defined as

A reference tool whose Dictionary Writing System is equipped with disruptive technologies. These allow lexicographers to store as much data as possible and users to retrieve only the data they need in specific use situations. Its articles are prepared by the same team with the basic aim of helping human and/or machine users to meet their needs in a quick and easy way. They contain both lexicographically prepared data and linked open data with lexicographic value. The lexicographic data is reusable, subject to a constant process of updating and can be used in conjunction with other tools, e.g. assistants. (Fuertes-Olivera 2016)

At the time of writing this paper, this dictionary portal is still under construction. For this purpose we are using three Dictionary Writing Systems (DWS), each of which is devoted to one particular dictionary type:

- (a) monolingual Spanish dictionaries;
- (b) bilingual English–Spanish/Spanish–English dictionaries; and
- (c) specialised dictionaries.

Each DWS contains an editor, database, Web interface, various management tools, and a kind of dictionary grammar that specifies the structure of the portal, e.g. by making ad-hoc connections of the different elements integrating the DWS (Kilgarriff 2006: 7). They are *currently* storing around 128, 000 general Spanish lemmas, approximately 22, 000 English lemmas, and some 12, 000 English and Spanish specialised lemmas. Around 15,000 more lemmas are added per year and it is expected to go public in 2022. By that time we expect to have finished the lexicographic description of some 75,000 lemmas and to have pre-

pared different access structures and interfaces. We aim to create dynamic access structures, i.e. users will have at their disposal different access possibilities, and easy-to-consult interfaces, e.g. all without publicity. Depending on its possible commercial success, this portal can be increased by more specialised and general dictionaries.

Comparing the *Diccionarios Valladolid-UVa* with the above types of dictionary portals, especially with OWID, there are several important differences; three are relevant for this paper. Firstly, this portal is integrated, which signifies the following:

- (a) that all the dictionaries included in the portal use a common grammar, that is, an abstract representation of dictionary structure, e.g. the same access structure (Koeva and Blagoeva 2013);
- (b) that much of the lexicographic data is reusable, i.e. it can be transferred from one Dictionary Writing System to another;
- (c) that all of them are conceptualised as reference tools (Fuertes-Olivera and Tarp 2014);
- (d) that they will also support and feed *Write Assistant* (see section 5 and Fuertes-Olivera and Tarp 2020), and
- (e) that they *include all members of any clearly-defined sets*, e.g. accounting methods, days of the week, poisonous snakes, non-poisonous snakes, chemical elements, etc. For instance, there are three lemmas for Spanish *pez* (English fish), and one lemma for including all the components of the "fish taxonomy": *pez rata, pez luna, pez rubia, pez sable, pez zorro, pez mujer, pez negra, pez gordo, pez sierra, pez ángel, pez payaso, pez obispo, pez piloto, pez resina, pez griega, pez espada, pez volador, pez reverso, pez machete, pez rémora, pez martillo, pez trompeta, pez mantequilla, pez aguja, pez ballesta, pez clavo, pez cofre, pez de San Pedro, pez limón, pez erizo, pez gato, pez gordo, pez guitarra, pez hacha, pez loro, pez pulmonado, pez rojo de China, pez torpedo, pez vela, estar pez* and *que le folle un pez que la tiene fría*.

The "treatment of *pez* illustrates some of the features of this dictionary portal. *Pez* has three grammars. *Pez* (English *fish*) is an animal that admits singular and plural forms, but *pez* (English *tar*) is also a substance that is used only in the singular and *pez* (English *have no idea*) is an invariable adjective that is typically used in multiword lemmas such as *estar pez*. Other lemmas are included because they refer to things, i.e. concepts. Lemmatising them allows us to include categories and taxonomies, e.g. the Spanish taxonomy *peces* (fish) shown above. All the members of these clearly defined sets are accompanied by their Latin formal names as synonyms and can thus be easily identified in several languages, e.g. in English and in Spanish. Finally, we have also included lemmas referring to facts, e.g. cities of the world, rivers, names of battles, festivities, and so on.

Secondly, each meaning of a lemma can be described with up to 67 different lexicographic data. Examples 2 and 3 show the dictionary article *balance*

sheet in two use situations (users will have more than 28 different use situations at their disposal): a general English–Spanish production situation for Spanish native speakers (example 2), and a specialised English production situation for English native speakers and translators (example 3):

balance sheet UK English

Gramática

Nombre contable, usado en singular y plural.

Flexiones

balance sheet, balance sheets

Nota de uso

La forma "balance-sheet" se usa menos (Ngram Viewer).

Definición

documento financiero que recoge el informe en el que se analiza el activo, o bienes o valores de los que dispone la empresa, y el pasivo, u obligaciones económicas de la misma como deudas, préstamos, etc., y su patrimonio neto, la diferencia entre el activo y el pasivo de la empresa; este análisis permite conocer la situación económica y financiera de una empresa o entidad en un momento determinado; se usa en contabilidad

Sinónimos

balance

balance-sheet

statement of financial position US English

Equivalent

balance de situación

Collocations

a condensed balance sheet

un balance de situación consolidado

account form balance sheet

balance con forma de cuenta

derecognise in the balance sheet

eliminar del balance

(...)

Example

The balance sheet is a statement of the enterprise's assets, equity and liabilities at the balance sheet date.

El balance de situación es un document que registra los activos, pasivos y patrimonio de la empresa en la fecha en la que se formula el balance.

Formations (compuestos y derivaciones)

balance sheet account

cuenta de balance

balance sheet at year-ends
fecha de cierre de balance
balance sheet date
fecha de cierre de balance
balance sheet entry
entrada en el balance
balance sheet format
formato de balance
balance sheet liability method
método de la deuda basado en el balance
balance sheet sum
suma del balance
balance sheet total
total del balance
balance sheet value
valor del balance
(...)

Example 2: The entry *balance sheet* in the *Diccionarios Valladolid-UVa: Production in an English–Spanish general situation* (English) (excerpts)

balance sheet UK English

noun

Inflections

A balance sheet, the balance sheet, balance sheets

Definition

The balance sheet is a statement of the enterprise's assets, equity and liabilities at the balance sheet date. The statement is a status report estimating the enterprise's assets, equity and liabilities as a snapshot at a certain date.

Synonyms

Balance

Statement of financial position US English

Collocations

- a comparative balance sheet as of the end of the immediately preceding financial year
- balance sheet amounts
- balance sheet at the end of the current interim period
- balance sheet at year-end
- balance-sheet layout
- disclose on the face of the balance sheet
- (...) (twenty more collocations, i.e. unfinished sentence examples)

Examples

A vertical balance sheet is one in which the balance sheet presentation format is a single column of numbers, beginning with asset line items, followed by liability line items, and ending with shareholders' equity line items.

Example 3: The entry *balance sheet* in the *Diccionarios Valladolid-UVA*: a specialised English production situation (excerpts)

The above examples 2 and 3 represent only 7.4% of the presentation possibilities available at the time of writing this paper (there will be around 28 different possibilities in the *Diccionarios Valladolid-UVA*). For reasons of space, suffice it to say that each of these is based on technological and lexicographic decisions (Fuertes-Olivera, Tarp and Sepstrup 2018), and together they account for the third important feature of this project. Technological and lexicographic aspects are based on the concept of *disruptive innovation* espoused in business theory and practice. Disruptive innovation was introduced in 1995 and has proved a powerful way of thinking about innovation-driven growth, insofar as disruption "describes a process whereby a smaller company with fewer resources is able to successfully challenge established incumbent businesses" (Christensen et al. 2015).

In our case, we have prepared, designed and compiled up-market reference tools, i.e. tools that aim at displacing established competitors by offering a whole new population of consumers at the base of a market access to a product or service that was only accessible to consumers with a lot of money or a lot of skill (Christensen 2011). For instance, preparing *dynamic dictionary articles*, i.e. different data for different users in different situations (examples 2 and 3 above), is a feature of up-market online dictionaries that can easily be implemented as a strategy for broadening the customer base of online dictionaries. Secondly, we have prepared our dictionaries as part of the "data-driven economy", e.g. we are using many sources and a large amount of data (Fuertes-Olivera 2015). An example of this is the type of definitions used in our general dictionaries. On average, they contain around 75 words per definition and resemble *terminological definitions* (see example 2). They are crafted following an ad-hoc typology that has been engineered with two main aims, namely, to help lexicographers in their daily work and to facilitate the creation of patterns that can be displayed by using AI technologies, especially machine learning and neural networks (see below).

4. Definitions

The concept of definition has been the subject of scrutiny in different fields, e.g. Philosophy, Logic, Law, Linguistics, Terminology and Lexicography. For the purpose of this paper, definitions describe the meaning of the lemma, i.e. the

"set of conditions which must be satisfied by a lexical unit in order to denote the extralinguistic reality/-ies which correspond(s) to each of its senses" (Fuertes-Olivera and Arribas-Baño 2008: 69). Hence, they refer to the "specific set of data that explains the meaning of a lemma and which is clearly addressed to the lemma" (Nielsen 2011: 202).

Rundell (2015: 314) indicates that, in the printed era, a focus on economy led to definitions "which achieve conciseness (and aspire to precision) through the use of standard formulae ("the act of X-ing), "characterised by Y", and so on) and through a "recursive" strategy." These strategies have costs which are passed on to the user, who has to learn these conventions in order to understand what the dictionary is saying. He adds that in the last 30 years publishers, and especially those in the UK, have addressed this issue by developing more open defining styles. These aim to offer enough information for understanding the definition without knowledge of *dictionarese*, i.e. the typical dictionary conventions such as *recursiveness*.

After reviewing the literature on definitions, Fuertes-Olivera and Arribas-Baño (2008: 70) report that so-called *lexical definitions*, *conceptual definitions*, *relational definitions*, *definitions by extension*, *definition by intension*, *partitive definitions*, and *encyclopaedic definitions* show only a few formal differences among the defining styles of the specialised dictionaries they study. In the *Diccionarios Valladolid-UVa* we have focused on these differences with the aim of constructing an ad-hoc typology that assumes the tenets of the so-called *integrationist approach* (Harris and Hutton 2007). The basic assumptions here are that all signs are semantically indeterminate, that meanings are lexicographers' or terminologists' creations, and that these rely on precise and specific contexts:

A lexicographical definition (...) does not identify a meaning independently existing in actual usage and *discovered* there by the lexicographer: it is deliberately *constructed* and *allocated* by the lexicographer on the basis of materials selected for study, and its allocation will depend on the viewpoint the lexicographer has chosen to adopt. (Harris and Hutton 2007: 78)

and

A definition can only be as effective as the context allows it to be, and the context includes the situation of the person seeking to understand the meaning. The notion of a definition adequate to all occasions and all demands is a semantic *ignis fatuus*. (Harris and Hutton 2007: 49)

Our typology comprises 8 types of definitions:

1. *Specialised definitions*: those which describe the meaning of abstract concepts; basically speaking, ideas and thoughts that are not part of the material world but human constructions used in specialised texts for denoting the basics of a particular domain. We use several full sentences that go from more general to more specific aspects. With these definitions we always *include* example sentences that help one to understand the concept.

All such definitions are included in the specialised dictionaries of the portal, crafted by experts in the field, and *mainly* target human users, e.g. experts and experienced translators (example 4):

coste atribuido (English: deemed cost)

Definición

El coste atribuido es el importe usado como subrogado del coste o del coste depreciado en una fecha determinada. En la depreciación o amortización posterior se supone que la entidad había reconocido inicialmente el activo o pasivo en la fecha determinada, y que este coste era igual al coste atribuido.

Equivalente

deemed cost < a deemed cost, the deemed cost, deemed costs >

Ejemplo

- Las partidas de edificios, instalaciones y equipos se valoran al coste como coste atribuido menos amortización acumulada y deterioros.
- Items of property, plant and equipment are measured at cost as deemed cost less accumulated depreciation and impairment losses.

Example 4: Example of *specialised definitions* in the *Diccionarios Valladolid UVa* (Spanish–English Bilingual specialised dictionaries).

2. *Semi-specialised definitions*: those which describe the meaning of abstract concepts that have been subjected to a process of terminologisation and that can, therefore, be *also* found in non-specialised texts. We use two main defining styles, a full-sentence Cobuild-style definition in a specialised dictionary and a gloss in a general dictionary (examples 5 and 6, respectively); they are crafted by lexicographers and basically target human translators and interested laypersons:

cancellation

Definición

Cancellation is the act of bringing an arrangement, e.g. a contract, to an end as from a particular date.

Example 5: Example of *semi-specialised definitions* in the *Diccionarios Valladolid UVa* (English specialised dictionaries).

cancelación (English: cancellation)

Definición

anulación del efecto de una obligación jurídica

Example 6: Example of *semi-specialized definitions* in the *Diccionarios Valladolid UVa* (Spanish general dictionaries).

3. *Special definitions*: those which describe the meaning of "entities" present in the physical world, e.g. animals, objects, plants, etc., i.e., someone or something that can be pictured. They contain chunks of clauses that are juxtaposed with semi colons. They describe the entity step by step and illustrate it, usually by linking users to an image (example 7) and by using metalanguage (*es decir, ejemplo*, etc.) that *always* explains the terms used in the definition. These are typically used in the general dictionaries of the portal and target native speakers and algorithms, i.e. these are used for creating patterns that will facilitate operations with *Write Assistant*, to which we will refer below:

balance (English: balance)

Definición

informe financiero que analiza el activo, o bienes o valores de los que dispone la empresa, y el pasivo, u obligaciones económicas de la misma como deudas, préstamos, etc., y su patrimonio neto, la diferencia entre el activo y el pasivo de la empresa; este análisis permite conocer la situación económica y financiera de una empresa o entidad en un momento determinado; se usa en economía

<https://upload.wikimedia.org/wikipedia/commons/thumb/a/aa/Balances.JPG/395px-Balances.JPG>

Example 7: Example of *special definitions* in the *Diccionarios Valladolid UVA* (Spanish general dictionaries).

4. *Descriptive definitions*: those which describe the meaning of qualitative lemmas, e.g. adjectives. There are two defining styles: a full Cobuild-style sentence or a chunk of words starting with "que" (English which or that) followed by one or more clauses. In both styles, we usually use explanations preceded by "i.e.", "esto es", or examples preceded by "e.x." with the aim of including denotative meaning (examples 8 and 9). The former type mainly targets human translators whereas the latter is addressed at native speakers and algorithms, as in the case of the previous definition style:

biannual

Definición

Biannual means half-yearly, i.e. every six months.

Example 8: Example of *Descriptive definitions* in the *Diccionarios Valladolid UVA* (English specialised dictionaries).

zazoso (English: stuttering)

Definición

que hace referencia a la persona que tartamudea, esto es, que tiene problemas para leer o hablar por repetir sílabas o pronunciar de forma entrecortada

Example 9: Example of *Descriptive definitions* in the *Diccionarios Valladolid UVa* (Spanish general dictionaries).

5. *Action definitions*: those which describe processes, e.g. verbs. These typically consist of one or more activities that interact to achieve a result, and they are also accompanied with exemplification (example 10); here both human speakers and algorithms are targeted:

abet

Definición

relacionarse con alguien con el único fin de lograr que dicha persona haga algo que no debería hacer, normalmente cometer un acto delictivo

equivalente

inducir

Example 10: Example of *Action definitions* in the *Diccionarios Valladolid UVa* (English–Spanish bilingual general dictionaries).

6. *Function definitions*: those which generally describe the meaning, if any, and function (usually) of lemmas such as adverbs, prepositions, interjections, and so on. We typically use chunks of words joined by semi-colons and target both human speakers and algorithms (example 11):

as at

Definición

expresión que indica un hecho que ocurre en un momento determinado; se utiliza normalmente en finanzas; se utiliza para conectar dos sintagmas, es decir un grupo de palabras que constituyen una unidad de funcionamiento, como por ejemplo "capaz de leer una novela", que es un sintagma adjetivo

equivalente

como en

Example 11: Example of *Function definitions* in the *Diccionarios Valladolid UVa* (English–Spanish bilingual general dictionaries).

7. *Pattern definitions*: those which describe the meanings in terms of semantic types (see Hanks' *Pattern Dictionary of English Verbs*). We use this defining style for explaining expressions, e.g. idioms, quotations, proverbs, and so on. We use chunks of words followed by a semantic type of the expression and its translation in bilingual dictionaries. The targets here are human speakers and algorithms (example 12):

play one's ace card

Definición

emplear una persona su mejor recurso con intención de conseguir ventaja en una situación

equivalente

jugar su mejor baza

- Someone plays his or her ace card
- Alguien juega su mayor baza

Example 12: Example of *Pattern definitions* in the *Diccionarios Valladolid UVA* (English–Spanish bilingual general dictionaries).

8. *Equivalents*: All the bilingual dictionaries have one equivalent per definition, as shown in examples 10, 11 and 12.

In addition, our ad-hoc typology has allowed us to equip our dictionary portal with technologies for using two innovative types of search systems, both of which are currently working with the general Spanish dictionaries and the English–Spanish bilingual dictionaries. The first system is a WordFinder, i.e. a search system that explores the meaning and lexical remarks fields for concepts, i.e. search "definitions". Users can adapt their search with Boolean operators. For example, a Spanish user may look up American poisonous serpents by writing "serpiente – veneno + Hispanoamérica" in the search engine. This will retrieve dictionary articles for more than 100 words, e.g. the dictionary articles for *macaurel*, *boa común*, *mazacuata*, *tragavenado*, and *Boa constrictor imperator*. All these are non-poisonous serpents that live in Latin America. If the users changes the search string for "serpiente + veneno + Hispanoamérica", more than 100 dictionary articles will also be retrieved, but in this case poisonous serpents such as *serpiente de cascabel*, *crótalo*, *víbora de foseta*, and so on. As previously indicated, this system works by exploring the meaning and lexical remark fields, and can find the concepts because of the *special definition* type used (example 13):

macaurel (English: large poisonous snake)

Definición

serpiente de la familia de los boidos o boas que habita en Venezuela y otras zonas de América Central y del Sur; es una serpiente

de hábitos nocturnos y no venenosa; es similar a la boa constrictor llamada tragavenado en Venezuela, pero más pequeña (de hecho, a menudo se considera que la tragavenados y la macaurel son el mismo tipo de serpiente); de color marrón, con manchas en forma de "H", musculosa y grande (puede llegar a medir cuatro metros); es temida por sus constantes ataques a las aves de corral y animales pequeños; habita, fundamentalmente, en las zonas cálidas y bajas de Venezuela

Example 13: Definition of *macaurel* in the *Diccionarios Valladolid-UVa* (Spanish general dictionary)

The second system also searches in the meaning field of the lexicographic database of the Dictionary Writing System, i.e. search "definitions". It is employed when a user logs on to *Write Assistant*, software developed to help Spanish native speakers to write English texts, and does not know an English word (Fuertes-Olivera and Tarp 2020). Examples 4, 6, 7, 9, 10, 11, 12 and 13 illustrate definitions that are very different from those existing in Spanish general dictionaries and English-Spanish bilingual ones. Our definitions will be particularly useful for the *Write Assistant*, which will be *mainly* fed by the *Diccionarios Valladolid-UVa*. *Write Assistant* is currently using statistics and will use machine learning and neural networks technologies for displaying patterns that will offer users of *Write Assistant* the most regular possibilities (Fuertes-Olivera and Tarp 2020). It basically works as follows: a Spanish native speaker is writing in English and he or she does not know or remember an English word. He or she can then write in Spanish and *Write Assistant* will offer them English possibilities, all of which are ordered according to frequency and possibility patterns that will be created with machine learning and neural network technologies that "learn" by analysing huge amounts of data; therefore, our definitions are long and complete. In this regard, although they are included in general dictionaries, the definitions in the examples above are basically similar to *terminological definitions*: they contain a lot of data, are very precise, and are linked to good contexts (chunks of words, collocations and examples) and images (e.g. facial recognition programs work with images). For instance, *coste atribuido* (example 4) is not present in *WordReference.com* (Spanish-English), *Diccionario de la Lengua Española* (RAE 2014), *Diccionario de Contabilidad Inglés-Español/ Español-Inglés* (Sánchez 2003), *Diccionario de uso del español* (Moliner 2007), nor *Diccionario del español actual* (Seco et al. 2011). Meanwhile, *zazoso* (example 9) is defined as *tartamudo* (*Diccionario de la Lengua Española*) and as *tartamudo* or *ceceoso* in the *Diccionario de uso del español* (Moliner 2007). The other dictionaries do not include this word. In the same vein, *macaurel* is poorly defined in the *Diccionario de la Lengua Española* (example 14):

macaurel (English: large poisonous snake)
Serpiente de Venezuela, no venenosa y parecida a la tragavenado,
pero de menor tamaño.

Example 14: Definition of *macaurel* in the *Diccionario de la Lengua Española*

To sum up, the ad-hoc typology of definitions is associated with certain topics that are currently addressed in *terminology*, whilst offering a proposal for working with tools such as *Write Assistant*, and illustrating the economic value of lexicographic and terminological data in the data-driven economy associated with digital knowledge.

5. Conclusion

The *Diccionarios Valladolid-UVA* represent an attempt to make lexicographic and terminological activities profitable. This combines the tenets of the Function Theory of Lexicography (dictionaries are nothing more and nothing less than reference tools covering language, facts and things) with developments in business management. In this context, the focus is especially on the use of disruptive technologies for "attacking" the market dominance of incumbents and increasing the customer base of a product or service, together with the coming of age of the data-driven economy, which values the use of huge amounts of data for establishing patterns that can be used for many different activities, e.g. for upgrading existing language tools such as writing assistants. Both theoretical standpoints are based on two broad assumptions. Firstly, we have assumed that the concept of an integrated dictionary portal will allow us to include as many lemmas as possible, without limiting them and their lexicographic treatment to existing dictionary typologies, e.g. to differences between general and specialised dictionaries. Secondly, our dictionary entries are comprised on average of 200 words plus links to open data per sense. This makes them very different from existing dictionaries which as a rule use around 30 words and almost no link per sense (Fuertes-Olivera, Tarp and Sepstrup 2018). Of particular interest is the definition employed for describing each lemma; we have created an ad-hoc typology of definitions that is geared towards allowing AI technologies to develop patterns that will greatly improve the construction of assistants and similar language technologies. Already existing definitions found in monolingual and bilingual online dictionaries are of little use in this area.

Both developments explain the confluence of terminology and lexicography commented on at the beginning of this article. Following Bowker's Table 1, the following aspects may be appreciated regarding the *Diccionarios Valladolid UVA*: lexicographers, terminologists, translators and experts in the field work side by side; we analyse both words and terms in a similar fashion; we cover general and specialised domains; users can also search for domains (neither an alphabetical nor a thematic approach is necessary in online dictionaries); we

equip *all* our lemmas with up to 67 different lexicographic data; our intended users are professionals and organisations interested in and with the resources for using these dictionaries in combination with other tools, e.g. assistants.

We have illustrated the convergence of both disciplines with an ad-hoc typology of definitions. This typology is not based on any specific theory but on methodological issues (it facilitates daily work), and on the assumption that definitions such as those of examples 4, 6, 7, 9, 10, 11, 12 and 13 can be used with AI technologies. Results carried out in several countries using a Test Driven Development methodology are confirming our initial intuition that software such as *Write Assistant* can be of much use and that its users are not interested in differentiating between lexicography and terminology but in answering their queries in the fastest and easiest way possible.

References

Dictionaries and Tools

- Diccionario de la Lengua Española*. 2014. Real Academia Española. 2014. *Diccionario de la lengua española*. 23rd edition. Madrid: Espasa. <http://www.rae.es/>. Accessed 18 September, 2020.
- Hanks, Patrick**. 2000. *Patterns Dictionary of English Verbs*. http://pdev.org.uk/#about_cpa. Accessed 18 September, 2020.
- Lexico.com*. <https://www.lexico.com/>. Accessed 18 September, 2020.
- Moliner, Maria**. 2007. *Diccionario de uso del español*. Third Edition. Madrid: Gredos. OWID: <http://www.owid.de/>. Accessed 31 January, 2020.
- Sánchez, Nora**. 2003. *Diccionario de contabilidad. Inglés-español/español-inglés*. Hoboken: John Wiley.
- Seco, Manuel, Olimpia Andrés and Gabino Ramos in collaboration with M^a Teresa de Unamuno, Juan Antonio Villafañez and Carlos Domínguez**. 2011. *Diccionario del español actual*. Madrid: Aguilar.
- The Free Dictionary*: <http://www.thefreedictionary.com/>. Accessed 18 September, 2020.
- WordReference.com*: <http://www.wordreference.com/>. Accessed 18 September, 2020.
- Write Assistant*: <https://www.writeassistant.com/es>. Accessed 18 September, 2020.

Other References

- Boelhouwer, Bob, Hindrik Sijens and Anne Dykstra**. 2018. Dictionary Portals. Pedro A. Fuertes-Olivera (Ed.). 2018: 754-766.
- Bowker, L.** 2018. Lexicography and Terminology. Pedro A. Fuertes-Olivera (Ed.). 2018: 138-151.
- Christensen, Clayton**. 2011. *Disruptive Innovation: The Christensen Collection*. Harvard: Harvard Business Review Press.
- Christensen, Clayton, Michael E. Raynor and Rory McDonald**. 2015. What is Disruptive Innovation? *Harvard Business Review*, December 2015. <https://hbr.org/2015/12/what-is-disruptive-innovation>. Accessed 18 September, 2020.

- Engelberg, Stefan and Carolin Müller-Spitzer.** 2013. Dictionary Portals. Rufus, H. Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 1023-1035. Berlin/Boston: Mouton de Gruyter.
- Fuertes-Olivera, Pedro A.** 2015. Lexicographical Storing: A Key Lexicographical Task in the Era of the Internet. *Lexicographica* 31: 67-89.
- Fuertes-Olivera, Pedro A.** 2016. *European Lexicography in the Era of the Internet: Present Situations and Future Trends*. Plenary talk, Beijing, 2 December 2016. Talk sponsored by the Commercial Press and the Chinese Association of Lexicography.
- Fuertes-Olivera, Pedro A. (Ed.)**. 2018. *The Routledge Handbook of Lexicography*. London: Routledge.
- Fuertes-Olivera, Pedro A.** 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: The *Diccionarios Valladolid-UVa*. *Lexicography* 6(2): 21-41.
- Fuertes-Olivera, Pedro A. and Ascensión Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography. The Representation of Meaning in English and Spanish Business Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Fuertes-Olivera, Pedro A. and Sven Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin/Boston: De Gruyter.
- Fuertes-Olivera, Pedro A. and Sven Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.
- Fuertes-Olivera, Pedro A., Sven Tarp and Peter Sepstrup.** 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the *Diccionarios Valladolid-UVa*. *Lexikos* 28: 152-176.
- Harris, Roy and Christopher Hutton.** 2007. *Definition in Theory and Practice. Language, Lexicography and the Law*. London/New York: Continuum.
- Kilgarriff, Adam.** 2006. Word from the Chair. Gilles Maurice de Schryver (Ed.). 2006. *DWS. Proceedings of the Fourth International Workshop on Dictionary Writing Systems, 5 September 2006, Turin, Italy (Pre-EURALEX 2006)*: 7. Pretoria: (SF)².
- Koeva, Svetla and Diana Blagoeva.** 2013. The Dictionary Writing System Lexit and its Application in Bilingual Lexicography. *Cognitive Studies/Études Cognitives* 13: 57-65.
- Müller-Spitzer, Carolin and Christine Möhrs.** 2008. First Ideas of User-adapted Views of Lexicographic Data Exemplified on OWID and eLexiko. *Coling 2008. Proceedings of the Workshop on Cognitive Aspects of the Lexicon (CogALex 2008), Manchester, 24 August 2008*: 39-46.
- Nielsen, Sandro.** 2011. Function- and User-related Definitions in Online Dictionaries. Kartashkova, F.I. (Ed.). 2011. *Ivanovskaya leksikograficheskaya shkola: traditsii i innovatsii [Ivanova School of Lexicography: Traditions and Innovations.] A Festschrift in Honour of Professor Olga Karpova*: 197-219. Ivanovo: Ivanovo State University.
- Rundell, Michael.** 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Sager, Juan C.** 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Storjohann, Petra.** 2018. German Lexicography in the Internet Era. Pedro A. Fuertes-Olivera (Ed.). 2018: 568-585.

Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners

Jun Gao, *School of Foreign Languages, Zhejiang Gongshang University, Hangzhou, China (gj821211@hotmail.com)*

and

Haitao Liu, *Institute of Quantitative Linguistics, Beijing Language and Culture University, Beijing, China; Department of Linguistics, Zhejiang University, Hangzhou, China; Ningbo Institute of Technology, Zhejiang University, Ningbo, China (Corresponding Author, htliu@163.com)*

Abstract: Valency is a major source of lexical errors in foreign language learning. Accordingly, the research question is how the syntactic and semantic properties of a word can be retrieved from the corpora and represented in a Chinese valency dictionary to facilitate foreign learners' vocabulary acquisition. Within the three aspects of the valency framework — logical-semantic, syntactic and semantic-pragmatic valency — this study examines 60 cases of Chinese lexical misuse extracted from the HSK (Chinese Language Proficiency Test) Dynamic Compositions Corpus. The results suggest that the majority of cases of misuse occur in the dimension of semantic-pragmatic valency and that this semantic-pragmatic misuse can be ascribed to various factors such as semantic collocations, emotive variables, text styles, registers, and other contextual factors. The results are then utilized as syntactic, semantic and pragmatic information to be presented in a Chinese valency dictionary. Specifically, the results obtained from a case study of a misused word by referring to a large-scale native Chinese speaker corpus help retrieve a relatively full list of complementation patterns, based on which the study designs a Chinese valency entry that embodies three basic elements — quantitative valency, qualitative valency and valency patterns.

Keywords: CHINESE VALENCY DICTIONARY, VALENCY ENTRY, LOGICAL-SEMANTIC VALENCY, SYNTACTIC VALENCY, SEMANTIC-PRAGMATIC VALENCY, CHINESE VOCABULARY ACQUISITION, LEXICAL MISUSE, COMPLEMENTATION PATTERNS, LEARNER CORPUS, NATIVE SPEAKER CORPUS

Opsomming: Valensiewoordeboeke en Chinese woordeskatverwerwing vir vreemdetaalvaarders. Valensie is 'n groot bron van leksikale foute in die aanleer van 'n vreemde taal. Gevolglik ontstaan die vraag hoe sintaktiese en semantiese eienskappe van 'n woord uit die korpus verkry en in 'n Chinese valensiewoordeboek weergegee kan word om woordeskatverwerwing vir vreemdetaalvaarders te vergemaklik. Met inagneming van die drie aspekte van 'n valensieraamwerk — logies-semantiese, sintaktiese en semanties-pragmatiese valensie — word 60 gevalle van Chinese leksikale foute wat uit die HSK (Chinese Taalvaardigheidstoets) Dinamies Saamgestelde Korpus onttrek is, bestudeer. Die resultate dui daarop dat die meeste van die foute

plaasvind in die semanties-pragmatiese valensie-dimensie en dat hierdie semanties-pragmatiese foute toegeskryf kan word aan verskeie faktore soos semantiese kollokasies, emotiewe veranderlikes, teksstyle, registers, en ander kontekstuele faktore. Die resultate word dan benut as sintaktiese, semantiese en pragmatiese inligting wat in 'n Chinese valensiewoordeboek weergegee moet word. Meer spesifiek, die resultate wat verkry word uit 'n gevallestudie van 'n verkeerd gebruikte woord wat onttrek is uit 'n grootskaalse Chinese moedertaalssprekerskorpus, help om 'n relatief volledige lys aanvullingspatrone, wat gebaseer is op die studieontwerpe van 'n Chinese valensie-inskrywing wat drie basiese elemente insluit — kwantitatiewe valensie, kwalitatiewe valensie en valensiepatrone — te verkry.

Sleutelwoorde: CHINESE VALENSIEWOORDEBOEK, VALENSIEINSKRYWING, LOGIES-SEMANTIESE VALENSIE, SINTAKTIESE VALENSIE, SEMANTIES-PRAGMATIESE VALENSIE, CHINESE WOORDESKATVERWERWING, LEKSIKALE WANGEBRUIK, AANVULLINGS-PATRONE, AANLEERDERSKORPUS, MOEDERTAALKORPUS

1. Introduction

A word, as a 'composite unit of form and meaning' (Lyons 2000: 23), has invariably been in the limelight of foreign language teaching and learning. Traditionally, the teaching of Chinese as a foreign language has centered on grammar while neglecting vocabulary to some extent. Sun (2006) argues that vocabulary should assume a fundamental role in Chinese teaching since it is on the basis of vocabulary that grammatical rules can be established. In practical learning, only a handful of lexical problems such as synonyms are posed and addressed in the classroom; whereas, a considerable number of lexical puzzles emerge from learners' daily study due to the specific features of Chinese vocabulary such as flexibility of word order and lack of inflections and derivations (Sun 2006). What follows is a case of misuse of *baifang* (拜访, visit) found in the HSK Dynamic Compositions Corpus (a corpus that collected the writings of non-native Chinese learners who participated in the HSK advanced level writing tests). The correct word is offered in the parentheses:

- (1) 爸 和 妈 我 很 想 你 们, ... 我 一 定 要 拜 访 (探 望) 你 们。
ba he ma wo hen xiang nimen, ... wo yiding yao baifang (tanwang) nimen.
Dad and Mom I very miss you, ... I surely will visit you.
Mom and Dad, I miss you so much. ... I will visit you no matter what happens.

In this example, the examinee intended to express the meaning 'to pay a visit to his or her parents'. However, despite the fact that both *baifang* (拜访) and *tanwang* (探望) are honorific verbs whose PATIENTs are the elder members of one's family, 'parents' are excluded from the list of semes (i.e. semantic features) presupposed by the PATIENT of *baifang*. In this respect, *baifang* and *tanwang* are synonyms and share the same English equivalent 'visit', but require different semantic features for their PATIENTs. Without solid lexical knowledge or proper guidance from teachers and reference books, learners tend to misuse the word.

As the 'silent teacher', learners' dictionaries are expected to provide systematic information of lexical usage. However, compared with the worldwide popularity of English learners' dictionaries, Chinese dictionaries for foreign learners (hereafter CLDs) have received scant attention. According to the investigation conducted by Xie et al. (2015: 4), more than 80% of foreign learners of Chinese 'do not know or barely know' CLDs. Most of them consult Chinese or Chinese–English dictionaries compiled for native speakers of Chinese. This situation may cause problems for foreign learners. Therefore, it is an urgent task to enhance learners' awareness of using CLDs.

Apart from enhancing learners' awareness, the quality and user-friendliness of CLDs need to be considered. Nevertheless, existing CLDs have some weaknesses that might not cater for learners' practical needs. For instance, Zhang (2011) points out that the system of CLDs, especially their definitions, largely follow the fashion of traditional Chinese dictionaries for native speakers; Xie et al. (2015) found that it is rather common that present CLDs lack systematic syntactic and pragmatic information. For example, the following entry of *baifang* in the *Commercial Press Learners' Dictionary of Contemporary Chinese* (Lu and Lv 2007) only shows one syntactic pattern — (NP) + VP + NP, and it lacks the pragmatic information of excluding 'parents' from the object slot of visiting:

拜访 *bàifǎng* (动) 客气话, 带着敬意看望: 拜访师长、亲友 | 我们周末去拜访王老师 | 上星期我们去拜访了一位老作家 | 有时间我一定去拜访您。

拜访 *bàifǎng* (verb) **polite words, to visit sb. with respect:**
to visit teachers, elders or relatives | We will visit teacher Wang at the weekend | We visited a veteran writer last week | I will visit you if I have time.

Another problematic situation is that foreign learners, as noted, are inclined to consult Chinese–English dictionaries and that most monolingual CLDs provide simple English equivalents for the Chinese entry-words. Thus, English, as the international language, may influence the acquisition of Chinese vocabulary to some degree. In Example 1, learners' incomplete lexical knowledge of *baifang* and their association with the English equivalent 'visit' may cause the negative transfer and make a syntactically-correct but semantically-and-pragmatically-incorrect sentence. The mastery of the three aspects of a word — syntactic, semantic and pragmatic — is essential in foreign language learning. To achieve this goal, foreign learners need a CLD that offers systematic information on words.

In this regard, valency theory and valency dictionaries, having proved to be effective in foreign language teaching and learning (Herbst and Götz-Voteler 2007; Helbig and Schenkel 1969), could lend theoretical and practical support

for foreign learners to acquire Chinese vocabulary in that valency constructions present the syntactic-semantic-pragmatic information of lexical units in a systematic and comprehensive manner.

It is generally acknowledged that modern valency theory was founded by French linguist Lucien Tesnière (1959) and was then systematically developed by German scholars. The notion of valency was borrowed from chemistry. As atoms have the ability to combine with a certain number of other atoms to constitute larger units, words have the property of attracting a selected number of words to form larger units. Accordingly, valency can be generally defined as the 'ability of words to combine in this way with other words' (Herbst et al. 2004: vii). The first valency dictionary of German verbs was compiled in 1969 by German linguists Helbig and Schenkel, and then Sommerfeldt and Schreiber published valency dictionaries of German adjectives and nouns respectively in 1974 and 1977. Not incidentally, the first English valency dictionary (VDE) was also compiled by scholars from Germany (Herbst et al. 2004). One of the principal aims of valency dictionaries, as argued by Herbst et al. (2004: vii), is to help 'advanced foreign learners to write grammatically correct and idiomatic English because it shows them in which constructions a word can be used'. In this respect, the nature of valency dictionaries tallies with that of learners' dictionaries as the latter's 'most interesting features are their efforts to develop new ways of defining words and provide information necessary for encoding' (Béjoint 2002: 73).

VDE is by nature a descriptive dictionary that provides a comprehensive depiction of the valency properties of the English lexicon. The representation of comprehensive lexical information is realized by its profuse use of grammatical and semantic codes extracted and synthesized from the Bank of English and the COBUILD-corpus of present-day English. Metalexicographers have expressed their concern over the use of codes in general-purpose learners' dictionaries: while compilers spare no efforts to include information for encoding in the dictionary, users take far less interest in consulting this type of information since they find information in coded form is too dense, confusing and time-consuming for grasping (Béjoint 2002; Cowie 2002). However, this concern about the usability of coded information for language production can possibly be lessened when taking into account users' study activities and proficiency level as well as the purpose of the dictionary. Bareggi's (1989) study shows that first-year undergraduates of English tend to use dictionaries mainly for decoding activities while from the third year on, they begin to attach equal importance to encoding. Bareggi (1989) also indicates that only 50 percent of the freshmen are able to comprehend grammatical codes, but for juniors, the figure rises to almost 100 percent. Neubach and Cohen (1988) also confirm that learners with high proficiency are better at making use of the dictionary information than those with low proficiency. At this point, VDE, as stated, primarily aims to serve the advanced learners in their encoding activities. It is expected that the intended users of VDE, if properly trained, can better understand and exploit the information for encoding activities.

Despite its strengths, Fillmore (2008) points out some weaknesses of VDE, which may impair its practicability. For instance, VDE treats verbs in detail and at length while not giving equal weight to nouns and adjectives. Particularly, the nouns derivationally related to verbs have combinatory properties similar to those of the corresponding verbs and thus need to be treated equally. This critical point is a significant reminder for future research and compiling of a Chinese valency dictionary undertaken by the present study. Moreover, 'by being corpus-based and therefore non-prescriptive, VDE has no way to introduce negative evidence, and the entries are not set up to include warnings about mistakes' (Fillmore 2008: 78). This indicates the importance of including learner corpora, in addition to native speaker corpora, in the construction of a Chinese valency dictionary to provide prescriptive guidelines for users. Therefore, apart from the description of valency information gained from a native speaker corpus (CCL), the current study will provide 'Note' blocks, alerting the users to common lexical errors extracted from a learner corpus (HSK Corpus). (See examples in Section 3.4 and Figure 3 in Section 4.)

Just as the compilation of VDE was greatly influenced by German thoughts of valency, the introduction of valency theory into China was initially promoted by Chinese scholars of German language (Wu 1996). Valency theory was widely used in the study of Chinese grammar (Shen 2000; Shen and Zhen 1995; Yuan 2010; Zhou 2011) and teaching Chinese as a foreign language (Lu 1997; Shao 2002). Correspondingly, some scholars proposed that the theoretical framework of valency be employed to construct CLDs. Mei (2003) believes that valency theory can be utilized to help set up the grammatical information of modern Chinese learners' dictionaries. With regard to the difficult issue of synonym discrimination that often challenges foreign learners of Chinese, Zhang's (2007) study holds that the syntactic, semantic and pragmatic levels explicitly presented by valency models can provide elaborate discriminatory information of synonyms. Xu (2012) examines divalent Chinese nouns in the *Contemporary Chinese Dictionary* (native-speaker-oriented) and the *Commercial Press Learners' Dictionary of Contemporary Chinese* (foreign-learner-oriented). Xu (2002) found that about 40% of the lexicographic definitions and examples in these two dictionaries do not incorporate valency elements. This calls for the urgent task of systematically organizing the complements of divalent nouns in CLDs. Furthermore, Han and Han (1995a and 1995b) endeavoured to compile a 'contrastive German–Chinese valency dictionary of verbs'. Taking Mannheim school's valency theory as the foundation, they adopted a semantic view of valency for describing Chinese verbs' valency patterns. Similarly, Zhan (2000) proposed to develop a valency-based dictionary — *A Valency-Based Semantic Dictionary*. Zhan (2000) holds the view that the representation of the semantic valency information of a Chinese synonymous word in the dictionary can help differentiate the semantic collocations in its different senses, thus assisting the computer to identify accurate English equivalents.

In light of the previous review of valency dictionaries, the research into

and the compilation of a Chinese valence dictionary for foreign learners, which primarily aims to provide comprehensive syntactic-semantic-pragmatic information of lexical units, is of practical significance. For one thing, the inclusion of valency into dictionaries can solve, at least to a considerable degree, the problems of current CLDs such as lack of syntactic information, under-differentiated senses and ill-arrangement of examples according to syntactic features of the headwords. For another, there are only limited studies that relate valency theory to CLDs. Some of these studies base the syntactic and semantic collocation information on researchers' intuition and introspection rather than on corpus evidence. In this regard, the present study follows the compilation principle of the *Valency Dictionary of English*, which combines 'corpus research and the theoretical background of valency theory' (Herbst et al. 2004: xxii). On account of these problems and facts, the research question of this study is how the syntactic and semantic properties of a word can be retrieved from the corpora and be presented in a Chinese valency dictionary to facilitate foreign learners' vocabulary acquisition. To answer this question, the study firstly elaborates on valency theory and key concepts of the user perspective (Section 2). It then employs two Chinese corpora and the *Valency Dictionary of English* to analyze the misuse of Chinese vocabulary within the valency framework, for the purpose of setting up a database for the construction of a Chinese valency dictionary (Section 3). Finally, the study tentatively designs a valency entry of a Chinese word based on the results of analysis (Section 4).

2. Valency theory and the user perspective

This section lays the theoretical basis from two parts. One is the construction of a valency framework for data analysis and discussion as well as the design of a valency entry in a proposed Chinese valency dictionary. Another part is the user perspective from which the user-friendliness of the Chinese valency dictionary undertaken by this research project can be enhanced.

2.1 The valency framework

Different researchers compare valency to different but similar concepts. Tesnière, the founder of valency theory, perceives a verb as an '*atome crochu*' (atom with hooks) which can attract a certain number of actants as its '*dépendance*' (1959: 238). De Groot (1949) focuses on the idea of 'restriction', arguing that different classes of words have different patterns of syntactic valency and that valency refers to the possibility and impossibility of headwords restricting or being restricted by other words. Therefore, valency is not an exclusive property of verbs but a shared capacity of all other word classes such as nouns, adjective, adverbs, prepositions, numerals, etc. In fact, Sections 73 to 77 in Tesnière's (1959) work discusses the valency structure of nouns, adjectives and adverbs. Kac-

nel'son's (1948 and 1988) key idea is 'potentiality', defining a word as a lexical unit that has syntactic potentiality, which enables content words to combine with other words. Hence, valency is regarded as a means to express potential syntactic relations or to uncover potential grammatical phenomena.

Based on Tesnière and other forerunners, German scholars developed the valency theory into a system. In general, the valency framework constructed by German scholars can be summarized with regard to three dimensions: logical-semantic, syntactic and semantic (Gao and Liu 2019; Han 1993 and 1997). Bondzio (1978) upholds the concept of logical-semantic valency. A headword has the ability of governing other words (i.e. complements) by assigning different semantic roles to them. For example, the verb *visit* reveals the relation between 'visitor' and 'the visited'. Bondzio (1978) adopts the term 'slot' from logic to refer to the relations between the headword and its dependents. In this respect, *visit* has two slots determined by its semantic components. Accordingly, these conceptually based slots constitute the valency of a word, and the number of slots corresponds to the number of valency. Hence, logical valency relates to the quantitative aspect of valency.

Syntactic valency, put forward by Helbig (1992), is the formal realisation of logical valency in a language. In a particular sentence, the logical-conceptual relations are transformed into syntactic relations. However, syntactic structure and logical structure do not coincide on many occasions because sometimes the former cannot fully realize the latter. For instance, logically, *give* is a trivalent verb that can govern three elements 'agent', 'patient' and 'recipient'. Syntactically, *give* can be a monovalent, divalent or trivalent verb as in the sentences *To give or take is a choice*, *They were given a box to carry* and *My teacher gives me a book*. Therefore, a logical valency structure may have different syntactic representations.

Different from logical valency, semantic valency is concerned with the semantic or collocational properties of the headword. The semantic roles assumed by the complements of the headword, as depicted in logical valency, need to have compatible properties in order to combine with the headword (Han 1993). As seen in the Introduction, *baifang* (拜访, visit) prerequisites its PATIENTs to have the semantic feature of 'elder members of one's family, not including parents'. As a result, although *I baifang my parents* is logically and syntactically correct, it is semantically improper in terms of the valency framework.

In addition to logical, syntactical and semantic valency, another aspect — pragmatic valency — is put forward and discussed by some linguists. Růžička (1978) connects valency with communication, pointing out that the selection of complements, mainly optional complements, depends on the context of communication. In certain communicative contexts, obligatory complements can also be removed from the sentence. Helbig (1992) further probes into the connections between valency and communication. Apart from Růžička's idea of contextual selection of complements, Helbig includes two other factors: text

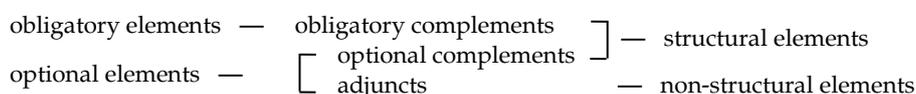
style and semantic collocation. Complementation patterns are determined by various styles of texts, which, in turn, are decided by communicative purposes. For semantic collocation, it refers to the semantic valency mentioned earlier. As the headword presupposes the semantic features of its complements, when they are collocated in such a way, pragmatic errors can be avoided.

Based on German scholars' study on valency, the present study employs a similar but slightly different valency framework. As Section 3 analyzes lexical misuse committed by foreign learners of Chinese due to their misapplication of valency patterns, the analysis will be approached from three aspects of valency: logical-semantic, syntactic and semantic-pragmatic (Gao and Liu 2019). Logical-semantic valency and syntactic valency are drawn respectively from Bondzio (1978) and Helbig (1992) as mentioned earlier in this section, with the former referring to the repertoire of semantic roles assumed by the complements of a headword and the latter referring to the formal realisation of logical-semantic valency in a sentence. However, the sentence formed out of syntactic arrangement of semantic roles is sometimes incorrect in terms of other semantic and pragmatic requirements. These factors, such as semantic features (\pm Human, \pm Elder members of the family), text styles (Literary, Explanatory), registers (Written-Formal, Spoken-Informal), emotive variables (Derogatory-Commendatory) and other contextual factors, need to be taken into account. Therefore, semantic-pragmatic valency integrates these factors to guarantee the generation of semantically and pragmatically correct sentences. In summary, the valency framework adopted by this study is illustrated in Table 1 (Gao and Liu 2019: 331):

Table 1: The valency framework

The valency framework	
Logical-semantic valency	Quantitative aspect of valency: the whole array of semantic roles of complements presupposed by headwords, such as 'agent', 'patient' and 'recipient'
Syntactic valency	Formal aspect of valency: syntactic realisation of logical-semantic valency
Semantic-pragmatic valency	Qualitative aspect of valency: semantic features (\pm Human, \pm Elder members of the family), genres (Literary, Explanatory), registers (Written-Formal, Spoken-Informal), emotive variables (Derogatory-Commendatory), and other contextual factors

There is a pair of concepts requiring further explanation, namely complement and adjunct. Although complements and adjuncts are elements of a sentence, they have different structural status. Helbig and Schenkel (1969) have differentiated these two elements as shown below:



Obligatory and optional complements can be distinguished through an 'Eliminierungstest' (elimination test). For instance, in the sentence *He gives me a smile*, if all the complements are deleted one by one, only sentence b is correct:

- a. ___ gives me a smile.
- b. He gives ___ a smile.
- c. He gives me ___.

In this case, 'He' and 'a smile' are obligatory complements while 'me' is optional.

Schumacher (1986) uses the method of an 'Implikationsprobe' (implication) to differentiate optional complements from adjuncts. This method is based on logical-semantic valency, which gives a full list of semantic components of the governing word. Although not all the semantic components are selected in a particular sentence within a particular context, they are implied or presupposed by their governor. For example, *bring* (within the sense of CARRY) implies three semantic components: somebody brings something to somebody else. As a result, for the sentence *He brought a dictionary to me three days ago*, after all the elements are deleted one by one, sentences c and d are correct:

- a. ___ brought a dictionary to me three days ago.
- b. He brought ___ to me three days ago.
- c. He brought a dictionary ___ three days ago.
- d. He brought a dictionary to me ___.

'To me' in sentence c is a semantic component implied by the governing word *bring*; whereas 'three days ago' in sentence d is not within the list of implied complements. Thus, 'to me' is an optional complement and 'three days ago' an adjunct.

The example entry of a Chinese learners' dictionary in Section 4 will be designed within the valency framework and find a practical way to treat obligatory complements, optional complements and adjuncts in the dictionary. Section 3 utilizes the valency framework established as such to analyze and discuss the lexical misuse committed by foreign learners of Chinese, the purpose of which is to prepare authentic data from learner and native speaker corpora for the construction of a Chinese valency dictionary. However, the compiling of a dictionary needs not only guidance from linguistic theory, but also lexicographical principles in order to cater to the needs of intended users. For this reason, the next section adopts the user perspective to set up the basis for its usability.

2.2 Valency dictionaries and the user perspective

Valency dictionaries have been claimed to be a useful reference tool for foreign language learning owing to its comprehensive and systematic description of lexical information. For example, Herbst et al. (2004: vii) state in the VDE that the following questions, which may baffle English learners, can be answered by consulting the dictionary: 'Is it *avoid to do something* or *avoid doing something?*', 'is *try to do something* the same as *try doing something?*', or 'Can you say *the exhibition opened* in English or not?'. Nevertheless, there is no empirical evidence to testify to the usability of valency dictionaries. This study also endeavours to conceive a valency dictionary for foreign learners of Chinese. This conception is, to some extent, limited in that it involves no user surveys, which may not give the intended users what they want in the dictionary. For this limitation, the present authors attempt to fulfil the minimum requirement for lexicographers as proposed by Béjoint (2002: 112): 'Lexicographers must give to the public what the public expects, or at least what they think the public expects, at the expense if necessary of what a truly scientific description of the language would require'. For this purpose, the study employs some of the key concepts of the user perspective to design the dictionary, which is constructed within the valency framework and on the basis of corpus data, for meeting the needs of the expected users. As well, in the last section, this limitation is included as a suggestion for further research into users' expectation in regard to valency dictionaries.

Hartmann (2005) summarises six user perspectives: pedagogical lexicography, dictionary awareness, user sociology, reference needs, reference skills and user training. As these perspectives are interrelated, the study discusses them from four viewpoints.

Firstly, the Chinese valency dictionary proposed in this study is pedagogical in nature. It is a tendency for modern British pedagogical dictionaries to include more information for encoding, such as syntactic patterns, collocations and registers (Béjoint 2002; Svensén 1993). This reflects an increasing need for productive activities on the part of foreign learners. While traditional learners' dictionaries need to consider the simplicity of the setup of encoding information to cater to the need of common users, which may lead to the omission of some vital information (Cowie 2002), a valency dictionary can explicitly present as much encoding information as possible by essentially focusing on helping learners with their encoding tasks. This pedagogical purpose with encoding orientation is reflected in the design of a valency entry of a Chinese valency dictionary in Figure 3, Section 4. Moreover, as argued by Fillmore (2008), valency dictionaries for foreign learners need not only descriptive representation of encoding information, but also prescriptive guidelines. In this regard, the proposed Chinese valency dictionary provides 'Note' blocks, which warn users against common lexical errors found in a learner corpus. (See examples in Section 3.4 and Figure 3 in Section 4.)

Secondly, according to Hartmann (2005), users of pedagogical dictionaries

generally have a low level of awareness of dictionary contents and typology. They, especially low-level users, are more familiar with and take more interest in such information categories as meaning and spelling while neglecting those for encoding activities, such as frequency, syntactic patterns and collocations (Béjoint 2002). The low level of dictionary awareness leads to users' failure to make full use of the information, thus reducing the potential usefulness of a dictionary. Furthermore, the valency dictionary, specially designed for encoding, is relatively new in lexicographical typology, and little is known of its population of active users, as well as their knowledge, proficiency level and skill at using the dictionary. The urgent task, as expressed by some metalexicographers (Béjoint 2002; Cowie 2002; Hartmann 2005), is to introduce user training programs to improve user skills, which entails the joint efforts of the whole education system. For instance, user training programs need lessons and instructions from teachers, the setup of lexicographical courses by academic institutions, the inclusion of user training in the national curriculum by educational departments, the supply of easy-to-read users' guides and workbooks by publishers, and improvements in the user-friendliness of dictionaries on the part of lexicographers.

Thirdly, user sociology and needs are closely connected, as Hartmann (1989: 103) hypothesizes that 'different user groups have different needs'. Among the six aspects regarding reference needs (Hartmann 2005: 88), two are relevant to the present study: text production (semantic or syntactic problems for writers) and language acquisition. In order to investigate these needs for a better design of the Chinese valency dictionary, an intricate set of sociological variables of users — age, gender, first language background, foreign language proficiency level, educational background, habit of using dictionaries, attitude toward dictionaries, ownership, dictionary awareness, etc. — should be taken into consideration in future surveys as suggested in Section 5. For achieving as much user-friendliness as possible, the study, in its present form, attempts a short profile of prospective users, including the explanation of their proficiency level and the contexts in which they are expected to use the dictionary. (See Section 4.)

Fourthly, a good command of reference skills is the prerequisite for successful dictionary use and the ensuing satisfaction of user needs. The types of linguistic activities conducted by users determine the types of skills and strategies needed in the look-up process (Hartmann 2005, Wiegand 1998). As mentioned, the Chinese valency dictionary examined in this study aims to meet the user needs of language acquisition in general and text production in particular. It is suggested that in future empirical surveys, special attention be paid to the skills required in the consultation of information in the Chinese valency dictionary for writing activities engaged in by foreign learners of Chinese. Specifically, the survey could be implemented according to the seven essential components of the consultation process specified by Hartmann (2005: 90-92): activity problem, determining problem word, selecting dictionary, external search (macrostructure), internal search (microstructure), extracting relevant

data and integrating information. Another possible way of training users' reference skills for consulting the Chinese valency dictionary is to provide relevant training exercises in the front or back matter.

This section first establishes the linguistic foundation — the valency framework — for designing a Chinese valency dictionary and then touches on the user perspective in the hope that the design and the future compilation of the dictionary could be friendly to prospective users, on the basis of which the data analysis and discussion in the next section are conducted.

3. An analysis of lexical misuse caused by misapplication of valency patterns

3.1 Research design

The present section of this study comprises three main parts. In section 3.2 misused cases of Chinese words are collected from the HSK Dynamic Compositions Corpus (hereafter HSK Corpus). In section 3.3 an analysis of the collected data is conducted along the three dimensions of the valency framework. The analysis is assisted by consulting three Chinese dictionaries and the Beijing University Corpus of Modern Chinese Language (CCL). In section 3.4 the results of the analysis, which have implications for a Chinese valency dictionary, are discussed, and then a comparison is drawn between the valency patterns of misused words and those of English equivalents in order to identify the influence of improper transfer of valency structure on lexical use, which suggests the idea of a contrastive bilingual valency dictionary.

It is necessary to make a brief introduction to HSK Corpus and CCL before the analysis unfolds. As the source of data collection, HSK Corpus is a learner corpus developed by Beijing Language and Culture University. It collected the compositions of non-native learners of Chinese who participated in HSK high-level tests from the year 1992 to 2005. Its scope now reaches 4,240,000 words and 11,569 compositions. Besides the collection of original compositions through scanning, it includes annotated materials in which interlanguage misuse is manually labelled. The annotation covers five levels of the Chinese language: Chinese characters, punctuation, words, sentences and texts. This study stays at the level of words whose annotation ranges from misused words, missing words to unnecessary words, supplemented by the statistics of various types of lexical misuse.

As the reference corpus, CCL is a native speaker corpus developed by the Center for Chinese Linguistics, Beijing University. It collects various sources of modern Chinese language such as spoken language materials from TV dialogues and interviews as well as written language materials from history, government white papers, economic reports, health and medicine, dictionaries, newspapers, films, literature, translation, essays, etc. Its scope now reaches 581,794,456 words.

3.2 Data collection

As mentioned earlier, HSK Corpus provides its users with statistical facts about interlanguage misuse at all levels. This study focuses on the lexical level, but due to the limitation of time and space, it does not attempt to examine all the types of lexical misuse committed by foreign learners of Chinese, only selecting some samples. First, Chinese words whose *pinyin* (Chinese pronunciation system) begins with letter Z are chosen as the level-one sample for investigation. The Z-group occupies the largest portion of Chinese words in the corpus (grouped together according to initial *pinyin* letters), containing 2,369 words, among which 622 are annotated as misused cases. These 622 annotated words are then selected as the level-two sample for examination. The annotation covers all types of misuse such as misspellings, missing words and unnecessary words. As well, the annotation covers all classes of words like nouns, verbs, adjectives, adverbs, conjunctions, prepositions, etc. For a systematic and thorough inquiry into complementation patterns, the study chooses verbs as the level-three sample. We conduct a close screening of these verbs one by one in order to trace interlanguage misuse caused by improper transfer of valency patterns from English to Chinese.

The criterion of screening is that the misused word and the correct one provided by the corpus share the same English equivalent(s). The equivalents are confirmed by consulting some Chinese dictionaries for foreign learners such as *A Dictionary of Chinese Usage: 8000 Words* (Chinese Proficiency Center Beijing Language and Culture University 2000) (hereafter HSK 8,000) or bilingual dictionaries for native learners such as *A New Century Chinese-English Dictionary* (Hui 2004). In Example 2, *zhuanbian* (转变) and *gaibian* (改变) have the same English equivalent 'change', but *zhuanbian* is labelled as a misused case. The reason is that although *zhuanbian* and *gaibian* can both govern such complements as 'attitudes' and 'ideas', 'living environment' is not within the governing power of *zhuanbian*. The meaning nuance of this pair of near-synonyms is implicitly embodied in their semantic collocations, and this implicitness can be unearthed from the perspective of semantic-pragmatic valency.

- (2) 我 现在 生活 的 环境 转变 (改变) 了 很多。
wo xianzai shenghuo de huanjing zhuanbian (gaibian) le henduo.
My current living (aux.) environment change (-ed) a lot.
My current living environment has changed a lot from my parents'.

However, in Example 3, *zhuijiu* (追究, find out/investigate) and *zhuiqiu* (追求, pursue/go after) do not share the same English equivalent. The obvious meaning differences between this pair of lexical items do not necessarily call for a close examination within the valency framework. Thus, cases of this kind are excluded from the study.

- (3) 有些人还追究 (追求) 美食。
youxie ren hai zhuijiu (zhuiqiu) meishi.
 Some people still investigate (pursue) table delicacies.
 Some people pursue table delicacies.

Furthermore, the corpus provides all the misused cases of a word owing to its polysemous nature. Hence, the misuse of a polysemous word can be classified into one or more than one group, and each group is represented by a synonym of one of the senses of the word in question. For instance, the misuse of the word *zaocheng* (造成, create; cause/give rise to; bring about) fall into three groups represented respectively by its synonyms *chuangzao* (创造, create), *chansheng* (产生, cause/give rise to) and *dailai* (带来, bring about). Each group is treated as a case. Finally, after screening, there are 60 cases found in level-three samples, as displayed in Table 2:

Table 2: The list of misused cases for analysis

No.	Misused verbs	Correct verbs	Shared equivalents	No.	Misused verbs	Correct verbs	Shared equivalents
1	造成	创造	create	31	主张	提出	propose
2	造成	产生	cause, give rise to	32	主张	坚持	maintain
3	造成	带来	bring about	33	主张	维护	maintain
4	遭到	遇到	encounter	34	转	推	shift
5	赞扬	赞赏	speak highly of	35	转变	改变	change
6	赞成	支持	approve	36	转换	改变	change
7	战胜	取胜	win	37	祝愿	祝	wish
8	展开	睁开	open	38	注重	关注	pay attention to
9	展开	张开	spread	39	注意	关注	pay attention to
10	照顾	考虑/顾及	consider	40	注意	关心	care for
11	招揽	招	recruit	41	招来	得到	attract
12	掌握	把握	grasp	42	注意	专心	concentrate one's attention on
13	掌握	控制	control	43	注目	注意	keep one's eyes on
14	掌	拿	hold	44	追	追求	pursue
15	珍重	珍惜	treasure	45	撞到	碰到	run into
16	珍惜	爱惜	cherish	46	装	装作	pretend
17	争斗	争取	fight for	47	装满	充满	fill up
18	执行	进行	carry out	48	着想	来看	considering
19	知悉	了解/知道	know	49	滋长	形成	develop
20	知道	认识	know	50	总结	结束	conclude
21	知道	了解	know	51	组织	组建	organize
22	知道	明白	understand	52	组成	营造	form
23	止	戒	stop, quit	53	组成	构建	form
24	制作	制造	produce	54	走	去	go to
25	指点	指出	point out	55	走	走路	walk
26	指导	引导	guide, direct	56	走入	步入/进入	step into, enter into
27	治疗	治好	cure	57	遵守	保持	keep
28	治疗	解决	cure	58	遵守	坚持	stick to
29	助长	帮助	facilitate	59	作出	得出	work out
30	住	留	stay	60	作出	制定	make

3.3 Data analysis

This section scrutinizes lexical misuse within the valency framework and identifies the dimension(s) — logical-semantic, syntactic or semantic-pragmatic — in which the misuse occurs. The scrutiny of lexical misuse is first conducted by consulting both native-speaker-oriented (the *Contemporary Chinese Dictionary*) and foreign-learner-oriented (the *Commercial Press Learners' Dictionary of Contemporary Chinese*; HSK 8,000) Chinese dictionaries. In Example 4, *zhansheng* (战胜, win) is usually a transitive divalent verb that governs an object such as an enemy, a team or difficulty (from the *Contemporary Chinese Dictionary*). Therefore, it is incorrect to use *zhansheng* as a monovalent verb without objects; whereas *qusheng* (取胜, win), an intransitive monovalent one, is appropriate. This kind of misuse involves logical-semantic valency.

- (4) 是 为了 提高 自己, 并 不是 只 为了 战胜 (取胜)。
shi weile tigao ziji, bing bushi zhi weile zhansheng (qusheng).
Be to improve oneself, and not just to win.
The purpose is to improve oneself rather than to win.

In Example 5, both *zanyang* (赞扬, praise) and *zanshang* (赞赏, praise) are divalent verbs, but only *zanshang* can be used in the syntactic pattern 'dui (对, treat) + NP + hen (很, very) + VP' (from HSK 8,000). Therefore, these misused cases involve syntactic valency.

- (5) 我 对 这种 人 很 赞扬 (赞赏)。
wo dui zhezhong ren hen zanyang (zanshang).
I to this kind person very praise.
I speak highly of this kind of person.

In Example 6, it is semantically correct for *zhidao* (知道, know) to be collocated with 'one's thoughts' (from the *Contemporary Chinese Dictionary*). However, in this specific context, *mingbai* (明白, understand) is more appropriate. These cases of misuse, which are largely due to contextual factors, involve semantic-pragmatic valency.

- (6) 子女 ... 才 知道 (明白) 父母 的 想法。
zinv ... cai zhidao (mingbai) fumu de xiangfa.
Children ... just understand parents (aux.) thoughts.
Children began to understand parents' thoughts.

In Example 7, *zhuanhuan* (转换, change) can be both a monovalent and divalent verb and its usage conforms to syntactic rules. However, its object is usually attitude, direction, topic, or pattern, not including environment which, nevertheless, can be governed by *gaibian* (改变, change) (from HSK 8,000). Hence, the semantic collocation is improper. These cases of misuse involve semantic-pragmatic valency.

- (7) 为了 转换 (改变) 一下 工作 环境 ...
weile zhuanhuan (gaibian) yixia gongzuo huanjing ...
To change one time working environment ...
In order to change working environment ...

In Example 8, *zhiliao* (治疗, cure) is a transitive divalent verb and its semantic collocates include 'disease' (from HSK 8,000). Nevertheless, in this context, the speaker implies that the disease needs to be treated and the patient can thus recover. In this respect, *zhihao* (治好, cure) is correct because it incorporates both of the semes 'to treat a patient' and 'to help a patient gain recovery'. Thus, these cases of misuse are motivated by semantic and contextual factors and fall into the dimension of semantic-pragmatic valency.

- (8) 病人 的 病 严重, 没有 希望 治疗 (治好)。
bingren de bing yanzhong, meiyou xiwang zhiliao (zhihao).
Patient (aux.) disease serious, does not have hope be cured.
The patient is seriously ill and has no hope of being cured.

However, some analyses of lexical misuse cannot be accomplished by simply consulting dictionaries. As mentioned in the Introduction, the majority of the definitions and examples in these dictionaries do not contain full lists of complementation patterns (Xu 2012). To solve this problem, the present study turns to CCL, a large-scale native Chinese speaker corpus, by adopting the method of studying semantic prosody (a corpus-based study of the semantic environment of a given word). The approach is a combination of corpus-based and corpus-driven investigations, which involves four steps (Partington 1998; Sinclair 1991 and 1996; Wei 2002). Firstly, a number of concordance lines are randomly retrieved from the corpus. The next step is to determine the span of the node word and then establish colligation(s) (i.e. syntactic structure) by observing the collocates around the node word. Then, semantic features of these collocates are analyzed. The last step is to draw out the semantic prosodies of the key word. This approach can lend support to the analysis of valency structure. The colligations established by observing collocates help identify complementation patterns. The revelation of semantic features of collocates helps to work out a relatively full list of semes of the headword. More importantly, the conclusion of semantic prosodies helps uncover pragmatic information of the headword such as emotive variables.

In Example 9, the usage of *zhaolai* (招来, attract) cannot be found in these dictionaries. For this reason, the study opts for a corpus-based and corpus-driven method and it entails three steps.

- (9) 流行 歌曲 ... 招来 (得到) 人们 的 喜欢 和 支持。
Liuxing gequ ... zhaolai (dedao) renmen de xihuan he zhichi.
Popular songs ... attract people (aux.) fondness and support.
Popular songs attract people's fondness and support.

Step 1: The establishment of colligations

The frequency of occurrence of the node word *zhaolai* in CCL is 1,444, from which we randomly select 100 concordance lines. The span is set as -7/+7, and the words within this span are collocates for observation. Our study shows that there are 5 types of colligations for the node word *zhaolai*:

1) NP + *zhaolai* + NP. This colligation is the most frequent, accounting for 83% of the total concordances.

2) *zhaolai* + *de* (的, structural auxiliary). The colligation accounts for 8% of the total concordances. This construction functions mostly as adjectives (7%) to modify nouns, such as *zhaolai de gongren* (招来的工人, recruited workers), *zhaolai de xuesheng* (招来的学生, recruited students) and *zhaolai de pengyou* (招来的朋友, invited friends). Only one case acts as a noun (1%), such as *zhaolai de ze shi yidui fen'nu de qianze* (招来的则是一堆愤怒的谴责, What our action incurred was a pile of furious denunciation).

3) NP + *shi* (是, be) + *cong* (从, from) + someplace + *zhaolai* + *de* (的, structural auxiliary). This colligation takes up 4% of the total. For example, *zhanshi duoshu shi cong nongcun zhaolai de* (战士多数是从农村招来的, Most of the soldiers were conscripted from rural areas).

4) *ba* (把, used to advance the object of a verb to the position before it) + NP + *zhaolai* + *le* (了, structural auxiliary). This colligation only occurs once (1%). For example, *ba gongren zhaolai le* (把工人招来了, Prospective employees were recruited).

5) Idioms or fixed phrases. There are four cases of idioms centered on *zhaolai* (4%), such as *zhaolai huiqu* (招来挥去, to call in and send away sb. at will) and *congshi zhaolai* (从实招来, admit it; make a clean breast of everything).

Step 2: Collocates and their semantic properties

The first type of colligation forms the overwhelming majority of all the concordances, and the complements of *zhaolai* (招来, attract) are divided into three groups:

Group one: people (such as staff, talented human resources, soldiers, customers and readers) and vehicles. The colligations with this group of complements occupy 31% of all the concordances.

Group two: negative comments, emotions and attitudes (such as condemnation, reproach, indifference, sarcasm, quarrel, catcall, dissatisfaction, criticism and rude language) as well as unfavourable things (such as disaster, trouble, punishment, ill consequence and mosquito). The frequency of this group of complements is higher than that of group one, reaching 46%.

Group three: resisting power (such as opponent, counter-attack and resistance).

The colligations with this group of complements take up only 6%.

The complements in the second type of colligation, which account for 8%, are divided into two groups:

Group one: people (such as personnel, talented human resources, soldiers and friends). The colligations with this group of complements take up 7%.

Group two: negative comments (such as quarrel). The colligation represented by this group occurs once (1%).

The complements in the third type of colligation, which account for 4% of concordances, fall into one group, that is, people such as workers, child labourers and soldiers.

For the fourth type of colligation, which accounts for 1% of concordances, the complement of the governing word *zhaolai* is employees.

The fifth type of colligation (4%) has no explicit complements.

The semantic features of the complements in those colligations can be summarized as follows. The governing verb *zhaolai* presupposes three classes of complements whose respective semantic components are: people; negative comments, emotions, attitudes, unfavourable things; and resisting power.

Step 3: Prosodic structure

Based on the above observations concerning the semantic features of complements, the prosodic structure of *zhaolai* (招来, attract) is summarized in Table 3:

Table 3: The prosodic structure of *zhaolai*

招来	Group of complements	Positive prosody	Negative prosody	Neutral prosody
Colligation 1 (83%)	people, vehicles (31%)	—	—	31%
	negative comments, emotions and attitudes, unfavourable things (46%)	—	46%	—
Colligation 2 (8%)	resisting power (6%)	—	2%	4%
	people (7%)	—	—	7%
Colligation 3 (4%)	negative comment (1%)	—	1%	—
	people (4%)	—	—	4%
Colligation 4 (1%)	people (1%)	—	—	1%
Colligation 5 (4%)	fixed phrases (4%)	—	4%	—
	Total	—	53%	47%

It can be seen from the table that *zhaolai* is not surrounded by a positive semantic environment but by the nearly even configuration of negative and neutral prosodies. Apart from the groups of 'fixed phrases' and 'resisting power', when *zhaolai* is collocated with "people", it carries a neutral prosody, and when collocated with the group of complements — negative comments, emotions and attitudes — it carries a negative prosody. When it comes to Example 9 where *zhaolai* collocates with two positive emotive and attitudinal words *xihuan* (喜欢, fondness) and *zhichi* (支持, support), the foreign learner obviously misuses the word in terms of semantic-pragmatic valency.

With the joint assistance of dictionaries and corpora, the 60 cases of lexical misuse listed in Table 2 are examined and the misuse of valency are identified. The results are summarized in Tables 4, 5 and 6 (due to the limitation of space, examples from the HSK Corpus are not provided):

Table 4: Lexical misuse in the dimension of logical-semantic valency

No.	Misused verbs	Correct verbs	Shared equivalents	Causes for misuse
1	战胜	取胜	win	As a divalent verb, it lacks a 'patient' and cannot be used as an intransitive verb.
2	总结	结束	conclude	As a divalent verb, it lacks a 'patient' and cannot be used as an intransitive verb.

Table 5: Lexical misuse in the dimension of syntactic valency

No.	Misused verbs	Correct verbs	Shared equivalents	Causes for misuse
1	赞扬	赞赏	praise	It cannot be used in the syntactic pattern: 对 (verb: treat) + NP + 很 (adverb: very) + ~.
2	招揽	招	recruit	It cannot be used in the passive pattern: 被 (preposition: by) + ~ + 入 (verb: enter) + someplace.
3	指导	引导	guide, direct	It cannot be used in the negative syntactic pattern: ~ + NP + 不(adverb: not) + VP.
4	注目	注意	fix one's eyes on	It is usually used in the idiom '引人注目' which acts as 'modifiers' or 'predicative'.
5	装	装作	pretend	It usually combines with '出' or '作' (together as verb phrase: pretend to be/as) to express the meaning 'pretend to be'.
6	着想	来看	considering	It cannot be used in the syntactic pattern: 以 (preposition: according to) + NP + ~.
7	走	去	go to	It cannot be used in the syntactic pattern: ~ + 了 (structural auxiliary) + someplace.

Table 6: Lexical misuse in the dimension of semantic-pragmatic valency

No.	Misused verbs	Correct verbs	Shared equivalents	Causes for misuse
1	造成	创造	create	Semantically, it cannot be collocated with 'future'.
2	造成	产生	cause, give rise to	Semantically, it cannot be collocated with 'the feeling of dissatisfaction', though it can be collocated with 'dissatisfaction'.
3	造成	带来	bring about	Semantically, it cannot be collocated with 'side effects (caused by pesticide)'.
4	遭到	遇到	encounter	Semantically, it cannot be collocated with 'troubles or difficulties'.
5	赞成	支持	approve	Semantically, it cannot be collocated with 'green food'.
6	展开	睁开	open	Semantically, it cannot be collocated with 'eyes'.
7	展开	张开	spread	In the context, its seme [spread (arms)] does not agree with the action of 'supporting sb. to walk with hands'.
8	照顾	考虑/顾及	consider	Semantically, it cannot be collocated with 'others' feelings'.
9	掌握	把握	grasp	Semantically, it can be collocated with 'opportunities'; but in the context, this collocation is improper.
10	掌握	控制	control	Semantically, it cannot be collocated with 'neurons'.
11	掌	拿	hold in one's hand	Semantically, it cannot be collocated with 'lucky money wrapped in red envelopes'.
12	珍重	珍惜	treasure	Semantically, it cannot be collocated with 'resources or food'.
13	珍惜	爱惜	cherish	Semantically, it can be collocated with 'me'; but in the context, this collocation is improper.
14	争斗	争取	fight for	Semantically, it cannot be collocated with 'chance'.
15	执行	进行	carry out	In the context, it cannot be used in informal or unofficial writing.
16	知悉	了解/知道	know	In the context, it cannot be used in informal writing.
17	知道	认识	know	Semantically, it can be collocated with 'my classmates'; but in the context, this collocation is improper.
18	知道	了解	know	Semantically, it can be collocated with 'each other's family background or habit'; but in the context, this collocation is improper.
19	知道	明白	understand	Semantically, it can be collocated with 'one's thoughts'; but in the context, this collocation is improper.

20	止	戒	stop, quit	Semantically, it cannot be collocated with 'smoking'.
21	制作	制造	produce	Semantically, it cannot be collocated with 'cigarettes'.
22	指点	指出	point out	Semantically, it cannot be collocated with 'signs of development or one's mistake'.
23	治疗	治好	cure	In the context, it lacks the seme [to help a patient gain recovery].
24	治疗	解决	cure	Semantically, it cannot be collocated with 'problems'.
25	助长	帮助	facilitate	It usually governs the type of complements — 'negative emotions, attitudes, desires, or unfavourable trends, behaviours', but not positive ones.
26	住	留	stay	Semantically, it can be collocated with 'one's heart'; but in the context, it is improper to be used in a non-literary style of writing.
27	主张	提出	propose	Semantically, it cannot be collocated with 'one's own ideas'.
28	主张	坚持	maintain	Semantically, it cannot be collocated with 'one's own ideas'.
29	主张	维护	maintain	Semantically, it cannot be collocated with 'one's rights'.
30	转	推	shift	Semantically, it cannot be collocated with 'responsibility'.
31	转变	改变	change	Semantically, it cannot be collocated with 'living environment'.
32	转换	改变	change	Semantically, it cannot be collocated with 'working environment'.
33	祝愿	祝	wish	Semantically, it can be collocated with 'good health'; but in the context, this collocation is improper.
34	注重	关注	pay attention to	Semantically, it cannot be collocated with 'issue or contradiction'.
35	注意	关注	pay attention to	Semantically, it can be collocated with 'others' pains'; but in the context, this collocation is improper.
36	注意	关心	care for	Semantically, it can be collocated with 'academic performance'; but in the context, this collocation is improper.
37	注意	专心	concentrate on	Semantically, it cannot be collocated with 'learning'.
38	招来	得到	attract	It can govern the type of complements — 'emotions or attitudes', but only negatives ones, not positive ones.
39	追	追求	pursue	Semantically, it cannot be collocated with 'economic development, career or profits'.
40	撞到	碰到	run into	In the context, its seme [collide with] does not agree with the object 'eye'.

41	装满	充满	fill up	Semantically, it cannot be collocated with 'smell of cigarettes'.
42	滋长	形成	develop	It can govern complements with a derogatory sense — 'malpractice, bad habit, arrogance, or crimes', but not those with a positive sense.
43	组织	组建	organize	Semantically, it can be collocated with 'a family'; but in the context, this collocation is improper.
44	组成	营造	form	Semantically, it cannot be collocated with 'atmosphere'.
45	组成	构建	form	Semantically, it can be collocated with 'a wonderful world'; but in the context, this collocation is improper.
46	走	走路	walk	It is not proper in the context 'a baby learns to walk'.
47	走入	步入/ 进入	step into, enter into	Semantically, it cannot be collocated with 'modernisation or a scientific age'.
48	遵守	保持	keep	Semantically, it cannot be collocated with 'lifestyle'.
49	遵守	坚持	stick to	Semantically, it cannot be collocated with 'an idea or opinion'.
50	作出	得出	work out	Semantically, it cannot be collocated with 'different views'.
51	作出	制定	make	Semantically, it can be collocated with 'measures'; but in the context, this collocation is improper.

3.4 Discussion

In this section the results of the above analysis which have implications for the information structure of a Chinese valency dictionary are firstly discussed. It then compares the valency patterns of misused words with those of English equivalents in order to trace the influence of negative transfer of valency structure on lexical use, which may shed light on the construction of a contrastive bilingual valency dictionary.

From Tables 4, 5 and 6, it can be seen that the majority of cases of misuse occur in the dimension of semantic-pragmatic valency, taking up 85%. 11.7% of vocabulary misuse cases occur in the dimension of syntactic valency and only 3.3% in the dimension of logical-semantic valency. As argued by Han (1993), logical-semantic valency of a certain concept in different languages is generally the same because of people's shared experience. That is, the semantic roles of complements presupposed by the governing verb in different languages are nearly the same, or the governing verb has the same valency number in different languages. Thus, it is easy to explain that there are only two cases of misuse involving logical-semantic valency.

Different languages have different grammatical or syntactic systems. For

example, Chinese words lack inflections and derivations and the word order is relatively flexible, which is in stark contrast to English vocabulary. Consequently, there might be distinct differences between Chinese and English syntax. However, as the study shows, there are only seven cases of vocabulary misuse caused by improper application of syntactic patterns. One possible reason is that the language proficiency of these foreign learners who took part in HSK tests was not that high, so that they tended to use less complicated sentence structure. Another possible reason is that the study does not examine special constructions organized by structural auxiliaries such as *ba* (把, used to advance the object of a verb to the position before it), *bei* (被, used in the passive voice to introduce the doer of the action), *shi* (使, make/cause/have), *you ... you ...* (又 ... 又 ..., expressing the coexistence of several conditions or qualities), etc.

The bulk of the cases of misuse occurs in the semantic-pragmatic dimension. This type of misuse manifests a great diversity and can be attributed to the factors listed in Table 1 (The valency framework). Table 6 shows that most cases of semantic-pragmatic misuse can be ascribed to semantic collocation. Semantically, some headwords cannot govern particular complements. For instance, *zaodao* (遭到, encounter) cannot be collocated with 'troubles or difficulties', *zhiliao* (治疗, cure) cannot govern 'problems', and the object of *zhuangman* (装满, be filled with) should not be 'smell of cigarettes'.

Another leading factor is context, which gives rise to a number of cases of misuse. Some words, without contextual restrictions, can be in collocation with a range of words. However, in a particular context, the range of collocation is restricted. For example, in a general context, it is appropriate to say *Please zhuyi* (注意, pay attention to) *the pains of patients*. Nonetheless, in the sentence *I hope people from all walks of life can pay attention to the pains of patients*, the word *guanzhu* (关注, pay attention to) is more appropriate. Moreover, generally the expression *zhankai* (展开, spread) *arms* makes sense; but in the situation of helping children to walk with one's hands, the sentence *My father zhankai his arms to support me to walk* is awkward in the light of common sense.

The third semantic-pragmatic factor is emotive variables. The semantic environment around the headword decides its collocation with its complements. For example, *zizhang* (滋长, develop) tends to govern complements with derogatory sense — 'malpractice, bad habit, arrogance or crimes', but not those with positive sense.

As for register or formality of words, there are two cases of misuse. For instance, *zhixi* (知悉, know) is a word with high degree of formality and it cannot be used in informal writing.

The last factor is text style, with only one case of misuse. *zhu* (住, stay) is usually a word that is adopted in literary writing like poems. Therefore, it is stylistically improper to use this word in the sentence *His love zhu at my heart* in daily conversation.

In short, lexical misuse committed by foreign learners of Chinese can be analyzed from all three dimensions of the valency framework, with semantic-pragmatic errors being the most frequent. These results can be re-organized to be published as dictionary information. Specifically, the data gained from the native speaker corpus (CCL) can be presented as the basic information of a valency dictionary: quantitative valency, qualitative valency and valency patterns (Helbig and Schenkel 1969; Herbst et al. 2004). This treatment will be described and processed in detail in Figure 3, Section 4. The data gained from the learner corpus (HSK Corpus) can be arranged as tips or notes at the end of an entry so as to enhance learners' awareness of common lexical errors. This 'Note' block can be introduced by labels such as 'logical-semantic error', 'syntactic error' and 'semantic-pragmatic error'. For instance,

Note: Logical-Semantic Error It is incorrect to use '战胜' as an intransitive verb and add no object to it: ※ 比赛是为了提高自己, 并不是只为了战胜。The purpose of competition is to improve oneself rather than to win. '取胜', an intransitive verb, is more appropriate.

(Adapted from Example 4)

Note: Syntactic Error '赞扬' cannot be used in the syntactic pattern '对 + NP + 很 + VP': ※ 我对这种人很赞扬。I speak highly of this kind of person. '赞赏' would be appropriate in this pattern.

(Adapted from Example 5)

Note: Semantic-Pragmatic Error The objects of '转换' can be attitude, direction, topic, pattern or goal, not including environment: ※ 为了转换一下工作环境... In order to change working environment ... '改变' would be appropriate to govern the object 'environment'.

(Adapted from Example 7)

The possible causes for these cases of misuse are further discussed in the rest of this section by comparing the valency patterns of two misused words with those of English equivalents, which may help trace the influence of improper transfer of valency structure on lexical use. As noted, most contrastive valency studies are conducted among Indo-European languages such as German-English (Emons 2006, Mittmann 2007, Roe 2007), German-French (Plewnia 2006), German-Italian (Bianco 2006), German-Spanish (Fandrych 2006), German-Russian (Nübler 2006), German-Bulgarian (Baschewa 2006), German-Romanian (Stănescu 2006) and German-Polish (Schatte 2006). This study contributes to the few contrastive studies between Indo-European and Sino-Tibetan languages.

In Example 4, the English equivalent of the misused word *zhansheng* (战胜) is 'win'. According to the *Valency Dictionary of English* (Herbst et al. 2004: 948), the maximum number of valency complements of the headword is four and the minimum number is one, which means that 'win' can be used as an intransitive

verb. From Figure 1, the monovalent verb 'win' has two complementation patterns: NP + *win* and NP + be (expected, etc.) to + *win*:

- M He *won* at Oulton on 26 June, overtaking Andrews on the first lap. • The Christian Democrat-led government of Chancellor Helmut Kohl – who came to power in West Germany eight years ago – is widely expected to *win*.

Figure 1: Monovalent patterns of *win*

When composing a Chinese sentence which expresses the meaning — 'The purpose is to improve oneself rather than to win', the student may resort to his/her bilingual mental lexicon and retrieve the word *zhansheng* from his/her knowledge about Chinese. In the meantime, his/her inherent knowledge of the English word *win* is activated. Following the wrong association of the divalent Chinese verb *zhansheng* and the monovalent pattern of the English verb *win*, lexical misuse may occur.

In Example 7, the English equivalent of the misused word *zhuanhuan* (转换) is 'change'. According to VDE (Herbst et al. 2004: 122), this headword can govern at most four complements. From Figure 2, it can be seen that one possible object of divalent *change* is 'school', which indicates a particular environment or location:

- D1 + N_p
A Emotions can *change* the perception of pain.
Its flowers *change* colour as they mature.
B On this occasion she paused long enough to bathe her blistered feet and *change* her blouse.
C Spain is only a little over five hours from Paris, though travellers must *change* trains at the frontier.
D Ronnie didn't want to *change* schools, so he travelled with his dad to and from school every day even though it was a long way.
E So this morning we went to *change* some dollars in a shop.

Figure 2: Possible complements of the divalent verb *change*

In order to express 'to change my working environment' in Chinese, the examinee retrieves the word *zhuanhuan* and activates his/her knowledge about the semantic properties of *change*. The juxtaposition of *zhuanhuan* (whose list of objects excludes 'environment') and the semantic properties of *change* (which

empower the verb *change* to govern the object 'environment') may result in a pragmatic failure.

The above discussion shows that through the analytic lenses of the valency framework, the influence of improper transfer of the valency structure from one language to another can be revealed. This inspires the idea of a bilingual Chinese–English dictionary that introduces the differences and similarities between Chinese and English lexical items in different aspects. This idea is briefly represented in Table 7 and may be worthy of further research, which needs larger corpus data support.

Table 7: A brief design of a contrastive valency entry of '转换' and 'change' as divalent verbs

转换 \square (divalent)	change \square (divalent)
+ NP: [Mind: idea, perception, attention, attitude, topic, focus, goal, etc.] [Pattern] [Direction or Location] etc.	+ NP: [Mind: idea, perception, attention, attitude, topic, focus, etc.] [Pattern] [Clothes] [Transportation Tools] [Direction, Location, Destination or Environment] [Currency] etc.
Note: <u>Semantic-Pragmatic Error</u> The objects of '转换' can be attitude, direction, topic, pattern or goal, not including environment: ※ 为了转换一下工作环境... In order to <u>change</u> working environment ... '改变' would be appropriate to govern the object 'environment'.	

4. A valency dictionary for foreign learners of the Chinese language

With the analysis in Section 3, different levels of lexical misuse emerge from the data. The question arises: how to help foreign learners of Chinese to avoid making such mistakes. The answer may partly lie in a valency dictionary that systematically provides syntactic-semantic-pragmatic information of words, which might enhance learners' language proficiency. This section first introduces the basic elements of a Chinese valency dictionary for foreign learners and then designs one valency entry with a brief profile of prospective users.

Similar to German and English valency dictionaries, the Chinese valency dictionary has three basic elements: quantitative valency, qualitative valency and valency patterns.

Quantitative valency relates to the 'number of complements required for the verb to occur in an acceptable sentence' (Herbst et al. 2004: x). The complement inventory incorporates both obligatory and optional complements. Another quantitative feature of the dictionary is that it introduces the concept of 'probabilistic valency' (Herbst et al. 2004; Liu 2009). Labels, such as *rare*, *freq.* and *very freq.*, are used to indicate the frequency of valency patterns based on the CCL corpus.

Qualitative valency describes the characteristics of complements. Both obligatory and optional complements' semantic roles and properties are elaborated. Optional complements are placed in parentheses '()' to be distinguished from obligatory complements. Furthermore, a 'Note' block is provided at the end of the entry to warn learners against the common valency errors that might be committed in encoding tasks.

Valency patterns specify the syntactic structure of the governing word and its complements. Patterns are displayed in the form of combinations of phrases or clauses, such as NP + ~ + NP, substantiated by concrete examples. Furthermore, those adjuncts that can help exemplify a complete sentence structure are also included and placed in angle brackets '< >'.
Taking *zhaolai* (招来) from Example 9 as the entry word, these three elements of a valency dictionary are organized as follows:

招来 zhāolái 动词

① 吸引、引来或引进 to employ, recruit, attract, solicit or call in

I. 价数 Maximum valency: (a) + b = 2

II. 语义角色 Semantic roles

a Agent b Patient

III. 语义成分 Semes

a [Human] [Entity] [Thought]

b [Human: worker, talented human resources, student, soldier, reader, investor, customer etc.] [Animal] [Entity: vehicle]

IV. 句法模式 Syntactic patterns

i. (NP) + ~ + NP (*very freq.*): 1879年他从世界各地~实验员。In 1879 he employed experimenters from all over the world. ◇ 你看你能~学生不? Do you think you are able to recruit prospective students? ◇ 新的运营模式~了很多顾客。The new business model attracted many customers. ◇ 军警迅速~一辆救护车。The military police quickly called in an ambulance.

ii. ~ + 的 + NP ('招来', when followed by structural auxiliary '的', becomes an *adj.* used to modify a NP; '招来' is the governor of NP): ~的朋友 ◇ ~的技术人员

iii. NP + 是 + <Prep N_{place}> + ~ + 的 (*rare*): 战士多数是<从农村>~的。Most soldiers were conscripted from the countryside.

iv. 把 + NP + ~ + 了 (*rare*): 把工人~了。We have employed the workers.

② 引起或招致 to incur or bring

I. 价数 Maximum valency: (a) + b = 2

II. 语义角色 Semantic roles

a Agent b Affected

III. 语义成分 Semes

a [Human] [Behaviour] [Thought]

b [Negative comments, emotions, attitudes, reactions etc.] [Unfavourable things] [Resistance]

IV. 句法模式 Syntactic patterns

i. (NP) + ~ + NP (*very freq.*): 广告~了一片异议。The advertisement stirred up objections. ◇ 项目的失败~了精英们的冷嘲热讽。The failure of the project has incurred sarcastic remarks. ◇ 要忍耐, 以免~祸患。Be patient to avoid disasters. ◇ 好名声也给他~了许多麻烦。Reputation also brought him troubles. ◇ 破坏绿林会~大自然的惩罚。Deforestation may incur punishment from nature. ◇ 军队因供给不足, ~了反攻。Due to the shortage of supply, our army suffered a counterattack.

ii. ~ + 的 + 是 + NP (*rare*) ('招来', when followed by structural auxiliary '的', becomes a *noun*; '招来' is the governor of the NP after the linking verb '是'): ~ 的则是一堆愤怒的谴责。What we met was a heap of condemnation.

③ 习语 idioms: ~挥去 to call in and send away sb. at will ◇ 从实~ to make a clean breast of everything

Note: Semantic-Pragmatic Error '招来' cannot govern objects with 'positive or favorable' meanings: ※ 流行歌曲招来人们的喜欢和支持。Popular songs attract people's fondness and support. '得到' would be appropriate.

Figure 3: A design of the valency entry

It should be noted that this design of a valency entry is an example based on the results of data analysis in Section 3. While the data analyzed is a limited range of sample concordance lines retrieved from the corpora, it cannot exhaust all the possible valency patterns. This Chinese valency dictionary in a bilingualized form is mainly intended for advanced foreign learners of the Chinese language. The dictionary can be used in the setting of classroom teaching and learning or in self-study. It is expected that learners use the dictionary to retrieve relevant information to write in a correct and idiomatic manner. It is also expected that learners read the 'Instructions to Users' and finish corresponding exercises in the front matter for the sake of efficient and effective use of the dictionary.

5. Conclusion

Lexical misuse occurs on different levels in foreign language acquisition and can be analyzed in terms of the three dimensions of the valency framework — logical-semantic, syntactic and semantic-pragmatic. Following the analysis, this study found that 85% of the cases of misuse involve semantic-pragmatic valency, 11.7% involve syntactic valency and only 3.3% involve logical-semantic valency. In addition, there are diverse factors responsible for lexical misuse in the semantic-pragmatic dimension, among which semantic collocation and contextual factors are most influential. Other pragmatic factors include emotive variables, text styles and registers. These cases of lexical misuse and analytical results discovered from authentic corpus material are largely lacking in treatment in popular Chinese learners' dictionaries, and thus need to be properly represented and foregrounded in CLDs. The study hence designed a valency entry that includes all these information types to aid foreign learners' acquisition of Chinese words.

There are four suggestions for further research on valency dictionaries and Chinese vocabulary learning. Firstly and primarily, in addition to verbs, it might be important to study the valency of other lexical classes such as nouns, adjectives, adverbs, prepositions, numerals and quantifiers so that the Chinese valency dictionary can provide comprehensive information concerning all classes of Chinese vocabulary. Secondly, pragmatic valency still has room for further studies. Most of previous valency studies focus on syntactic and semantic aspects. Although the present study found that various factors contribute to lexical misuse in semantic-pragmatic valency, they are not scrutinized completely and thoroughly and there might be other unknown pragmatic factors. Thirdly, as suggested in Section 3.4, the research project of a contrastive Chinese–English valency dictionary could be promoted to help learners discern lexical errors caused by interlingual transfer of valency structure. Finally, it is suggested that empirical studies be conducted to verify the usability of Chinese valency dictionaries. The feedback from user surveys may help improve the design and compilation of Chinese valency dictionaries.

Acknowledgements

We are grateful to two anonymous reviewers and the editor of *Lexikos* for their valuable suggestions and comments. The current research was funded by the Zhejiang Provincial Planning Project of Philosophy and Social Sciences (No. 17NDJC224YB) and the New Key Professional Think Tank of Zhejiang Province — the Zhejiang Gongshang University Cultural Industry Think Tank.

References

- Ágel, V. et al. (Eds.). 2006. *Dependenz und Valenz: ein internationales Handbuch der zeitgenössischen Forschung*. Berlin/New York: Walter de Gruyter.

- Bareggi, C.** 1989. Gi Studenti e il Dizionario: Un'inchiesta presso gli Studenti di Inglese del Corso di Laurea in Lingue e Letterature Straniere Moderne della Facoltà di Lettere di Torino. *Prat Zagrebelsky, M.T. et al. (Eds.). 1989. Dal Dizionario ai Dizionari. Orientamento e Guida all'uso per Studenti di Lingua Inglese: 155-179.* Torino: Tirrenia Stampatori.
- Baschewa, E.** 2006. Kontrastive Fallstudie: Deutsch–Bulgarisch. Ágel, V. et al. (Eds.). 2006: 1229-1233.
- Béjoint, H.** 2002. *Modern Lexicography: An Introduction.* Beijing: Foreign Language Teaching and Research Press.
- Bianco, M.T.** 2006. Kontrastive Fallstudie: Deutsch–Italienisch. Ágel, V. et al. (Eds.). 2006: 1187-1196.
- Bondzio, W.** 1978. Abriß der semantischen Valenztheorie als Grundlage der Syntax. III. Teil. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 31(1): 21-33.
- Chinese Proficiency Center Beijing Language and Culture University (Ed.).** 2000. *A Dictionary of Chinese Usage: 8000 Words.* Beijing: Beijing Language and Culture University Press.
- Cowie, A.P.** 2002. *English Dictionaries for Foreign Learners: A History.* Beijing: Foreign Language Teaching and Research Press.
- De Groot, A.W.D.** 1949. *Structurele Syntaxis.* Den Haag: Servire.
- Emons, R.** 2006. Contrastive Case Study: Predicates in English and German. Ágel, V. et al. (Eds.). 2006: 1170-1176.
- Fandrych, C.** 2006. Kontrastive Fallstudie: Deutsch–Spanisch. Ágel, V. et al. (Eds.). 2006: 1197-1206.
- Fillmore, C.J.** 2008. A Valency Dictionary of English. *International Journal of Lexicography* 22(1): 55-85.
- Gao, J. and H.T. Liu.** 2019. Valency and English Learners' Thesauri. *International Journal of Lexicography* 32(3): 326-361.
- Han, W.H.** 1993. Basic Concepts and Research Methods of Valency Theory. *Journal of Tianjin Foreign Studies University* First Issue: 22-32.
- Han, W.H.** 1997. Views and Differences Concerning the Essential Issues from the Main Schools of German Valency Theory. *Linguistics Abroad* 3: 12-20, 31.
- Han, W.H. and Y.X. Han.** 1995a. The Theoretical Basis and Compilation Program of A Contrastive German–Chinese Valency Dictionary of Verbs. *Journal of Tianjin Foreign Studies University* 1: 12-20.
- Han, W.H. and Y.X. Han.** 1995b. A Sequel to the Theoretical Basis and Compilation Program of a Contrastive German–Chinese Valency Dictionary of Verbs. *Journal of Tianjin Foreign Studies University* 2: 17-27.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie. Volume I: 102-111.* Berlin: Walter de Gruyter.
- Hartmann, R.R.K.** 2005. *Teaching and Researching Lexicography.* Beijing: Foreign Language Teaching and Research Press.
- Helbig, G.** 1992. *Probleme der Valenz- und Kasus theorie.* Tübingen: Niemeyer.
- Helbig, G. and W. Schenkel (Eds.).** 1969. *Wörterbuch zur Valenz und Distribution deutscher Verben.* Leipzig: VEB Bibliographisches Institut.
- Herbst, T. and K. Götz-Votteler.** 2007. *Valency: Theoretical, Descriptive and Cognitive Issues (Trends in Linguistics Studies and Monographs 187).* Berlin/New York: Mouton de Gruyter.

- Herbst, T., D. Heath, I.F. Roe and D. Götz (Eds.).** 2004. *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Berlin/New York: Mouton de Gruyter.
- Hui, Y. (Ed.).** 2004. *A New Century Chinese-English Dictionary*. Beijing: Foreign Language Teaching and Research Press.
- Kacnel'son, S.D.** 1948. O Grammatičeskio Kategorii. *Vestnik Leningradskogo Universiteta, Serija Istorii, Jazyka i Literatury* 2: 114-134.
- Kacnel'son, S.D.** 1988. Zum Verständnis von Typen der Valenz. *Sprachwissenschaft* 13: 1-30.
- Liu, H.T.** 2009. *Dependency Grammar: From Theory to Practice*. Beijing: Science Press.
- Lu, J.M.** 1997. Valency Grammatical Theory and Teaching Chinese as a Second Language. *Chinese Teaching in the World* 1: 3-13.
- Lu, J.J. and W.H. Lv (Eds.).** 2007. *Commercial Press Learners' Dictionary of Contemporary Chinese*. Beijing: The Commercial Press.
- Lyons, J.** 2000. *Linguistic Semantics: An Introduction*. Beijing: Foreign Language Teaching and Research Press.
- Mei, X.J.** 2003. Valency Grammar and the Compilation of Modern Chinese Learners' Dictionaries. *Lexicographical Studies* 5: 39-46.
- Mittmann, B.** 2007. Contrasting Valency in English and German. Herbst, T. and K. Götz-Votteler (Eds.). 2007: 271-286.
- Neubach, A. and A.D. Cohen.** 1988. Processing Strategies and Problems Encountered in the Use of Dictionaries. *Dictionaries* 10: 1-19.
- Nübler, N.** 2006. Kontrastive Fallstudie: Deutsch-Russisch. Ágel, V. et al. (Eds.). 2006: 1207-1213.
- Partington, A.** 1998. *Patterns and Meanings*. Amsterdam/Philadelphia: John Benjamins.
- Plewnia, A.** 2006. Kontrastive Fallstudie: Deutsch-Französisch. Ágel, V. et al. (Eds.). 2006: 1177-1186.
- Roe, I.** 2007. Valency and the Errors of Learners of English and German. Herbst, T. and K. Götz-Votteler (Eds.). 2007: 217-228.
- Růžička, R.** 1978. Three Aspects of Valence. Abraham, W. (Ed.). 1978. *Valence, Semantic Case, and Grammatical Relations*: 47-54. Amsterdam: John Benjamins.
- Schatte, C.** 2006. Kontrastive Fallstudie: Deutsch-Polnisch. Ágel, V. et al. (Eds.). 2006: 1214-1216.
- Schumacher, H. (Ed.).** 1986. *Verben in Feldern, Valenzwörterbuch zur Syntax and Semantik deutscher Verben*. Berlin/New York: Walter de Gruyter.
- Shao, J.** 2002. Valency Theory and Lexical Teaching of Chinese as a Foreign Language. *Language Teaching and Linguistic Studies* 1: 43-49.
- Shen, Y.** 2000. *Valency Theory and Studies on Chinese Grammar*. Beijing: Language and Culture University Press.
- Shen, Y. and D.O. Zhen (Eds.).** 1995. *Studies on Valency Grammar of Modern Chinese*. Beijing: Beijing University Press.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.** 1996. The Search for Units of Meaning. *Textus* 9(1): 75-106.
- Sommerfeldt, K.-E. and H. Schreiber (Eds.).** 1974. *Wörterbuch zur Valenz und Distribution deutscher Adjektive*. Leipzig: VEB Bibliographisches Institut.
- Sommerfeldt, K.-E. and H. Schreiber (Eds.).** 1977. *Wörterbuch zur Valenz und Distribution der Substantive*. Leipzig: VEB Bibliographisches Institut.
- Stănescu, S.** 2006. Kontrastive Fallstudie: Deutsch-Rumänisch. Ágel, V. et al. (Eds.). 2006: 1234-1243.

- Sun, D.J. (Ed.).** 2006. *A Study on Teaching Chinese Vocabulary as a Foreign Language*. Beijing: The Commercial Press.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-making*. Oxford/New York: Oxford University Press.
- Tesnière, L.** 1959. *Éléments de Syntaxe Structurale*. Paris: Klincksieck.
- Wei, N.X.** 2002. General Approaches to Studying Semantic Prosody. *Foreign Language Teaching and Research* 34 (4): 300-307.
- Wiegand, H.E. (Ed.).** 1998. *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Tübingen: Max Niemeyer.
- Wu, W.Z.** 1996. A Review of Studies on Chinese Verb Valency. *Journal of Sanming University* 2: 1-11.
- Xie, H.J., X.Y. Zheng and L.P. Zhang.** 2015. *An Investigation of and Study on Chinese Dictionaries for Foreign Learners*. Beijing: Foreign Language Teaching and Research Press.
- Xu, X.H.** 2012. Lexicographical Definitions and Exemplifications of Bivalent Nouns. *Lexicographical Studies* 4: 32-37.
- Yuan, Y.L.** 2010. *A Study on Chinese Valency Grammar*. Beijing: The Commercial Press.
- Zhan, W.D.** 2000. Valency-Based Chinese Semantic Dictionary. *Applied Linguistics* 1: 37-43.
- Zhang, S.J.** 2007. The Construction of Synonym Discrimination Information in Foreign-oriented Chinese Learners' Dictionary. *Foreign Language Research* 3: 94-97.
- Zhang, Y.H.** 2011. *Studies on Lexicographic Definitions for Foreign-oriented Chinese Learners' Dictionaries*. Beijing: The Commercial Press.
- Zhou, G.G.** 2011. *A Study on Valency Grammar of Contemporary Chinese*. Beijing: Higher Education Press.

Special Field and Subject Field Lexicography Contributing to Lexicography

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)*

Abstract: Restricted dictionaries are fully-fledged dictionaries and their contribution to lexicography should never be underestimated. Because restricted dictionaries often are neglected in lexicographic discussions this article emphasises the significance of their role as members of the lexicographic family. Within a comprehensive dictionary culture the focus should not only be on dictionaries dealing with languages for general purposes but also on dictionaries in which languages for special purposes are treated. This paper firstly offers some terminological clarity and distinguishes between subject field dictionaries and special field dictionaries. The user-perspective is then discussed before it is shown how aspects of a general theory of lexicography also prevail in these dictionaries. This applies among others to the subtypological classification as well as different lexicographic functions. Using a dictionary from each of the categories of subject field and special field dictionaries it is indicated how dictionary structures are employed and further developed in an innovative way. Attention is given to structures like the article structure and the frame structure and to a transtextual approach in monolingual dictionaries with a bilingual dimension. The focus in the discussion of the subject field dictionary is on different aspects of the macrostructure. An explanation is given of double-layered sublemmata and it is shown how integrated macrostructures are employed in this dictionary. It is indicated how this section of the lexicographic practice can enrich the field of metalexicography and dictionary research.

Keywords: BACK MATTER TEXTS, DOUBLE-LAYERED SUBLEMMATA, FIRST LEVEL SUBLEMMATA, GUIDING ELEMENT, INTEGRATED MACROSTRUCTURE, LEARNER'S DICTIONARY, PARTIAL ARTICLE STRETCHES, SPECIAL FIELD DICTIONARY, SPECIALISED DICTIONARY, SUBJECT FIELD DICTIONARY, SUBLEMMATA, TRANSTEXTUAL APPROACH, USER-PERSPECTIVE

Opsomming: Spesialeveldleksikografie en vakleksikografie lewer 'n bydrae tot leksikografie. Beperkteveldwoordeboeke is volwaardige woordeboeke en hulle bydrae tot die leksikografie mag nie onderskat word nie. Aangesien beperkteveldwoordeboeke dikwels in leksikografiese gesprekke afgeskeep word, beklemtoon hierdie artikel die belang van hulle rol as lede van die woordeboekfamilie. In 'n omvattende woordeboekkultuur moet die fokus nie net op woordeboeke vir taal vir algemene doeleindes wees nie, maar ook op woordeboeke waarin die taal vir spesifieke doeleindes bewerk word. Dié artikel bied eerstens 'n terminologiese vereenduidiging en onderskei tussen vakwoordeboeke en spesialeveldwoordeboeke. Daarna word die gebruikersperspektief bespreek voordat daar aangetoon word hoe aspekte van die algemene leksikografie ook in hierdie woordeboeke ter sake is. Dit geld onder meer aspekte van die subtipologiese verde-

ling asook verskillende leksikografiese funksies. Aan die hand van 'n lid van elk van die kategorieë van vak- en spesialeveldwoordeboeke word daar gewys hoe veral woordeboekstrukture op 'n innoverende manier in hierdie woordeboeke benut en aangepas word. Aandag word onder meer gegee aan die artikelstruktuur en die raamstruktuur asook 'n transtekstuele benadering in eentalige woordeboeke met 'n tweetalige dimensie. Die fokus in die bespreking van die vakwoordeboek is veral op verskillende aspekte van die makrostruktuur. 'n Beskrywing word gebied van dubbelvlak-kige sublemmata. Daar word ook gewys hoe geïntegreerde makrostrukture in die betrokke woordeboek gebruik word. Daar word aangedui hoe vanuit hierdie afdeling van die leksikografie-praktyk die terrein van die metaleksikografie en woordeboeknavorsing verruim kan word.

Slutelwoorde: AANLEERDERWOORDEBOEK, AGTERTEKSTE, DEELARTIKELTRAJEKTE, DUBBELVLAKSUBLEMMATA, EERSTEVLAKEBLEMMATA, GEBRUIKERSPERSPEKTIEF, GEÏNTEGREERDE MAKROSTRUKTURE, GIDSELEMENT, SPESIALEVELDWOORDEBOEK, SUBLEMMATA, TRANSTEKSTUELE BENADERING, VAKWOORDEBOEK

1. Introduction

Kilgarriff (2012: 27) says: "When ordinary people refer to dictionaries, they mean general language dictionaries like the Oxford English Dictionary, Le Grand/Petit Robert, Duden, Webster, etc." He criticises an approach where a quantitative comparison between general language dictionaries and what he calls special language dictionaries is used to indicate that many dictionaries do not have language as their object. According to Kilgarriff such a comparison:

"is like noting that there are more local airstrips than international airports in the world, so basing an account of aviation on local airstrips. Numbers of publications alone do not give a good overall picture, and I remain convinced that general language dictionaries are central to the lexicographical firmament."

This central position of general language dictionaries, and to regard them as default or prototypical dictionaries, especially due to their high frequency of use, cannot and should not be disputed. However, one also has to acknowledge the existence of numerous dictionaries that are not general language dictionaries, e.g. those dealing with the terminology of different subject fields. They do not primarily focus on language and do not have the same frequency of use as general language dictionaries, but this does not exclude them from the domain of lexicography. It would be an injustice to underestimate their value as lexicographic products and their contribution to the development of the broad field of lexicography.

According to Wiegand (1989: 251) lexicography is a practice, aimed at the production of dictionaries in order to initiate another practice, i.e. the cultural practice of dictionary use. The lexicographic practice gives evidence of a variety of dictionary types with different aims, objectives, functions and users. The lexicographic practice is comprehensive and all types of dictionaries fall within its scope. This comprehensive nature does not only apply to the lexicographic

practice but also to the theoretical component of lexicography, i.e. metalexicography and dictionary research. Metalexicography and dictionary research are concerned with dictionaries and not only with general language dictionaries. Dictionary research is directed at all types of dictionaries. In this regard Schierholz (2003: 10) says that the total of all theories directed at lexicography and dictionaries as well as the scientific practice are regarded as dictionary research.

Lexicography, both the practice and the theory, contains a variety of sub-domains, including domains that have their focus on e.g. general language dictionaries, subject-specific dictionaries or dictionaries focusing on a specific data type, e.g. pronunciation dictionaries or spelling dictionaries.

This paper aims to show some mutual aspects of special field lexicography and general language lexicography. This can be illustrated by means of typological diversity but it can also be seen in the way in which lexicographic functions and different dictionary structures come to the fore in these different domains of the lexicographic practice. Bergenholtz (1995: 53) aptly captures the approach to be followed in this paper, when he says that lexicography has both general language and the language of subject fields in its scope.

2. Towards terminological clarity

2.1 Terminography and special field lexicography

In the fields of lexicography and terminology there are different points of view regarding the use of the terms *terminography* and *special field lexicography* and whether they should be regarded as synonyms or not, cf. the well-balanced discussion in Bergenholtz (1995). This distinction will not be discussed in the present paper. The term *terminography* will not be used here. Instead, the terms *special field lexicography* and *subject field lexicography* will be used — but not as synonyms. In the subsequent paragraph the less precise terms *specialized dictionaries* and *specialized lexicography* will also be used, albeit that this use will then be discarded.

2.2 Specialised lexicography versus general lexicography

In the English lexicographic terminology it has been a problem to find an unambiguous term for that section of lexicography that does not have general language as its object. Terms like *special language dictionaries*, *subject-specific dictionaries*, *special field dictionaries*, *specialised dictionaries* and *technical dictionaries* have been used. Bergenholtz and Tarp's *Manual of Specialised Lexicography* gave a firm footing to the term *specialised dictionaries/lexicography*. This term is also used in publications like Schierholz (2003: 5), Fuertes-Olivera (2010), Jesenšek (2013) and in translating terms like *Fachwörterbuch* (*Specialized dictionary*) and *Fachlexikogra-*

phie (*Specialized lexicography*) in the *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research* (Wiegand et al. 2010–2020) as well as in the *Routledge Handbook of Lexicography* (Fuertes-Olivera 2018). In this handbook Humbley (2018: 317) explains the use and scope of this term as follows:

"Specialised dictionaries are defined here by the specialised nature of the subjects they treat, focusing on particular subject fields, professional practices or even leisure activities such as sports ..."

For the general dictionary user and even for the lexicographer it often remains unclear exactly what is meant by *specialised lexicography*. Does it only refer to dictionaries in which the lexicon of subject fields is treated or also to other dictionaries that do not have the full lexical stock of the general language as its object or do not present a treatment comparable to that of traditional general language dictionaries?

In his typological classification of dictionaries Zgusta (1971: 204) makes a distinction between "general dictionaries on the one side, and restricted (or special) dictionaries on the other". According to Zgusta the restriction lies therein that the lexicographer decides "that he will make his choice from only a certain part of the lexicon of the language." Zgusta argues that there is practically an endless number of different restrictions — reflecting any principle or combination of principles determined by the lexicographer of a dictionary. These include language varieties, slang, jargon, trades, crafts, sports, professional languages, etymology, loan words, abbreviations, etc.

In the comprehensive international encyclopedia of lexicography (Hausmann et al. 1989–1991) the German term *Spezialwörterbücher* is often used to refer to certain types of restricted dictionaries, e.g. in the table of contents of the second volume *Paradigmatische Spezialwörterbücher, Spezialwörterbücher zu markierten Lemmata* and *Spezialwörterbücher mit bestimmten Informationstypen*. The respective English translations of this term are completely unsatisfactory: *Paradigmatic dictionaries* (no attempt to translate the *Spezial-*), *Dictionaries dealing specifically with marked standard language entrywords* and *Dictionaries offering specific types of information*. English clearly presents no adequate equivalent for this occurrence of *Spezial-*.

The scope of Zgusta's term *restricted dictionary* could include dictionaries falling within the scope of both the German terms *Fachwörterbuch* and *Spezialwörterbuch* — something not achieved by either *specialised dictionary* or *special field dictionary*. In English a clearer terminological differentiation is needed. In this paper the terms *special field* and *subject field* will be used. *Subject field dictionary* will be used as an approximate equivalent of the German *Fachwörterbuch* and *special field dictionary* as an approximate equivalent of *Spezialwörterbuch*. The latter category includes the whole range of Zgusta's restricted dictionaries with the exception of dictionaries dealing with academic subjects. This implies that e.g. dictionaries of football terms, philately, idioms and pronunciation will be

regarded as special field dictionaries, whereas dictionaries of e.g. chemistry, linguistics and economics will be regarded as subject field dictionaries.

In the remainder of this paper the focus will be on various aspects of restricted dictionaries to illustrate that subject field and special field lexicography do not only fall within the scope of the overarching discipline of lexicography, but also that the practice of subject and special field lexicography contributes to the development of the broad domain of lexicography.

3. The metalexigraphic domain

In the emergence and early development of the discipline of metalexigraphic and dictionary research there had been a strong focus on general language dictionaries. This was to the detriment of special and subject field lexicography where insufficient in depth research into these domains prevailed, as indicated by Wiegand (1988). During the last decades the situation has improved — as is evident in research publications such as Tarp (1994; 2000), Schierholz (2003; 2013), Wiegand (2003; 2004; 2004a), Fuertes-Olivera and Arribas-Baño (2008), Fuertes-Olivera (2010), Jesenšek (2013), Humbley (2018). In various contributions in Hausmann et al. (1989–1991) different types of special field dictionaries have also been discussed. Gouws (2016: 107-109) argues in favour of a comprehensive dictionary culture that should have the full lexicographic spectrum, including both general language and special and subject field dictionaries, in its scope.

Albeit that research into and discussions about special and subject field dictionaries are not yet on equal par with that of general language dictionaries the contributions have shown that special and subject field lexicography belongs to the broader field of lexicography and these publications reflect a significant number of mutual issues between general language and special and subject field lexicography. Some of these aspects will be briefly referred to in the subsequent paragraphs.

4. The user-perspective and dictionary typology

Dictionaries are planned and compiled in response to the needs and reference skills of a well-defined target user group. In general language dictionaries it has been a longstanding approach that diverse target user groups demand different approaches and different dictionaries, even with regard to dictionaries belonging to the same broad category like monolingual or bilingual dictionaries. Bergenholtz and Tarp (2010: 28) indicate that terminographers and most lexicographers agree that terminography differs from lexicography in various ways including: "The target group of terminology is the expert, whereas in lexicography it is the layman." However, a look at the development in the field of lexicography clearly shows that when it comes to subject field lexicography as

executed by lexicographers this is not the case. Humbley (2018: 319) recognises different types of users of specialised dictionaries. These user groups are laypersons, educated laypersons and experts. Gouws (2016: 109) complements the classification of experts, semi-experts and laymen made by Bergenholtz and Tarp (1995: 19) with an additional group, namely the informed layperson. This is a layperson in a specific subject field but he or she has acquired some knowledge of that specific field. The user-perspective with regard to the fields of special and specialised lexicography also comes to the fore in modern-day metalexigraphical publications like Fuertes-Olivera and Arribas-Baño (2008), Fuertes-Olivera (2010) and Jesenšek (2013).

Cognizance of the user has also expanded the typological spectrum of subject field dictionaries. One of the major typological explosions in general language dictionaries has been the emergence and growth of pedagogical dictionaries, especially learner's dictionaries, during the last two decades of the previous century. Learner's dictionaries became a significant focal point in metalexigraphical discussions, cf. Lemmens and Wekker (1986), Cowie (1987), Hartmann (1992) Van der Colff (1996), Hollós (2004), Steyn (2004), Tarp (2004; 2008) and Runte (2015). This typological category has also found its way into the domain of subject field lexicography, once again cf. Fuertes-Olivera and Arribas-Baño (2008), Fuertes-Olivera (2010) and Jesenšek (2013). A growing number of subject field dictionaries have also been compiled for users belonging to the class of learners.

Learner, a more or less unambiguous term in general language lexicography, had also been employed in special and subject field lexicography where it acquired more than one sense. Gouws (2010) already referred to the ambiguity in the use of the word *learner* when discussing special and subject field dictionaries. In some of these dictionaries the target user is a learner of the specific subject or field, e.g. in the *Schüler Duden: Die Musik* (Kwiatkowski 1989). In other cases the user is a learner of the language of the dictionary who needs a specific subject field dictionary. An example of such a dictionary is the *Oxford Dictionary of Computing for Learners of English* (Pyne and Tuck 1996). As stated in its preface, this subject field dictionary has been compiled for users learning the English language. In some cases *learner* could refer to users with both these qualities. The *Ungarisch-Deutsches Deutsch-Ungarisches Fachwörterbuch zur Rentenversicherung* (Fata 2005) is a subject field dictionary but it is also intended to be a learner's dictionary used in institutes where German is taught as specialised foreign language.

The user and the subtypological nature of special and subject field dictionaries determine the contents. This is clear when comparing the treatment of the same term in a subject field dictionary for learners of the field and experts in the field. The *SASOL Science and Technology Resource* (Hartmann-Petersen et al. 2001), a subject field dictionary for school learners, defines the term *genome* as follows:

genome The sum-total of all of an individual's genes, and hence its entire chromosome set.

Fig. 1: SASOL Science and Technology Resource

The *Gene Technology Encyclopedic Dictionary* (Kaufmann and Bergenholtz 1998), a subject field dictionary aimed at the expert, defines the same term as:

genome *genoma* m
In eukaryotes the genome is the chromosomes or DNA present in a haploid cell, so if the chromosome number in the somatic cells of an organism is $2n$ and the amount of DNA is $2C$ then the genome comprises n chromosomes with an amount of DNA which is C . Sometimes the word is also used of all chromosomes or all DNA in a diploid cell, but in that case the expression diploid genome should be used. The genome of *E. coli*, i.e. the amount of DNA in one *E. coli* chromosome, has a size of $4 \cdot 10^6$ bp and the human genome has a size of about $3 \cdot 10^9$ bp. Mitochondria and chloroplasts possess small circular genomes encoding some of their components.
● The human genome, for example, are thought to contain fewer than 10^5 genes.
▲ an animal ~ *un* ~ *animal*; the host cell's ~ *el* ~ *de las células hospederas*; the human ~ *el* ~ *humano*; the mitochondrial ~ *el* ~ *mitochondrial*; the viral ~ *el* ~ *viral*
→ diploid; eukaryote; haploid
genome analysis *análisis* m *genómico*

Fig. 2: *Gene Technology Encyclopedic Dictionary*

5. Innovative improvements in the lexicographic practice

The interactive relation between theory and practice in lexicography implies that theory should guide the practice, but it should also be enriched by the practice. This applies to all domains of the lexicographic practice and to a theoretical approach underpinned by a comprehensive and all-inclusive dictionary culture. As a result special and subject field dictionaries need to be products of the application of a sound theoretical approach, and developments in lexicographic theory should also reflect changes initiated in an innovative lexicographic practice. In the remainder of this paper it will be shown how lexicographic theory has contributed to ensuring some good subject and special field dictionaries but also how advances in some of these dictionaries have resulted in improvements in the relevant lexicographic theory — where practice takes the lead and theory has to catch up.

Every dictionary should have an intended target user group and the lexicographic needs and reference skills of this group should determine the contents, functions and structures of the specific dictionary. As seen in the treatment of the term *genome* the extent of data presented in the item giving the paraphrase of meaning shows a vast difference that is determined by the needs of the target users. The lexicographic functions to be satisfied by a specific dictionary should also be determined in accordance with the user group of that dictionary. The *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research* (Wiegand et al. 2010–2020) is a subject field dictionary compiled for the expert and semi-expert. With regard to lexicographic functions, the typical user consults this dictionary with either a communicative need of text reception or translation or a cognitive need. The article of the lemma *sign expandiertes allgemeines Suchbereichsstrukturbild* (Figure 3) shows how the text reception function is achieved in the brief item giving a definition (presented in the first search zone) whilst a cognitive function is accounted for in the section introduced by the bullet (●), where additional data transfer and more encyclopedic explanations are presented. The translation function is satisfied by the listing of translation equivalents in the final search zone of the article. The variety of search zones in this article also indicates the way in which an article structure with an appropriate inner access structure gives evidence of a well-planned packaging of data. Theory is applied to practice.

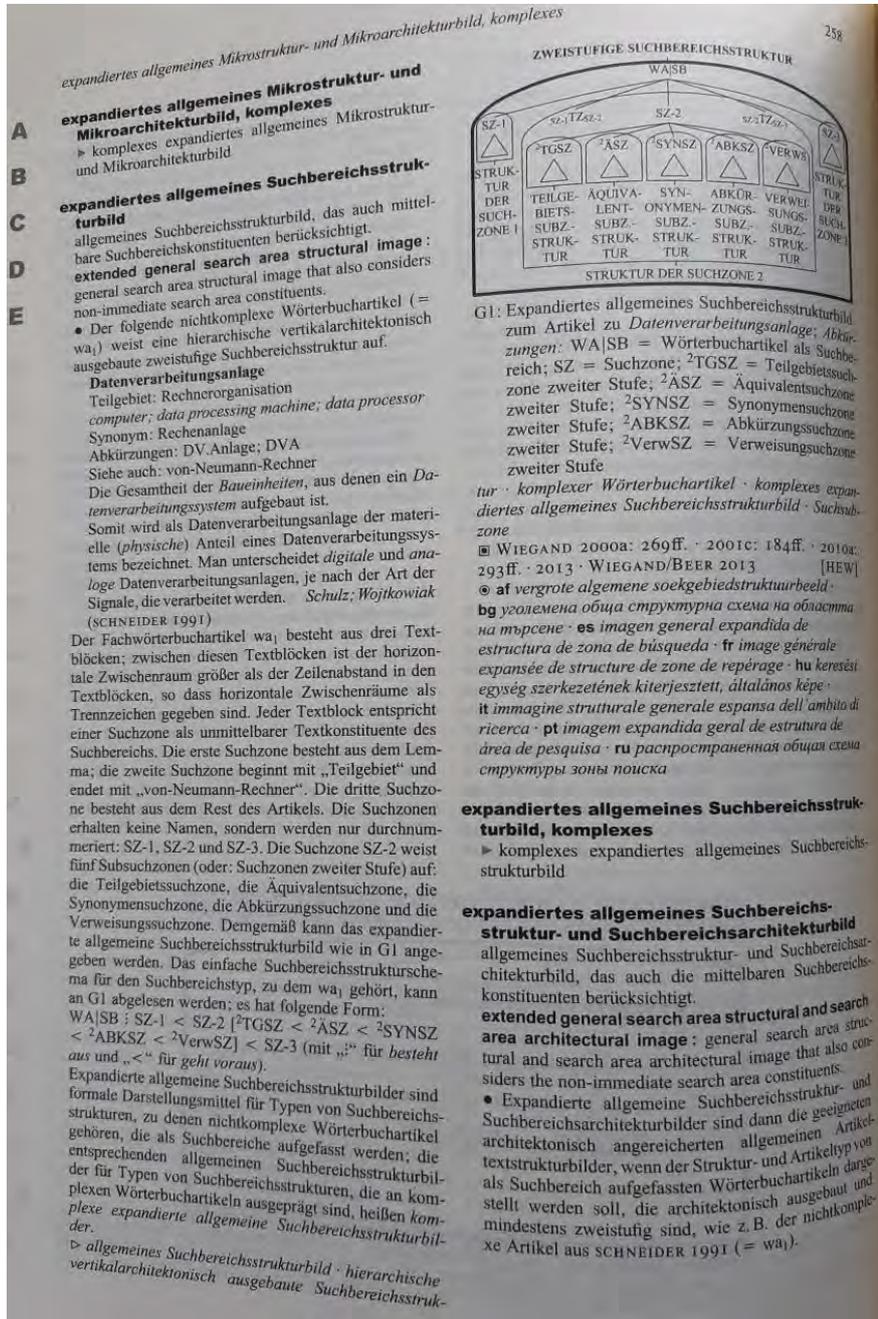


Fig. 3: Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research

In the subsequent sections it will be shown how the practice of special and subject field lexicography utilises insights from the field of dictionary research in an innovative way to enhance the quality of certain dictionary structures.

5.1 A special field dictionary

Idiomewoordeboek (De Villiers and Gouws 1988) is a special field dictionary that presents and treats a selection of Afrikaans idioms and fixed expressions. This dictionary illustrates that dictionaries dealing with a subsection of the lexicon of a given language can utilise and improve existing dictionary structures and can enrich the lexicographic practice as well as the field of metalexicography and dictionary research. Figure 4 displays a partial article stretch from this dictionary. Although idioms and fixed expressions are the primary treatment units, they are not the guiding elements of the articles in terms of the outer access structure of the dictionary. Due to practical problems in entering full idioms as guiding elements, this dictionary has opted for a system where a single core word from the expression is entered as guiding element in its alphabetical position. The idiom or fixed expression is entered, in bold, in a separate search zone. In partial article stretches with different articles of idioms and fixed expressions for which the same core word has been selected, there is a repeated occurrence of that core word as iterative guiding element. This can be seen in Figure 4 with regard to the core word *kat* (cat):

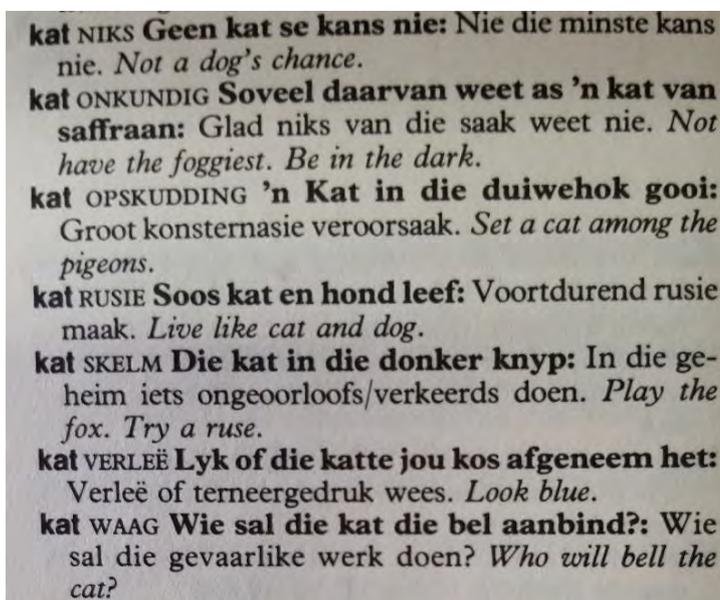


Fig. 4: *Idiomewoordeboek*

Figure 5 presents the article in which the idiom *soveel daarvan weet as 'n kat van saffraan* is treated:

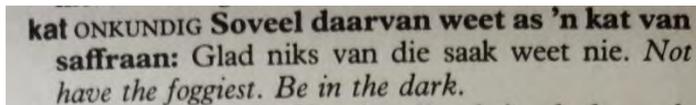


Fig. 5: *Idiomewoordeboek*

In terms of Wolski (1989: 365) the guiding element *kat* is a lemma part occurring in a lemma-external position, here specifically as guiding element in the article entrance position, albeit isolated from the full form of the lemma sign. Between the guiding element and the idiom functioning as primary treatment unit is the item *ONKUNDIG*. This item gives the semantic field to which the idiom belongs. The back matter section of this dictionary contains two texts that can assist the user to find idioms and expressions that fall within specific semantic fields. The first text gives the item indicating the semantic field as guiding element of an index article, as seen in Figure 6.

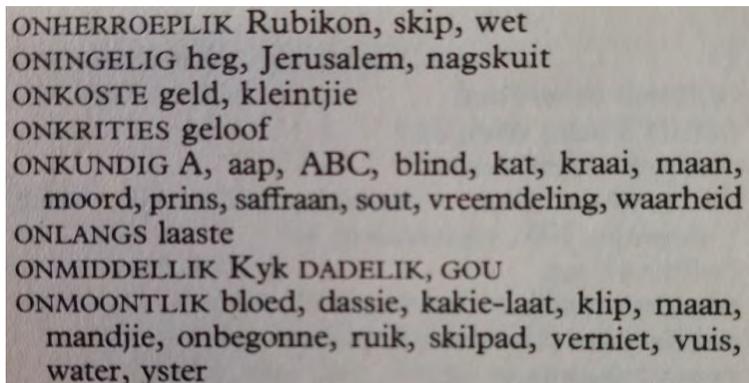
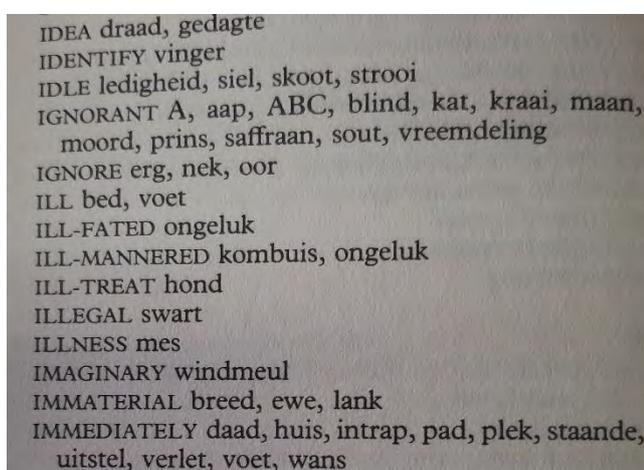


Fig. 6: *Idiomewoordeboek*

This is followed by a listing of words, that are guiding elements of articles in the main text, containing idioms and expressions that reflect a meaning falling within the specific semantic field. The article of the index lemma *ONKUNDIG* guides the user, among others, to the guiding element *kat* where the idiom *soveel daarvan weet as 'n kat van saffraan* is treated. Where a partial article stretch in the main text contains more than one idiom or fixed expression with the same guiding element, these articles are ordered according to the alphabetical value of the items giving the semantic field — in Figure 4 from *NIKS – WAAG*.

Idiomewoordeboek is a monolingual dictionary that has been partly bilin-

gualised (cf. Hartmann 1994). In the main text of the dictionary, the idiom or fixed expression is followed by a brief paraphrase of meaning in Afrikaans (in Figure 5 *Glad niks van die saak weet nie*). Then one or more items in English follow that present one or more equivalent idioms or fixed expressions in English, or one or more brief paraphrases of meaning (In Figure 5: *Not have the foggiest. Be in the dark*). This bilingual dimension is continued in the back matter section with a text that contains a list of English words representing the different semantic fields, functioning as guiding elements of the index articles. Figure 7 presents a partial article stretch from this back matter text.



IDEA draad, gedagte
IDENTIFY vinger
IDLE ledigheid, siel, skoot, strooi
IGNORANT A, aap, ABC, blind, kat, kraai, maan,
moord, prins, saffraan, sout, vreemdeling
IGNORE erg, nek, oor
ILL bed, voet
ILL-FATED ongeluk
ILL-MANNERED kombuis, ongeluk
ILL-TREAT hond
ILLEGAL swart
ILLNESS mes
IMAGINARY windmeul
IMMATERIAL breed, ewe, lank
IMMEDIATELY daad, huis, intrap, pad, plek, staande,
uitstel, verlet, voet, wans

Fig. 7: *Idiomewoordeboek*

The English item functioning as guiding element of the index article, e.g. *IGNORANT* in Figure 7, is followed by words directing the user to the relevant idioms or fixed expressions in the central text of the dictionary. These English items do not occur in the central list of the dictionary.

It is clear that the lexicographic process resulting in *Idiomewoordeboek* was devised in a planned and consistent way. In terms of the discussion of lexicographic processes in Wiegand (1998: 37-38) this dictionary adheres to significant criteria that characterise the lexicographic practice like calculability, analysability, verifiability, regulatability, learnability and testability.

With regard to adhering to and expanding the general theory of lexicographic theory, this dictionary employs and creatively enhances existing lexicographic structures and displays a variety of features accounted for in meta-lexicography as well as innovative adaptations. These innovations are seen in the macrostructural presentation, the article structure and the ordering of items within the articles where both alphabetical and semantic criteria play a

determining role, the frame structure, the transtextual use of a bilingual dimension and the poly-accessible nature of this dictionary.

5.2 A subject field dictionary

The structures of some subject field dictionaries also give evidence of the extent to which these restricted dictionaries are rooted in the overarching domain of lexicography and that they do not function alongside it. Existing structures are employed and often adapted in such an innovative way, that these changes activate further research and discussions in the field of metalexicography and dictionary research.

The South African *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences* (Brink and Lochner 2011), henceforth abbreviated as WGW, has experts and semi-experts in the field of the health sciences as its target user group. This dictionary displays a number of interesting macrostructural strategies.

5.2.1 Macrostructural strategies

5.2.1.1 Sublemmatisation

According to Wiegand and Gouws (2013: 78) the macrostructure of a dictionary is the textual structure that presents the ordering of all the elements that contribute to the macrostructural coverage of the specific dictionary. It is important to note that they refer to *all* elements of the macrostructural coverage. Bergenholtz, Tarp and Wiegand (1999) give a comprehensive discussion of the ordering of lemmata in subject field dictionaries. They also refer to the ordering of multiword terms as lemmata and specifically to their occurrence in non-grouped niches (Bergenholtz, Tarp and Wiegand 1999: 1817). The WGW also has to negotiate the inclusion and treatment of numerous multiword terms. Some of these terms are included as main lemmata in the default vertical ordering of the macrostructural presentation, but many are ordered in non-grouped clusters in either a nested or a niched format. According to Gouws (2012: 259) these clusters are often attached to the article of the preceding main lemma of which the lemma sign represents a term that also functions as the initial part of the multiword term. For space-saving reasons this first lemma part is given in an abbreviated format in the cluster, resulting in a partial lemma, cf. Bergenholtz, Tarp and Wiegand (1999: 1817), presented here as sublemma. Attached to the article of the lemma sign *nucleus* Figure 8 shows the condensed non-grouped niche of partial lemmata, presented as sublemmata: *n. arcuatus*, *n. basalis*, *n. caudatus*, *n. dentatus*, *n. puposus*, representing the multiword terms *nucleus arcuatus*, *nucleus basalis*, *nucleus caudatus*, *nucleus dentatus* and *nucleus puposus*.

nucleus:† *nucleus*. (L. *nucleus* neut) Latyns-afgeleide benaming wat dui op: **1.** die sentrale gedeelte van 'n anatomiese struktuur; of **2.** 'n begrensde versameling senuweeselle in die sentrale senuweestelsel wat grys van kleur is, en dus maklik makroskopies van die omliggende witstof van die brein onderskei kan word. Die meeste kerne is geassosieer met 'n bepaalde funksie of groep funksies. Die grysstof wat die korteks van die serebrale hemisfeer vorm, asook die grysstof wat die medulla van die rugmurg vorm, word tradisioneel nie "nuclei" genoem nie, ondanks die feit dat hul histologies en fisiologies nie van neurologiese kerne onderskei kan word nie. Kyk NUKLEUS.

n. arcuatus: *n. arcuatus*. (L. *arcuatus* boogvormig) Boogvormige grysstofkern in die intermediêre hipotalamus, o.a. betrokke by prolaktien- en groei-hormoonsekresie.

n. basalis: *n., basal*. (meerv. *nuclei basales*) Een van 'n groot verskeidenheid van spesifieke, ineengeskakelde versamelings van grysstof, gewoonlik bilateraal teenwoordig, diep in die serebrale hemisfeer en naasliggende gedeeltes van die breinstam. Voorbeelde hiervan is die corpus striatum (saamgroepering van die nucleus caudatus, putamen en globus pallidus), nucleus subthalamicus en substantia nigra, wat betrokke is by die beplanning, programmering en uitvoer van willekeurige liggaamsbewegings; ■ basale kern; basaalkern. (by **2**).

n. caudatus: *n. caudatus*. Kyk CORPUS, STRIATUM.

n. dentatus: *n., dentate*. Groot groep senuweeselle in die serebellum.

n. pulposus: *n. pulposus*. (L. *pulpa* pap) Halfvloeibare, gelatienagtige massa wat die sentrale gedeelte van die intervertebrale skyf van fibrokraakbeen vorm. Enigste oorblyfsel van die vrugrugmurg (notochorda) in die embryo. (by **1**).

Fig. 8: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Working with target users that are experts in this field and supposedly have already acquired good reference skills, these ordering and textual condensation procedures should be in order. To enhance the users' success in retrieving the required information the lexicographers also explain their lemmatisation procedures in the user guidelines text in the front matter section.

The strategy of including multiword terms as condensed partial sublemmata in non-grouped clusters is taken a step further in article nests that display a lemmatisation procedure where the nest is also attached to the article of a lemma sign that represents a term that is a final part of the multiword term. Figure 9 shows such a partial article stretch of the main lemma *sirkulasie* and an attached nest of partial sublemmata from *s., enterohepatiese* to *s., sistemiese*:

sirkulasie:* *circulation*. (L. *circulus* sirkel) Beweging van vog in 'n kringloop, soos in die geval van bloed, limf of serebrospinale vog.

s., enterohepatiese: *c., enterohepatic*. Kyk GAL, -SOUT.

s., kollaterale: *c., collateral*. (L. *cum* met + *latus* kant) Sirkulasie deur bykomstige, parallelverlopende bloedvate, veral ná afsluiting van die hoofvate. 'n Funksionele omleiding word dus daargestel.

s., mikro: *c., micro-*. (G. *mikros* klein) Die bloedperfusie deur klein bloed-

vaatjies, agtereenvolgens die arteriole, kapillêre en venules.

s., portale: *c., portal*. Veneuse terugvloei van die ingewande na die lewer.

s., pulmonêre: *c., pulmonary*. Bloedsomloop vanaf die regterhartkamers na die longe en terug na die linkerhartkamers.

s., sistemiese: *c., systemic*. Bloedsomloop deur die hele liggaam met die uitsondering van die portale en pulmonêre sirkulasie.

Fig. 9: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Each article in the nest has a guiding element that has an abbreviation of the preceding main lemma as first component. The comma after the abbreviated lemma part indicates that that lemma part functions as final component of the multiword term that has the subsequent word as first component. The condensed sublemma *s., enterohepatiese* should accordingly be read as *enterohepatiese sirkulasie*. This is an example of first level nesting, cf. Gouws (2001: 106), where the alphabetical ordering is maintained within the nest but not with regard to the surrounding main lemmata. In the guidelines to the dictionary this lemmatisation procedure is explained in non-metalexigraphic lay terms. It is employed to include sublemmata in close proximity of that main lemma that corresponds to the semantic core of the term presented by the sublemma. In the nested articles in Figure 9 *sirkulasie* would be the semantic core of the terms presented by the different sublemmata.

Taking cognizance of the diverse lemmatisation procedures in general language dictionaries the WGW does not only follow this example, but also employs other macrostructural strategies. In Figure 10 the main lemma is *nekrose* and the attached partial article nest contains the non-grouped condensed lemmata that have the partial lemma *n., akute tubulêre* as nest entrance lemma.

nekrose:* *necrosis; mortification*. (G. *nekrosis* afgestorwenheid; L. *mortifer* dodelik) Patologiese afsterwe van 'n sel, groep selle, afgebakende weefsel of 'n orgaan (-segment) in 'n lewende liggaam a.g.v. onomkeerbare skade berokken deur siektes, besering (chemies, fisiek), of iskemie. Indien 'n inflammatoriese respons intree om dooie materiaal op te ruim, eindig die proses dikwels in residuele fibrose en littekenvorming. Vgl. APOPTOSE

n., akute tubulêre (ATN): *n., acute tubular (ATN)*. 'n Kliniese beeld van akute nierversaking weens tubulêre selskade veroorsaak deur nefrotoksiese stowwe, renale iskemie, trauma weens vergruising, hipovolemiese skok, septisemie of uitgebreide brandwonde. Kyk NEFROTOKSIEN; NEFROTOKSIES; VERGRUIS, -SINDROOM; SKOK, HIPOVOLUMEIESE. Vgl. NEFROSE, LAERNEFRON-

n., aseptiese: *n., aseptic*. (G. *an neg.* + *sepsis* verrotting) Toenemende sklerotiese en sistiese veranderinge in die kop van die femur wat dikwels volg op traumatiese ontwinging van die heup. 'n Soortgelyke toestand kan ook in die kop van die humerus ontwikkel, asook in ander bene; ■ avaskulêre nekrose.

n., avaskulêre: *n., avascular*. (G. *an neg.* + L. *vasculum* vaatjie) Fokale infarksie van 'n skeletbeen sekondêr tot belemmerde bloedsirkulasie sonder kompli-

serende infeksie, bv. trauma (soos in dislokasie), etanolisme, prednisonmisbruik, sekelsiesiekte, bestraling, maligniteit, barotrauma, ens.;

■ aseptiese nekrose. Kyk **OSTEONEKROSE**, **PREDNISOON**, **SEKELSEL**.

n., dekubitus-: *n., decubital*. (L. *decumbere* lê) Dekubitus-ulkus. Kyk **DEKUBITUS**.

n., droë: *n., dry*. Nekrose waarby die weefsel verdroog.

n., druk-: *n., pressure*. Nekrose te wyte aan 'n onvoldoende lokale bloedtoevoer, a.g.v. druk op weefsel.

n., emboliese: *n., embolic*. Koagulasienekrose wat volg op embolisme.

n., fibrinoïede: *n., fibrinoid*. Homogene, granulêre, eosinofiele degenerasie, met die eienskappe van fibrinoïed, van renale arteriolêre wande, beskadig deur langstaande hipertensie.

n., fokale: *n., focal*. Teenwoordigheid van 'n klein fokus van nekrose.

n., fosfor-: *n., phosphorus*. Nekrose van die bokakebeen as gevolg van blootstelling aan fosfordampe.

n., hialiene: *n., hyaline*. Zenker se degenerasie (von Zenker, F.A., Duitse patoloog, 1825-1898). Hialiene degenerasie van bv. voorarmspiere, onderworpe aan iskemie, veroorsaak deur 'n styfpassende gipsspalk.

n., ikteriëse: *n., icteric*. Nekrose van die lewer tydens geelsug.

n., iskemiese: *n., ischaemic*. Onomkeerbare weefseldood as gevolg van ontoereikende arteriële bloedvoorsiening en –perfusie weens o.a. patologie van die betrokke bloedvate, bv. arteriosklerose, trombo-embolisme, ens. Kyk **ISKEMIE**, **MIOKARDIALE INFARKSIE**, **GANGREEN**.

n., kwik-: *n., mercurial*. Vorm van nekrose wat voorkom in kwikvergiftiging.

n., mediale: *n., medial*. Kyk **ERDHEIM SE NEKROSE**.

Fig. 10: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

This cluster also displays first level nesting. In populating the nest the primary focus of the lexicographers has not been the lemmatisation of only multiword terms. Besides multiword terms this cluster also contains compounds of which the condensed lemma part is the second stem, and semantic core element, and the non-initial part of the guiding element the first stem of the compound. The compound lemmata in this partial article stretch are *n., druk-*, *n., fosfor-* and *n., kwik-* of which the full forms are *druknekrose*, *fosfornekrose* and *kwiknekrose*.

Nests in general language dictionaries can also include sublemmata with the nest attached to an article of which the lemma sign represents the lexical item that is represented by the non-initial stem of a compound or even the non-initial word of a multiword term. Such a form of lemmatisation is found in Figure 11 from *Basiswoordeboek van Afrikaans* (Gouws et al. 1994).

musiek (-) *nw.* 'n Mens maak musiek deur te sing, te fluit of op instrumente te speel soos die klavier, die trom, die fluit, die orrel, die kitaar, ens. *Hulle luister na musiek. Daar klink musiek uit die vertrek. musiekkamer, musiekuitvoering; musikaal, musikant; marsmusiek, popmusiek; klaviermusiek, orrelmusiek, vioolmusiek; klassieke musiek, ligte musiek. UITDR.:* Soos musiek (in 'n mens se ore) klink: goed klink. *Die plan klink soos musiek in my ore. Daar sit/steek musiek in iets: 'n Mens sê: Daar sit/steek musiek in hierdie idee as 'n mens te kenne wil gee dat 'n mens die idee goed vind.*

Fig. 11: *Basiswoordeboek van Afrikaans*

Such a co-occurrence of complex words, e.g. *musiekkamer*, *musikant*, *popmusiek*, as well as multiword items, e.g. *klassieke musiek*, *ligte musiek*, as sublemmata in a single article nest also prevails in the partial article stretch given in Figure 10. This subject field dictionary also embraces macrostructural strategies devised for general language dictionaries.

5.2.1.2 Double-layered sublemmatisation

The WGW employs a well-known strategy when ordering some compound and multiword lemma candidates as partial lemmata in a cluster of sublemmata within article nests and niches. Relying on this strategy, for both space-saving and semantic reasons, this dictionary takes the initiative to move from a system adhering to single layers of sublemmata to a system where double-layered sublemmata occur (cf. Gouws 2012: 268). Figure 12, a partial article stretch from the WGW, gives evidence of this strategy.

hemostase:* *haemostasis*. (G. *haima* bloed + *stasis* stilstand) Stuiting van bloedverlies of bloedvloei deur 'n bloedvat of 'n orgaan.

h., eksogene: *h., exogenous*. (G. *exo* buite + *gennan* voortbring) Ingrepe van buite die liggaam om bloedverlies uit beskadigde bloedvat(e) te stuit, bv. noodchirurgie, of drukking op toevoerende arterieë waar dit oor ferm, onderliggende beenstrukture verloop, ens.

h., endogene: *h., endogenous*. (G. *endon* binne + *gennan* voortbring) Alle fisiologiese prosesse om bloeding uit 'n beskadigde bloedvat(e) te voorkom of te beëindig.

e.h. deur ekstravaskulêre meganismes: *e.h. by extravascular mechanisms*. Drukking op bloedvate deur reaksie van omringende elastiese bindweefsel, gladdespierweefsel en vel.

e.h. deur intravaskulêre meganismes: *e.h. by intravascular mechanisms*. Trombosietpropvorming, aktivering van bloedstolling en vaatspasma deur trombosietfaktore, asook stolselretraksie deur o.a. bloedstollingsfaktor XIII.

e.h. deur vaskulêre meganismes: *e.h. by vascular mechanisms.* Spasma van gladdespierweefsel in die beskadigde bloedvatwand wat veral deur die outonome senuweestelsel geaktiveer word.

Fig. 12: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Attached to the article of the lemma sign *hemostase* is a cluster containing a nest with two sublemmata, i.e. *h., eksogene* and *h., endogene* (representing the terms *eksogene hemostase* and *endogene hemostase*). These lemmata constitute the first layer of sublemmata. Attached to the article of the sublemma *h., endogene*, and not to the article of the main lemma *hemostase*, a further nest occurs, that contains articles of the sublemmata *e.h. deur ekstravaskulêre meganismes*, *e.h. deur intravaskulêre meganismes* and *e.h. deur vaskulêre meganismes*. (*endogene hemostase deur ekstravaskulêre meganismes*, etc.) These lemmata give evidence of a second level of sublemmata, constituting a process of double-layered sublemmatisation.

This procedure of double-layered sublemmatisation ensures a close proximity between the second layer of sublemmata and their preceding first level sublemma that represents the semantic core of the sublemmata in the second layer. However, it also introduces another significant macrostructural aspect, i.e. the notion of integrated macrostructures.

5.2.1.3 Integrated macrostructure

Wiegand (1989a: 482; 1996: 3) and Wiegand and Smit (2013: 176) discuss different types of microstructures. One of the distinctions they make, is between integrated and non-integrated microstructures. Where an integrated microstructure prevails each subcomment on semantics has its items giving example sentences positioned in the same integrate as the corresponding item giving the paraphrase of meaning or translation equivalent, as seen in Figure 13, a shortened article from the *Macmillan Dictionary* (<https://www.macmillandictionary.com/dictionary/>).

- line**
1. a long thin mark on the surface of something
Draw a straight line.
Each horizontal line on the graph represents fifteen minutes.
...
2. a border between two regions
the line between Northern Ireland and the Republic
...
3. an imaginary limit or border between two situations or conditions
The country is still divided along ethnic lines.

Fig. 13: *Macmillan Dictionary*

Putting the examples in the same subcomment on semantics as the paraphrase of meaning leads to proximity and a relation of direct addressing between the items giving the examples and the respective item giving the paraphrase of meaning. Proximity enhances comprehensibility.

The use of nesting and niching procedures in the ordering of lemmata shows some resemblance to the notion of integration because a cluster of articles are integrated into a partial article stretch by attaching them to the article of a vertically ordered main lemma. However, although arguments could be offered to support the notion of an integrated macrostructure here, the traditional forms of horizontal lemmatisation by means of niching and nesting are regarded in this paper as well-established lemmatisation procedures and not as instances of an integrated macrostructure. An integrated macrostructure prevails when the lemmatisation procedure that enables closer proximity between sublemmata and a preceding (sub)lemma goes a step further than the traditional forms of sinuous lemma files by means of horizontal ordering. Figure 12 illustrates how macrostructural items are brought into close proximity beyond the first layer of sublemmatisation and clustered as a second layer of partial sublemmata attached to the article of a sublemma. This is one of the types of an integrated macrostructure in the WGW.

The partial article stretch given in Figure 14 shows another example of an integrated macrostructure.

steriliteit: * *sterility*. (L. *sterilitas* onvrugbaarheid) **1.** Afwesigheid van lewende, patogene mikro-organismes. Kyk **ASEPSIE**. **2.** Onvermoë om kinders voort te bring, d.w.s. onvermoë van 'n vrou om swanger te word, of onvermoë van die man om swangerskap te verwek. Vgl. **INFERTILITEIT**.

s., absolute: *s., absolute*. (L. *absolutus* vry van beperkinge) Totale en onomkeerbare onvrugbaarheid (by **2**).

s., manlike: *s., male*. Absolute onvrugbaarheid van die man te wyte aan een van die volgende tipes afwykings (by **2**):

m.s., aspermatogeniese: *m.s., aspermatogenic*. Weens onvermoë om lewende, funksionele sperme voort te bring.

m.s., dispermatogeniese: *m.s., dysspermatogenic*. Weens abnormale spermatogenese.

m.s., normospermatogenic: Weens een of meer afwykings wat nie verband hou met abnormale spermatogenese nie.

s., primêre: *s., primary*. (L. *primus* eerste) **1.** Absolute onvrugbaarheid waar geen kind ooit voortgebring is nie. **2.** Absolute onvrugbaarheid weens die afwesigheid van 'n faktor wat noodsaaklik is vir voortplanting (by **2**). Vgl. **STERILITEIT**, **SEKONDÊRE**.

s., relatiewe: *s., relative*. (L. *relatio* om terug te bring) Kyk **INFERTILITEIT**.

s., sekondêre: *s., secondary*. (L. *secundus* tweede) **1.** Absolute onvrugbaarheid wat intree nadat die vrou 'n swangerskap beleef of die man dit verwek het. **2.** Absolute onvrugbaarheid weens 'n gebrek wat nie met geboorte aanwesig was nie (by **2**). Vgl. **STERILITEIT**, **PRIMÊRE**.

s., vroulike: *s., female*. Absolute onvrugbaarheid in die vrou weens 'n strukturele of funksionele afwyking van haar reprodktiewe stelsel (by 2).

Fig. 14: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Attached to the article of the main lemma *steriliteit* the first layer of sublemmata stretches from *s., absolute* to *s., vroulike*. The article of the sublemma *s., manlike* contains a further cluster of sublemmata, i.e. *m.s., aspermatogeniese; m.s., disspermatogeniese* and *m.s., normospermatogeniese*. The corresponding full forms of these partial sublemmata would be *aspermatogeniese manlike steriliteit*, *disspermatogeniese manlike steriliteit* and *normospermatogeniese manlike steriliteit*. In this example the second layer of sublemmata is not attached to the article of the preceding first layer sublemma but rather integrated into the article — as signalled by the colon positioned immediately before the second layer of sublemmata. These sublemmata represent terms that indicate defects that can cause the problem referred to by the term functioning as the preceding first level sublemma. The proximity stresses the semantic relation between these two layers of sublemmata.

Another type of integrated macrostructure can be seen in Figure 15.

tomografie: * *tomography*. (G. *tome* sny + *graphein* skryf) Gebruik van X- of gamma-strale, of ultrasoniese golwe vanuit twee of meer hoeke om inwendige liggaamstrukture grafies uit te beeld asof daar snitte dwarsdeur die liggaam gemaak is. Dit skep 'n beeld wat reghoekig geleë is tot 'n eenvoudige of tradisionele X-straalfotobeeld. Die elektroniese registrasie van teruggekaatste ultrasoniese golwe vanaf voorafbepaalde, gefokusde dieptevlakke bied 'n afbeelding van die anatomie van sodanige opeenvolgende snyvlakke (tomogramme); ■ konvensionele tomografie; planigrafie; stratigrafie.

t., rekenaar- (RT; CT): *t., computed; computerized (CT)*. Breë spektrum van tomografietegniese waarin rekenartegnologie toepassing vind vir die elektroniese verwerking en integrasie van seriële beelde verkry deur tomografie; ■ komputertomografie.

r., aksiale (KAT, RT; CAT; CT): *c. t., axial (CAT)*. (G. *akson* as) Spesiale RT-tegniek van historiese belang waar die deurgaande strale elektronies verwerk word om gerekonstrueerde dwarsnitte deur die liggaam uit te beeld. Die snitte word sodanig gefokus dat die tweedimensionele beelde ko-aksiaal om 'n bepaalde diepte-as wentel. ■ Ko-aksiale rekenartomografie; ko-aksiale tomografie.

r., dinamiese c. t., dynamic. (G. *dunamis* krag) RT-tegniek met vinnige inspuiting van kontrasmiddel met opeenvolgende skandering, gefokus op 'n voorafbepaalde gebied om die bloedvate te demonstree.

r., elektronstraal- (EST; EBT): *c. t., electron beam (EBT)*. RT-tegniek waarin die sirkelbeweging van die X-straal-buis vervang word deur vinnige elektroniese posisionering van die katodestraal rondom 'n sirkelvormige anode. Volle skandering geskied binne enkele millisekondes.

r., enkelfoton-emissie (EFERT; EFET; SPECT; SPET): *c. t., single-photon emission (SPECT; SPET)*. (L. *emittere* uitstuur) Kerngeneeskundige RT-teg-

niek, berus op die intraveneuse toediening van 'n gamma-foton-uitstralende radionuklid en die opsporing daarvan in die liggaam met 'n roterende gamma-kamera. EFERT weerspieël die verspreiding van radioaktiwiteit meer akkuraat as planare beelde en verbeter dus die presiese lokalisasie van patologie. Tweedimensionele beelde kan rekenaarmatig gerekonstrueer word om 'n drie-dimensionele voorstelling te gee.

Fig. 15: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

There is no niche or nest attached to the article of the main lemma *tomografie* — a niche or nest needs to contain more than one lemma. However, a single sublemma *t.,rekenaar (rekenaartomografie)*, with its abbreviated forms in Afrikaans and English *RT, CT*, is attached to the article of *tomografie*. Gouws (2012: 269) refers to such a single sublemma as a macrostructurally-isolated sublemma. The occurrence of this article with the single sublemma as guiding element offers a landing zone for a cluster of nine second level sublemmata of which only the first four are shown in Figure 15. These four second level sublemmata are *r., aksiale*; *r., dinamiese*; *r., elektronstraal-*; and *r., enkelfotonemissie*. The condensed lemma part *r.* represents the word *rekenaartomografie*, the full form of the sublemma of the article to which the second level sublemmata have been attached. The full form of the first of these sublemmata will be *aksiale rekenaartomografie*. In Figure 15 the integrated macrostructure employs a procedure of double-layering of sublemmata to attach a second level article nest to a macrostructurally-isolated sublemma. Yet again, the integrated macrostructure brings semantically related terms into close proximity — something that could not have been achieved as successfully without the double-layer of sublemmata.

There are two other types of integrated macrostructures prevailing in the WGW that will be discussed briefly — one with double-layered sublemmata and one with single level sublemmata. In the user guidelines of the WGW the editor says that in instances where the main lemma groups a big group of analogous combinations together each combination is treated as a sublemma, ordered alphabetically according to the first letter and numbered accordingly. This can be seen in Figure 16 (which is not an example of an integrated macrostructure) where 16 sublemmata are entered (of which only the first six are given here). They are not compound forms or multiword items in which the term represented by the main lemma occurs but they resemble hyponyms with the main lemma as superordinate. Their inclusion as a specific type of sublemma is determined on semantic grounds. These sublemmata are also entered in cross-reference articles in their respective alphabetical positions elsewhere in the relevant article stretches.

gastro-intestinale hormone:† *gastrointestinal hormones*. (G. *gaster* maag + L. *intestinum* darm) 'n Breë spektrum van polipeptiedhormone, dikwels molekuler homologe, primêr in die mukosa (slymvlies) van die maag en/of die darmkanaal (insluitend die endokriene pankreas) afgeskei, en wat op geïntegreerde

wyse dermkanaalfunksies sinchroniseer; ■ enterohormone; gastroïntestinale hormone.

1. cholestokiniën-pankreasimien (CSK-PS; CCK-PZ): *cholecystokinin-pancreozymin (CCK-PZ)*. (G. *khole* gal + *kustis* blaas + *kinein* beweeg + *zume* fermenteer, deursuur) Polipeptied, vrygestel uit die mukosa van die duodenum, jejunum en ileum, asook in die hipotalamus. Dit stimuleer die galblaaswand tot sametrekking, die eksokriene pankreas tot die afskeiding van spysverterings-ensieme, sinchroniseer dermkanaalbewegings en funksioneer as 'n neurotransmitter (senuwee-oordragstof) in die hipotalamus.

2. enteroglukagon: *enteroglucagon; gut glucagon*. (G. *enteron* ingewande + *glukus* soet + *ago* om te lei) Versamelnaam vir verskeie polipeptiede uit 'n voorloperhormoon (proglukagon) uit die L-selle van die mukosa van die boonste deel van die gastro-intestinale kanaal, afgeskei ná inname van voedsel en veral glukose. Een segment van die molekule is glisientien (ook deur die A-selle van die pankreas-eilandweefsel afgeskei), terwyl ander fragmente onderskeidelik glukagonagtige peptied 1 (GLP⁻¹) en glukagonagtige peptied 2 (GLP⁻²) genoem word. Molekulêre homologie tussen proglukagon uit die endokriene pankreas en enteroglukagon word gedemonstreer deur kruisreaksies met dieselfde immuunserum. Hierdie polipeptiede inhibeer maaglediging en stimuleer insuliensekresie, terwyl 'n trofiese invloed op die dermkanaal-mukosa gepostuleer is.

3. gastrien: *gastrin*. (G. *gaster* maag) 'n Reeks homoloë polipeptiede, hoofsaaklik afkomstig uit die G-selle van die mukosa van die antrum van die maag, duodenum, boonste ileum, asook die pankreaseilandweefsel, wat maagsuur- en pepsiensekresie stimuleer, terwyl dit esofagus- en maagbewegings, asook sfinkterfunksie modifiseer. Die funksies van gastrien in die adenohipofise, brein en perifere senuweevesels is onbekend.

Kyk ZOLLINGER-ELLISON-SINDROOM; ANTRALE G-SELHIPERPLASIE; GASTRIENVRYSTELLINGSPEPTIED (GVP); PENTAGASTRIEN.

4. gastrienvrystellingspeptied (GVP; GRP): *gastrin-releasing peptide (GRP)*. Polipeptied uit veral die mukosa en senuweevesels van die maag-antrum en duodenum, molekulêr homolog aan bombesien (afgeskei deur paddavel). Dit is moontlik 'n neurotransmitter wat die sekresie van gastrien, CCK-PZ en pankreassap stimuleer, galblaas- en dermkanaalmotiliteit bevorder, en ook betrokke is by temperatuurbeheer en pynpersepsie in die brein.

5. gastrien-inhibitoriese polipeptied (GIP): *gastric inhibitory polypeptide (GIP)*. Historiese benaming. Staan tans bekend as glukose-afhanklike insulintropiese peptied. Kyk 7.

6. glisientien: *glisentin*. Polipeptied (ook bekend as glukagonagtige immuno-reaktiwiteit, GLI^{*}, omdat dit die glukagonsegment bevat), wat saam met glukagon uit die A-selle van die pankreaseilandweefsel afgeskei word, maar ook wydverspreid in die spysverteringskanaalmukosa aangetref word. Funksies onbekend. Kyk ENTEROGLUKAGON.

Fig. 16: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

The type of nest with numbered sublemmata as seen in Figure 16 creates an opportunity for an application of double-layered sublemmatisation and the use of additional integrated macrostructural procedures. This can be seen in Figure 17.

etiek:* *ethics*. (G. *ethikos* voortvloeiend uit *sedes*; *ethos* moraal) **1.** Die norme en reëls wat 'n samelewing se siening oor die onderskeid tussen goed en kwaad, asook regte en verkeerde gedrag weergee. **2.** In filosofie: die dissipline waarin die verskynsel bestudeer word dat daar ten opsigte van alle menslike handeling ge vra kan word of hulle reg of verkeerd, goed of kwaad is. Daarin word probeer om die teoretiese gronde vir die onderskeiding tussen regte en verkeerde handeling te definieer, met inagneming van die samelewing se siening.

e., geneeskunde- *e., medical; biomedical*. Die norme en voortvloeiende gedrag rakende gesondheidswerkers se regte en verpligtinge jeens pasiënte, die familie van pasiënte, die gemeenskap en kollegas, asook die versorging van pasiënte onder hulle toesig; ■ *etiek, biomediese*. **Kyk ETIEKTEORIE**. Hierdie norme berus op minstens die volgende vier fundamentele beginsels:

1. geregtigheid: *justice*. (L. *justitia* regverdigheid; billikheid; geregtigheid) ;

2. goedwilligheid: *benificence*. (L. *beneficentia* weldadigheid; goedgunstigheid) ;

3. niekwaadwilligheid: *nonmaleficence*. (L. *non* nie + *maleficere* kwaadwillig) ;

4. outonomie: *autonomy*. (G. *autos* self + *nomos* wet)

e., -komitee: *e. committee*. **Kyk** NAVORSING(S)ETIEKKOMITEE.

Fig. 17: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Figure 17 presents an occurrence of double-layered sublemmata where a second layer of sublemmata *geregtigheid*, *goedwilligheid*, *niekwaadwilligheid* and *outonomie* is integrated into the article of a first level sublemma. The lemmata in this second layer are terms representing fundamental principles of norms referred to in the article of the first level sublemma *geneeskunde-etiek*. This results in an integrated macrostructure — each term is a guiding element and primary treatment unit of an article.

The last type of an integrated macrostructure does not need an occurrence of double-layered sublemmatisation. It is the result of semantically related sublemmata being integrated into different subcomments on semantics of a dictionary article. The article of the polysemous term *buis* (Figure 18) has three subcomments on semantics, identified by roman numerals given in brackets as polysemy markers. The third subcomment on semantics contains an item giving an English translation equivalent and a cross-reference item guiding the user to the lemma *kanaal*. The first two subcomments on semantics, each has an item giving a translation equivalent and an item giving a brief paraphrase of meaning. Integrated into these subcomments on semantics are nests with sublemmata presenting both compounds and multiword terms. Each sublemma contains the element *buis* and they link semantically with that sense of the main

lemma *buis* that prevails in the given subcomment on semantics. The lexicographer has employed an integrated macrostructure to ensure that semantically related terms are entered not only in close proximity of the relevant main lemma but in close proximity of the relevant sense.

buis:* (I) *duct*. Kanaal, veral een vir die vervoer van eksokriene klierafskeidings. Kyk EKSOKRIEN.

b., gestreepte: *d., striated*. Deel van intralobulêre buissisteem van 'n speekselklier waarin die selle gekenmerk word deur talle invouings van die basale selmembraan, vandaar die ligmikroskopiese beskrywing as gestreep.

b.-karsinoom: *d. carcinoma*. Karsinoom van relatief lae kwaadaardigheid, wat sy oorsprong het in die epiteel van 'n buis, veral die melkbuis, gewoonlik naby die tepel.

b.-papilloom: *d. papilloma*. Papilloom wat ontstaan uit die epiteel van 'n buis, veral 'n melkbuis, gewoonlik naby die tepel.

(II) *tube*. Enige hol silindriese struktuur.

b., endotrageale: *t., endotracheal*. (G. *endon* binne + *trachus* grof) 'n Soepel buis van plastiek, rubber of kunsvesel vir inplasing deur die mond, neus of trageostomie vir o.a. toediening van narkosegasse, aspirasie van sekresies en ventilasie.

b., Eustachius-: *t., pharyngotympanic*. Kyk B., FARINGOTIMPANIESE.

b., faringotimpaniese: *t., pharyngotympanic*. Buis, ongeveer 36 mm lank, wat strek van die timpaniese holte tot by die nasofarinks.

b., maag-: *t., stomach*. Buis vir voeding of een waarmee die maag uitgespoel word.

b., nasogastriese: *t., nasogastric*. 'n Buigbare rubber- of kunsveselbuis, deur die nasofarinks en esofagus tot in die maag geplaas, waardeur sg. maagbuisvoeding (maagbuisalimentasie) gegee kan word. Vgl. MAAGBUIS.

b., orotrageale: *t., orotracheal*. (L. os mond) 'n Endotrageale buis wat deur die mond ingeplaas word.

b., uteriene: *t., uterine*. Kyk FALLOPIUS-BUIS.

b.-voeding: *t. nutrition*. (L. *nutrire* voed) Die enterale toediening van gebalanseerde, vloeibare, gehomogeniseerde voedingstowwe per nasogastriese, nasoduodenale, of orogastriese, sagte silikoon- of poliuretaanbuis aan 'n ver-swakte pasiënt, of as daar anatomiese indikasies is, bv. esofageale obstruk-sies of fistels. Plasing van die buis deur 'n gastrostomie- of jejunostomie-ingreep word oorweeg as die voedingregime vir langer as 6 weke voortgesit moet word.

(III) *tract*. Kyk KANAAL.

b., uteriene: *t., uterine*. Kyk FALLOPIUS-BUIS.

b., Sengstaken-Blakemore: *Sengstaken-Blakemore*. (Sengstaken, R.W., Amerikaanse neurochirurg, 1923-** + Blakemore, A.H., Amerikaanse chirurg, 1897-1970) 'n Klassieke buis met drie lumens en twee ballonne, histories gebruik vir noodbehandeling van bloeiende esofageale spatere Vgl. PORTALE HIPERTENSIE.

Fig. 18: *Woordeboek vir die Gesondheidswetenskappe / Dictionary for the Health Sciences*

Integrated macrostructures presuppose a system where the vertical ordering of lemmata is complemented by a horizontal ordering. Two main types of integrated macrostructures have been employed in the WGW. Figures 12, 14 and 15 display the first type of integrated macrostructure. The integration is due to a second layer of sublemmata where the sublemmata are either integrated into the partial article stretch of first level sublemmata (Fig. 12), integrated into a partial article stretch by being attached to macrostructurally-isolated first level sublemma (Fig. 14) or integrated into an article of a first level sublemma (Fig. 15). Figure 17 shows a variant of this first type but where there is a relation between a nest of numbered sublemmata integrated into the article of a first level sublemma and a concept referred to in the preceding treatment of the first level sublemma. They are integrated into the article of the first level sublemma.

The second type of integrated macrostructure is seen in Figure 18. Here the first layer of sublemmata are integrated into separate subcomments on semantics of the article of a main lemma, signalling that there is a close semantic relation between the sublemmata and the sense of the main lemma expressed in the specific subcomment on semantics.

6. In conclusion

Special field lexicography and subject field lexicography are subdomains of lexicography. Typical lexicographic procedures are used in these dictionaries. Where the compiler of such a dictionary takes the necessary cognizance of guidelines from a general theory of lexicography such a dictionary can become a good dictionary not only on account of the contents but also due to the appropriate dictionary structures and an adherence to the user-perspective and the relevant lexicographic functions. Lexicographers of special field and subject field dictionaries often employ lexicographic procedures and structures in an innovative way. This gives theoretical lexicographers the opportunity to analyse and discuss these innovations, and use them as a point of departure from where the fields of metalexigraphy and dictionary research can be enriched.

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (UID) 85434). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

References

Dictionaries

- Brink, A.J. and J. de V. Lochner (Eds.).** 2011². *Woordeboek vir die Gesondheidswetenskappe/Dictionary for the Health Sciences*. Cape Town: Pharos.
- De Villiers, M. and R.H. Gouws.** 1988. *Idiomewoordeboek*. Cape Town: Nasou.
- Fata, I.** 2005. *Ungarisch–Deutsches Deutsch–Ungarisches Fachwörterbuch zur Rentenversicherung*. Szeged: Grimm Kiadó.
- Gouws, R. et al. (Eds.).** 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- Hartmann-Petersen, P. et al.** 2001. *SASOL Science and Technology Resource*. Claremont: New Africa Education Publishing.
- Kaufmann, U. and H. Bergenholtz (Eds.).** 1998. *Encyclopedic Dictionary of Gene Technology*. Toronto: Lugus Libros.
- Kwiatkowski, G. (Ed.).** 1989². *Schüler Duden: Die Musik*. Mannheim: Dudenverlag.
- Macmillan Dictionary*. <https://www.macmillandictionary.com/dictionary/>. (Consulted 10 February 2020.)
- Pyne, S. and A. Tuck.** 1996. *Oxford Dictionary of Computing for Learners of English*. Oxford: Oxford University Press.
- Wiegand, H.E. et al. (Eds.).** 2010–2020. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Berlin/New York: Walter de Gruyter.

Other literature

- Bergenholtz, H.** 1995. Wodurch unterscheidet sich Fachlexikographie von Terminographie? *Lexicographica* 11: 50-59.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. and S. Tarp.** 2010. LSP Lexicography or Terminography? The Lexicographer's Point of View. Fuertes-Olivera, P.A. (Ed.). 2010: 27-38.
- Bergenholtz, H., S. Tarp and H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2: 1762-1832*. Berlin/New York: Walter de Gruyter.
- Cowie, A.P. (Ed.).** 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. *Lexicographica. Series Maior* 17. Tübingen: Max Niemeyer.
- Fuertes-Olivera, P.A. (Ed.).** 2010. *Specialised Dictionaries for Learners*. Berlin/New York: Walter de Gruyter.
- Fuertes-Olivera, P.A. (Ed.).** 2018. *The Routledge Handbook of Lexicography*. London/New York: Routledge.

- Fuertes-Olivera, P.A. and A. Arribas-Baño (Eds.).** 2008. *Pedagogical Specialised Lexicography: The Representation of Meaning in English and Spanish Business Dictionaries*. Terminology and Lexicography Research and Practice (TLRP) 11. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H.** 2001. The Use of an Improved Access Structure in Dictionaries. *Lexikos* 11: 101-111.
- Gouws, R.H.** 2010. The Monolingual Specialised Dictionary for Learners. Fuertes-Olivera, P.A. (Ed.). 2010: 55-68.
- Gouws, R.H.** 2012. Innovative Strategies in Macrostructural Choices. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa*: 251-269. Stellenbosch: Sun Press.
- Gouws, R.H.** 2016. Increasing the Scope of the Treatment of Specialised Language Terms in General Language Dictionaries. Schierholz, S. et al. (Eds.). 2016. *Wörterbuchforschung und Lexikographie*: 101-118. Berlin: De Gruyter.
- Gouws, R.H. et al. (Eds.).** 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin/New York: De Gruyter.
- Hartmann, R.R.K.** 1992. Lexicography, with Particular Reference to English Learners' Dictionaries. *Language Teaching* 25(3): 151-159.
- Hartmann, R.R.K.** 1994. Bilingualised Versions of Learners' Dictionaries. *Fremdsprachen Lehren und Lernen* 23: 206-220.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1-5.3. Berlin/New York: Walter de Gruyter.
- Hollós, Z.** 2004. *Lernerlexikographie: syntagmatisch. Konzeption für ein deutsch-ungarisches Lernerwörterbuch*. Tübingen: Max Niemeyer.
- Humbley, J.** 2018. Specialised Dictionaries. Fuertes-Olivera, P.A. (Ed.). 2018: 317-334.
- Jesenšek, V. (Ed.).** 2013. *Specialised Lexicography. Print and Digital, Specialised Dictionaries, Databases*. Berlin/Boston: De Gruyter.
- Kilgarriff, A.** 2012. Review of Pedro A. Fuertes-Olivera and Henning Bergenholtz (Eds.). *e Lexicography: The Internet, Digital Initiatives and Lexicography*. *Kernerman Dictionary News* 20: 26-29.
- Lemmens, M. and H. Wekker.** 1986. *Grammar in English Learners' Dictionaries*. Tübingen: Max Niemeyer.
- Runte, M.** 2015. *Lernerlexikographie und Wortschatzerwerb*. *Lexicographica*. Series Maior 150. Berlin/Boston: De Gruyter.
- Schierholz, S.J.** 2003. Fachlexikographie und Terminographie. *Zeitschrift für Angewandte Linguistik* 39: 5-28.
- Schierholz, S.J.** 2013. New Developments in Lexicography for Special Purposes I: An Overview of Linguistic Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 431-442.
- Steyn, M.** 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.
- Tarp, S.** 1994. Funktionen in Fachwörterbüchern. Schaefer, B. and Bergenholtz, H. (Eds.). 1994. *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*: 229-246. Tübingen: Gunter Narr.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.

- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Van der Colff, A.** 1996. *Zur Konzeption eines einsprachigen deutschen Lernerwörterbuchs*. Unpublished Doctoral Dissertation. Stellenbosch: Stellenbosch University.
- Wiegand, H.E.** 1988. Was eigentlich ist Fachlexikographie? Haider Munske, H. et al. (Eds.). 1988. *Deutscher Wortschatz. Lexikologische Studien*: 729-790. Berlin: De Gruyter.
- Wiegand, H.E.** 1989. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Eds.). 1989-1991: 246-280.
- Wiegand, H.E.** 1989a. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989-1991: 462-501.
- Wiegand, H.E.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger lexikographischen Kolloquium*: 1-82. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1998. *Wörterbuchforschung*. Berlin: De Gruyter.
- Wiegand, H.E.** 2003. Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Städler, Thomas (Ed.). 2003. *Wissenschaftliche Lexikographie im deutschsprachigen Raum. Im Auftrag der Heidelberger Akademie der Wissenschaften*: 417-437. Heidelberg: Winter.
- Wiegand, H.E.** 2004. Über die Unterschiede von Fachlexikographie und Terminographie. Am Beispiel des Wörterbuchs zur Lexikographie und Wörterbuchforschung. Wiegand, H.E. (Ed.). 2004. *Studien zur zweisprachigen Lexikographie mit Deutsch IX*: 135-152. Germanistische Linguistik 178. Hildesheim/Zürich/New York: Georg Olms.
- Wiegand, H.E.** 2004a. Reflections on the Mediostructure in Special-Field Dictionaries. Also According to the Example of the *Dictionary for Lexicography and Dictionary Research*. *Lexikos* 14: 195-221.
- Wiegand, H.E. and R.H. Gouws.** 2013. Macrostructures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 73-110.
- Wiegand, H.E. and M. Smit.** 2013. Microstructures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 149-214.
- Wolski, W.** 1989. Das Lemma und die verschiedenen Lemmatypen. Hausmann, F.J. et al. (Eds.). 1989-1991: 360-371.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

An Overview of English Dictionaries of Abbreviations

Mojca Kompara Lukančič, *Faculty of Criminal Justice and Security
and Faculty of Tourism, University of Maribor, Slovenia*
(mojca.kompara@gmail.com) (mojca.kompara@um.si)

Abstract: This article addresses the position of specialized dictionaries of abbreviations for English. Special attention is given to the macro- and microstructure of English dictionaries of abbreviations, and current available online sources for users are presented. The article compares the micro- and macrostructure of seventeen English dictionaries of abbreviations, including general and language-specific ones, and three online general dictionaries of abbreviations. In the article two contemporary online attempts are also presented; namely, *Slovarček krajšav* (Kompara Lukančič 2006) and *Slovar krajšav* (Kompara Lukančič 2011). The article points out that traditional lexicography, mainly its orientation toward compiling functional and detailed entries, should be joined with modern concepts of lexicography oriented toward an extensive number of entries and constant development and improvement of a dictionary as new abbreviations are gradually added. The combination of traditional and modern concepts can develop resources that users need most — namely, a contemporary dictionary of abbreviations.

Keywords: ABBREVIATIONS, LEXICOGRAPHY, DICTIONARY, ENGLISH, SLOVENE, LANGUAGE, ENGLISH DICTIONARIES OF ABBREVIATIONS, DICTIONARY COMPILATION, ONLINE DICTIONARY, MICROSTRUCTURE, MACROSTRUCTURE

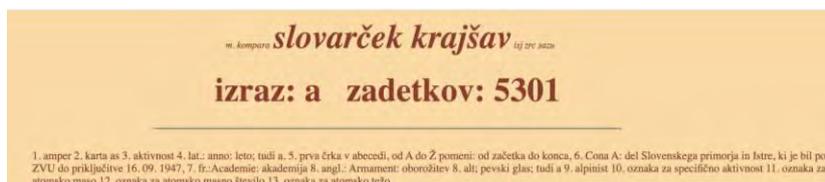
Opsomming: 'n Oorsig van Engelse afkortingswoordeboeke. Hierdie artikel bespreek die stand van gespesialiseerde Engelse afkortingswoordeboeke. Spesiale aandag word aan die makro- en mikrostruktuur van Engelse afkortingswoordeboeke geskenk en huidige beskikbare aanlyn bronne vir gebruikers word voorgestel. Die mikro- en makrostruktuur van sewentien Engelse afkortingswoordeboeke, met algemene sowel as taalspesifieke woordeboeke daarby ingesluit, asook drie aanlyn algemene afkortingswoordeboeke word met mekaar vergelyk. In hierdie artikel word ook twee eietydse aanlyn pogings, nl. *Slovarček krajšav* (Kompara Lukančič 2006) en *Slovar krajšav* (Kompara Lukančič 2011), voorgestel. Daar word daarop gewys dat die tradisionele leksikografie, veral die gerigtheid op die skep van funksionele en gedetailleerde inskrywings, versoen moet word met moderne leksikografiese benaderings wat gefokus is op die uitbreiding van die aantal inskrywings en op konstante ontwikkeling en hersiening van 'n woordeboek soos nuwe afkortings geleidelik bygevoeg word. Die kombinasie van tradisionele en moderne benaderings kan 'n hulpbron skep wat gebruikers die nodigste het — naamlik, 'n eietydse afkortingswoordeboek.

Sleutelwoorde: AFKORTINGS, LEKSIKOGRAFIE, WOORDEBOEK, ENGELS, SLOWEENS, TAAL, ENGELSE AFKORTINGSWOORDEBOEKE, WOORDEBOEKSAMESTELLING, AANLYN WOORDEBOEK, MIKROSTRUKTUUR, MAKROSTRUKTUUR

Introduction

Abbreviations are complex and problematic (Gabrovšek 1994), and they frequently develop, appear, and disappear with no prior notice (e.g., *COVID-19*, *SARS*, etc.). Some remain in a language forever or a very long time (e.g., *FBI*), and some disappear after a certain time (e.g., *LDS*). It is difficult to determine who really needs them and, due to their perishability, the issue of including them in dictionaries is discussed among lexicographers, who agree that dictionaries of abbreviations must be up to date when created because new abbreviations arise rapidly in every language (Gabrovšek 1994). Abbreviations are a growing phenomenon (Kompara Lukančič 2010), and in the Slovene linguistic framework they are mainly dealt with by normative dictionaries (Kompara Lukančič 2009; 2010; Logar 2005). An extensive study on the inclusion of abbreviations in monolingual and terminological dictionaries was prepared in 2017 (Kompara Lukančič 2017). In Slovene, four abbreviation dictionaries are available: two are paper dictionaries, *Kratice* (Župančič 1948) and *Rečnik jugoslovenskih skračenica* (Židar 1971), but both are outdated, and two more are contemporary online attempts; namely, *Slovarček krajšav* (Kompara Lukančič 2006), lacking a proper structure, and *Slovar krajšav* (Kompara Lukančič 2011), lacking an extensive number of entries. The *Slovarček krajšav* dictionary (Kompara Lukančič 2006) is an online dictionary composed of 5,700 Slovene and foreign abbreviations. The dictionary was prepared as a traditional paper dictionary that was later transferred into a digital format and went online. Among the foreign abbreviations we can find French, Italian, German, English, Spanish, and Latin abbreviations that are used in Slovene. All abbreviations are composed of expansions and translations, in the case of foreign abbreviations. Some abbreviations have more than one expansion, all are numerically ordered and included within the same dictionary entry, as in Example 1. The dictionary entry is simple, all foreign abbreviations are composed of language qualifiers, followed by the foreign expansions, as in Example 1 (i.e., 4. lat.: anno). Some additional explanations are also provided for some expansions to help the user understand the meaning, (i.e., 6. Cona A). The dictionary is a good attempt at merging traditional and modern lexicographic concepts, but unfortunately the dictionary lacks a proper structure. Within the dictionary entry the expansions are not alphabetically ordered, the number of dictionary entries should be increased, and a detailed revision of all entries is needed. In addition, special attention should be made to cross-references, which are currently lacking.

Example 1



The *Slovar krajšav* dictionary (2011) is an online dictionary compiled with the aid of the Termania dictionary mask, and is composed of 2,500 dictionary entries. In the dictionary we can find Slovene and foreign abbreviations and their expansions. All foreign abbreviations are composed of language qualifiers, but there are no translations provided, as in Example 2. The dictionary was compiled entirely automatically, thanks to the algorithm for automatic recognition of abbreviations, the algorithm for lemmatization of Slovene expansions, and the algorithm for language recognition (to include language qualifiers automatically). *Slovar krajšav* (2011) is a good example of modern lexicography, but it lacks a more extensive number of dictionary entries, as well as the inclusion of translations and/or additional data, as seen in *Slovarček krajšav* (Kompara Lukančič 2006).

Example 2

AA

Architectural Association (en)

arhitekturni atelje (sl)

Vir: [Slovar krajšav - Mojca Kompara](#)

A more extensive dictionary of abbreviations for Slovene is in preparation (Kompara Lukančič 2017). This article addresses the position of specialized dictionaries of abbreviations for English and takes them as an example of good practice for compiling future dictionaries of abbreviations. Special emphasis is placed on the macro- and microstructure of English dictionaries of abbreviations, and current available online sources for users are presented.

1. Overview of the macrostructure of English dictionaries of abbreviations

This overview of English dictionaries of abbreviations examines dictionaries published from 1942 to 2019. The following dictionaries of abbreviations are examined and presented:

- *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations* (Partridge 1942)
- *World Guide to Abbreviations of Organizations* (Buttress 1976)
- *Everyman's Dictionary of Abbreviations* (Paxton 1983)
- *Abbreviations Dictionary* (De Sola 1986)
- *Dictionary of Abbreviations in Medical Sciences* (Heister 1989)
- *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning* (Jung 1991)

- *The Oxford Dictionary of Abbreviations* (Alexander et al. 1992)
- *The Barnhart Abbreviations Dictionary* (Barnhart 1995)
- *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols* (Mattia 1997)
- *Abbreviations and Acronyms* (Geddes & Grosset 1999)
- *The Wordsworth Dictionary of Abbreviations and Acronyms* (Dale and Puttick 1999)
- *The New Penguin Dictionary of Abbreviations* (Fergusson 2000)
- *Dictionary of Financial Abbreviations* (Paxton 2003)
- *Acronyms, Initialisms and Abbreviations Dictionary* (Romaniuk 2006)
- *Abbreviations* (King 2014)
- *Dorland's Dictionary of Medical Acronyms and Abbreviations* (Dorland 2016)
- *Abbreviations Dictionary* (Stahl and Landen 2019)

The main characteristics of the macrostructure are presented in Table 1.

Table 1: Characteristics of the macrostructure of the English dictionaries of abbreviations

Dictionary	Entries	Appendix	Foreign abbrs.	Transl. of foreign abbrs.	Qualifier	Lang.	Encyclopedic data
Partridge 1942	5,000	No	Yes	Yes*	No	Yes	Yes
Buttress 1976	60,000	No	Yes	No	No	No	No
Paxton 1983	25,000	Yes	Yes	Yes*	Yes	Yes†	Yes
De Sola 1986	23,000	Yes‡	Yes	Yes*	No	Yes	Yes, rare
Heister 1989	8,000	No	Yes	Yes*	Yes	Yes	No
Jung 1991	50,000	No	Yes	No	Yes†	No	No
Alexander et al. 1992	20,000	Yes‡	Yes	Yes*	Yes	Yes	Yes
Barnhart 1995	60,000	No	Yes	Yes*	Yes	Yes	Yes
Mattia 1997	60,000	Yes	Yes	No	No	No	No
Geddes & Grosset 1999	10,000	No	Yes	Yes*	Yes	Yes	No
Dale and Puttick 1999	20,000	Yes, too ‡	Yes	Yes*	Yes	Yes	Yes
Fergusson 2000	40,000	Yes‡	Yes	Yes*	Yes	Yes	Yes, rare
Paxton 2003	7,000	Yes	Yes	Yes*	No	Yes	Yes
Romaniuk 2006	1,000,000	No	Yes	Yes*	Yes	No	Yes
King 2014	7,000	Yes‡	Yes	Yes*	No	No	Yes
Dorland 2016	30,000	Yes	Yes	Yes*	No	Yes	No
Stahl and Landen 2019	60,000	Yes‡	Yes	Yes*	No	Yes	No

Note: * = consistent, † = not consistent, ‡ = extensive

As seen in Table 1, the English dictionaries of abbreviations presented and analyzed differ in terms of macrostructure in the number of entries, whereby *Acronyms, Initialisms and Abbreviations Dictionary* (Romaniuk 2006) reaches the outstanding number of a million entries. The smallest number of entries is found in *Abbreviations and Acronyms* (Geddes & Grosset 1999), with only ten thousand entries, and some LSP dictionaries on abbreviations which are oriented entirely towards specialized fields (Partridge 1942; Heister 1989; Paxton 2003). Generally speaking, in the dictionaries examined, the number of entries is between twenty thousand and sixty thousand. In monolingual, bilingual, and multilingual dic-

tionaries, abbreviations often appear in an appendix (Kompara Lukančič 2010, 2017). Among the dictionaries of abbreviations examined in Table 1, ten of them have an appendix, which can be seen as a macrostructural component (Nielsen 1990; Čermák and Blatná), and in six cases these can be characterized as extensive or too extensive. All the dictionaries examined also include foreign abbreviations, but three of the dictionaries do not provide a translation of the foreign abbreviations. Translations were included into the above analysis at the macrostructural level mainly as an editorial decision. The lack of translation is observed in *World Guide to Abbreviations of Organizations* (Buttress 1976), *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning* (Jung 1991) and *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols* (Mattia 1997). Only nine dictionaries include field qualifiers, in *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning* (Jung 1991) they are not consistently included, and twelve dictionaries include language qualifiers. In *Everyman's Dictionary of Abbreviations* (Paxton 1983) the language qualifiers for foreign languages are not consistently used. Encyclopedic data are included in ten dictionaries, but *Abbreviations Dictionary* (De Sola 1986) and *The New Penguin Dictionary of Abbreviations* (Ferguson 2000) rarely present such data.

2. Overview of the microstructure of English dictionaries of abbreviations

This section presents an overview of the microstructure in the English dictionaries of abbreviations discussed above.

2.1 *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations* (Partridge 1942)

A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations (Partridge 1942) includes over five thousand field-specific (wartime) abbreviations. The dictionary mainly contains English abbreviations, but there are also some foreign abbreviations. All foreign abbreviations are systematically translated. The entry is simple, composed of an entry word in bold followed by the expansions. As seen in Example 3, the entry is composed of several expansions, numerically ordered — for example, (2) — except that the first number is missing. At the end of every expansion there is a period. When an expansion is foreign, as in the case of the fifth and sixth expansions, an abbreviated language qualifier is used (e.g., *L.* 'Latin') followed by a foreign expansion in italics (e.g., *anno mundi* and the English translation "in the year of the world"). The inclusion of additional information (encyclopedic data) is also visible in the first expansion in Example 3, in which the expansion is followed by a description (i.e., "For gallantry in saving life, esp. at sea").

Example 3

A.M. Albert Medal. For gallantry in saving life, esp. at sea.
(2) Air Ministry.
(3) Air Marshal.
(4) Army manual.
(5) L. *anno mundi*, in the year of the world.
(6) L. *Atrium Magister*, Master of Arts — an American form of *M.A.*
(7) Ave Maria.
(8) Associate Member.
(Partridge 1942: 12)

2.2 *World Guide to Abbreviations of Organizations* (Buttress 1976)

World Guide to Abbreviations of Organizations (Buttress 1976) contains sixty thousand alphabetically ordered entries of British/American and foreign organizations. The entry is simple; the entry word in bold is followed by the abbreviation's expansion. A negative aspect of this dictionary is the fact that no translations of foreign expansions are provided. Within the entry there are also no language and field qualifiers, but it is mentioned (even if not consistently) which country the abbreviation is used in. Some inconsistencies are visible in Example 4.

Example 4

CCD Conseil de Coopération Douanière (Belgium)
CCF Crédit Commercial de France
(Buttress 1976: 100)

2.3 *Everyman's Dictionary of Abbreviations* (Paxton 1983)

Everyman's Dictionary of Abbreviations (Paxton 1983) is composed of approximately twenty-five thousand alphabetically ordered entries and more than thirty-seven thousand expansions, and it has an appendix. The majority of the abbreviations included are English, but the presence of foreign ones (e.g., Latin, German, French, etc.) is also seen among the entries. The entry is simple: the entry word in bold is followed by alphabetically ordered expansions, divided by semicolons, as in Example 5. Example 5 shows the use of field qualifiers, abbreviated and in parentheses — for example, (*elec.*) and (*nav.*) — which follow the expansions. The language qualifier, also abbreviated, precedes the expansion (e.g., *It.*). The foreign expansion is written in italics (e.g., *da capo*) and is followed by the

English translation; for example, "repeat from the beginning (mus.)." Within the entry the expansions are alphabetically ordered.

Example 5

<p>d.c. dead centre; direct current (elec.); double column; double crochet; drift correction (nav.); It. <i>da capo</i>, repeat from the beginning (mus.) (Paxton 1983: 100)</p>

Example 6 shows a German abbreviation. In the entry there is no language qualifier, which shows the inconsistency of the dictionary compared to Example 5, but (*E. Ger.*) is an interesting datum showing the area of usage of the abbreviation.

Example 6

<p>DBD <i>Demokratische Bauernpartei Deutschlands</i>, Democratic Peasants' Party of Germany (<i>E. Ger.</i>) (Paxton 1983: 100)</p>

In Example 7 the entry word is followed by the expansion and an additional description of the term that might be useful for the user (i.e., "unit of noise measurement").

Example 7

<p>d/B decibel, unit of noise measurement (Paxton 1983: 100)</p>

If an abbreviation's expansion has more than one abbreviation, in the dictionary both abbreviations are included within one entry, as in Example 8 (i.e., *dbl.*, *dbble.*). The second abbreviation, *dbble.*, does not appear in a separate entry (an exception is seen in Example 9) because there is no entry for *dbble.* in the dictionary.

Example 8

<p>dbl., dbble. double (Paxton 1983: 100)</p>

Example 9 shows some inconsistency because there is no expansion of the abbreviation *d.b.* 'day book' in the entry *d.b.*, which follows the entry *D/B, d.b.*

Example 9

<p>D/B, d.b. day book d.b. double bed; double-breasted; draw bar (Paxton 1983: 100)</p>

2.4 *Abbreviations Dictionary* (De Sola 1986)

Abbreviations Dictionary (De Sola 1986) has twenty-three thousand alphabetically ordered entries and an extensive appendix. The dictionary mainly contains English abbreviations, but some foreign ones are also present. All foreign abbreviations are systematically translated. The entry is simple, composed of the entry word in bold followed by the expansions, alphabetically ordered and divided by semicolons, as in Example 10.

Example 10

<p>aud audible; audit; audition; auditor; auditorium (De Sola 1986: 100)</p>

There are no field qualifiers in the entries. As seen in Example 11, the entry word is followed by the foreign expansion written in italics (e.g., *ab urbe condita*) and the language qualifier, in parentheses and fully spelled out, the English translation, and an additional explanation, also inside the parentheses: "(Latin – from the founding of the city, usually refers to Rome)."

Example 11

<p>a.u.c. <i>ab urbe condita</i> (Latin – from the founding of the city, usually refers to Rome) (De Sola 1986: 100)</p>

2.5 *Dictionary of Abbreviations in Medical Sciences* (Heister 1989)

Dictionary of Abbreviations in Medical Sciences (Heister 1989) includes approximately eight thousand entries. The dictionary mainly includes English abbreviations in medical science, but some foreign ones are also present. All foreign

abbreviations are systematically translated. The entry is simple, composed of an entry word in bold and an expansion. Individual expansions are given in separate entries, as in Example 12. Example 12 also shows the use of field qualifiers in italics and abbreviated in parentheses; for example, (*pharm*), (*anat*). For the sixth expansion, "ACA (*pharm*) (epsilon)-aminocaproic acid(= EACA)," the prefix (*epsilon*)- is also added and an equivalent abbreviation in parentheses: (EACA).

Example 12

<p>ACA American Chiropractic Association ACA American College of Allergists ACA American College of Anesthesiologists ACA American College of Angiology ACA American College of Apothecaries ACA (<i>pharm</i>) (epsilon)-aminocaproic acid(= EACA) ACA (<i>anat</i>) anterior cerebral artery ACA automatic clinical analyzer ACA <i>ratio</i> ratio of accommodative convergence and accommodation (Heister 1989: 12)</p>

As already mentioned, foreign abbreviations are translated. The foreign abbreviations are mainly abbreviations of Latin origin, as in Example 13. The language qualifier is represented with only an abbreviated letter in parentheses (e.g., (*L*) 'Latin'). Field-specific qualifiers are also present in parentheses, in italics and abbreviated; for example, (*pharm*). As seen in Example 13, the entry word in bold is followed by abbreviated field-specific qualifiers in italics and in parentheses — for example, (*pharm*) — a foreign expansion and language qualifier — for example, (*L*) — and after a semicolon the English translation.

Example 13

<p>a.c. (<i>otorhin</i>) ad concham (<i>L</i>); just audible a.c. (<i>pharm</i>) ante cenam, ante cibos (<i>L</i>); before meals (Rx) (Heister 1989: 12)</p>
--

2.6 *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning* (Jung 1991)

The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning (Jung 1991) has fifty thousand entries. Every individual meaning or

expansion is in a separate entry. The entry word is written in bold and followed by the expansion. There are no translations of foreign abbreviations, as in Example 14 (e.g., *CST*, *CTA*), but there are field qualifiers in the entry — for example, (*Ss*) — but not consistently. Language qualifiers are not present, as for the abbreviation *CST*, but in the expansion *Comité des travailleurs Algériens* the qualifier (*F*) could denote the country (France) and/or the language (French).

Example 14

<p>CST Československá Televize CT Critical Theory (<i>Ss</i>) CTA Chicago Transit Authority CTA Comité des travailleurs Algériens (<i>F</i>) (Jung 1991: 100)</p>

2.7 *The Oxford Dictionary of Abbreviations* (Alexander et al. 1992)

The Oxford Dictionary of Abbreviations (Alexander et al. 1992) includes over twenty thousand abbreviations and is composed of an appendix with fourteen different fields (e.g., monetary, meteorological, and mathematical symbols, international vehicle registration marks, British postal codes, etc.). The dictionary mainly contains English abbreviations, but some foreign ones are also present. All foreign abbreviations are systematically translated. The entry is simple, composed of an entry word in bold and followed by expansions that are divided by the symbol •, as in Example 15. Example 15 also uses field qualifiers (e.g., *Psychol. Med.*) following the entry word, abbreviated and in italics, and also some additional information (e.g. "Siding Spring, NSW").

Example 15

<p>AAT <i>Psychol.</i> achievement anxiety test • <i>Med.</i> alpha-1-antitrypsin (treatment for cystic fibrosis) • Anglo-Australian Telescope (Siding Spring, NSW) • (Alexander et al. 1992: 4)</p>

The usage of language qualifiers is visible in Example 16, where the entry is composed of the entry word in bold followed by the expansion, and then in parentheses the language qualifier (e.g. *Latin*) followed by the foreign expansion in italics.

Example 16

AASS Associate of the American Antiquarian Society
(*Latin Americanae Antiquarianae Societatis Socius*)
(Alexander et al. 1992: 4)

2.8 *The Barnhart Abbreviations Dictionary* (Barnhart 1995)

The Barnhart Abbreviations Dictionary (Barnhart 1995) has over sixty thousand entries and is composed of two parts: the dictionary of abbreviations and the dictionary of expansions. The latter is mainly useful in the encoding process, when looking for an abbreviation. The dictionary mainly contains English abbreviations, but some foreign ones are also present. All foreign abbreviations are systematically translated, as in Example 17 and 18, where the foreign expansion is provided after the English translation. The entry is composed of an entry word in bold followed by the numbered, alphabetically ordered expansions and period. The dictionary includes field and language qualifiers, as in Example 17 and 18. If an expansion has more than one abbreviation, in the dictionary both abbreviations are included within one entry, as in Example 17 (i.e., AAC, A.A.C.).

Example 17

AAC
1 aeronautical approach chart. 2 Alaskan Air Command. 3 American Alpine Club. 4 *also* **A.A.C.** Association of American Colleges. 5 automatic amplitude control. 6 automatic approach control (in landing aircraft). 7 *also* **A.A.C.** in the year before Christ. [*Latin anno ante Christum*]
(Barnhart 1995: 2)

Example 18

A.B.U.
United Bible Societies. [*French Alliance Biblique Universelle*]
(Barnhart 1995: 6)

2.9 *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols* (Mattia 1997)

Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols (Mattia 1997) has sixty thousand entries, and every expansion or meaning is in a separate

entry. The entry is composed of an entry word in bold followed by the expansion. The dictionary does not include translations of foreign abbreviations, as in Example 19 (i.e., AAA, Accademia Arti Applicate) and field qualifiers (i.e., Military Science), but there is an appendix, where the abbreviations are organized in separate fields. The entry is simple; each individual expansion appears in a separate entry, as in Example 19. If an expansion has more than one abbreviation, in the dictionary both abbreviations are included within one entry, as in Example 19 (i.e., AAA A.A.A.).

Example 19

<p>AAA Accademia Arti Applicate — Italy; since 1956. AAA Agricultural Adjustment Administration. AAA Amateur Athletic Association. AAA American Accounting Association. AAA American Arbitration Association. AAA; A.A.A. American Association of Anatomists (U.S.A.). AAA American Automobile Association. AAA Antiaircraft Artillery. <i>Military Science.</i> (Mattia 1997)</p>
--

2.10 *Abbreviations and Acronyms* (Geddes & Grosset 1999)

The pocket dictionary *Abbreviations and Acronyms* (Geddes & Grosset 1999) has around ten thousand alphabetically ordered entries. The entry word in bold is followed by the expansions. The dictionary has English and foreign abbreviations, and all foreign abbreviations are consistently translated, as seen in Example 20. Language and field qualifiers are included in the dictionary, and there is no appendix.

Example 20

<p>BMW <i>Bayerische Motoren Werke</i>, German 'Bavarian Motor Works' (Geddes & Grosset 1999: 38)</p>
--

2.11 *The Wordsworth Dictionary of Abbreviations and Acronyms* (Dale and Puttick 1999)

The Wordsworth Dictionary of Abbreviations and Acronyms (Dale and Puttick 1999) includes twenty thousand alphabetically ordered entries and has an extensive

alphabetically ordered appendix of over 150 pages, where abbreviations from music, medicine, veterinary, zoology, geography, meteorology, and chemistry are found. The abbreviations from the appendix are not included among the entries. The appendix is useful, but it may confuse readers due to its extensiveness. It is not clear where to find the abbreviations: among the entries or in the appendix. The dictionary mainly contains English abbreviations, but there are also foreign abbreviations and they are consistently translated. The entry is composed of the entry word in bold followed by expansions alphabetically ordered and divided by semicolons, as in Example 21. Within an entry all expansions of an abbreviation sharing the same root are included in the same entry separated by semicolons, as seen in Example 21.

Example 21

<p>mech mechanical; mechanism; mechanic; mechanics; mechanical (Dale and Puttick 1999: 100)</p>
--

Completely divided in a separate entry are all other expansions with a different word root. They share the same abbreviation, as seen in Example 22. Example 22 uses field qualifiers in parentheses — for example, (*MIL*) — and language qualifiers (e.g., *Latin*) following the foreign expansion; the language qualifiers are written as full words in italics and are followed by a translation.

Example 22

<p>MC Member of Congress (USA) MC Master of Ceremonies MC medium capacity (<i>bomb</i>) (<i>MIL</i>) MC Military Cross (<i>MIL</i>) MC medium coeli (<i>Latin: mid-heaven</i>) (<i>SCI</i>) (Dale and Puttick 1999: 100)</p>

If one expansion has two abbreviations, both are included in one entry as two entry words, as seen in Example 23. The second entry word (i.e., *MMDS*) is not separately included as an individual entry.

Example 23

<p>MDS, MMDS multipoint microwave distribution system (FPA) (Dale and Puttick 1999: 101)</p>

The inclusion of encyclopedic data is visible in Example 24 — that is, (*paper which has been surfaced while on the papermaking machine*) — where it is written in italics in parentheses, and it follows the expansion.

Example 24

<p>MF machine-finished (<i>paper which has been surfaced while on the papermaking machine</i>) (COM) (Dale and Puttick 1999: 101)</p>
--

2.12 *The New Penguin Dictionary of Abbreviations* (Fergusson 2000)

The New Penguin Dictionary of Abbreviations (Fergusson 2000) includes approximately forty thousand alphabetically ordered entries, and at the beginning of the dictionary there is an extensive appendix including abbreviations used in advertising, foreign abbreviations from business, chemical elements, and other items. The useful appendix is divided into alphabetically ordered fields. The dictionary mainly contains English abbreviations, but there are also some foreign abbreviations. All foreign abbreviations are consistently translated. The entry is simple, composed of the entry word in bold followed by the expansions divided by semicolons, as in Example 25. In Example 25 the foreign expansion is written in italics, and the language qualifier (e.g., *French*) is written in full and placed after the foreign expansion in parentheses and followed by the translation. In the entry there are field qualifiers written in full in square brackets (e.g., *[astronomy]*). The entry also indicates the area of usage of the abbreviation, also in square brackets (e.g., *[USA]*).

Example 25

<p>CME Chicago Mercantile Exchange; <i>Conférence mondiale de l'énergie</i> (French: World Power Conference); [astronomy] coronal mass ejection; [USA] cost and manufacturability expert (Fergusson 2000: 100)</p>

The inclusion of encyclopedic data or additional information is also visible among a limited number of entries, as in Examples 26 and 27; for example, (*in parliamentary procedures*), (*shares*).

Example 26

Cmd Command Paper (1919–56) (in parliamentary procedures)
(Fergusson 2000: 100)

Example 27

cm. pf. cumulative preference (shares); cumulative preferred (shares)
(Fergusson 2000: 100)

2.13 *Dictionary of Financial Abbreviations* (Paxton 2003)

Dictionary of Financial Abbreviations (Paxton 2003) is composed of over seven thousand entries from finance and an appendix. The six units of the appendix include monetary symbols, financial reporting standards, statements of standard accounting practice, and other topics. The entry is simple; each individual expansion appears in a separate entry, as in Example 28. The entry is composed of an entry word in bold followed by the expansion.

Example 28

AA Associate in Accounting
AA attendance allowance (UK)
(Paxton 2003: 11)

The dictionary mainly contains English abbreviations, but some foreign ones are also present. All foreign abbreviations are systematically translated, as in Example 29. Example 29 contains three entries for the same abbreviation but different expansions. Each abbreviation has a separate entry; the entry word in bold is followed by the foreign expansion in italics (e.g., *accepté*), the translation (after a period), and the country of origin in parentheses (e.g., *France, Germany*).

Example 29

A *accepté*.
Accepted. (France)
A accepted
A *akzeptiert*. Accepted. (Germany)
(Paxton 2003: 11)

As seen in Example 30, some entries include additional information, where following the expansion there is information on when or where the association was founded (e.g., "Founded in London in 1869"). The additional data may function as encyclopedic data (e.g., "Association then influences the development of generally accepted accounting principles through published research of its professional membership").

Example 30

<p>AAA American Accounting Association. Association then influences the development of generally accepted accounting principles through published research of its professional membership. AAA Association of Average Adjusters. Founded in London in 1869. (Paxton 2003: 11)</p>

2.14 *Acronyms, Initialisms and Abbreviations Dictionary* (Romaniuk 2006)

Acronyms, Initialisms and Abbreviations Dictionary (Romaniuk 2006) is an extensive dictionary of abbreviations covering over a million abbreviations. The dictionary is published almost every year, and it is composed of an extensive number of entries for which every meaning or expansion is included as a separate entry. The entry is simple, composed of an entry word in bold and followed by the meaning or expansion (seen in Example 31). The entry also uses field qualifiers written as a full word in square brackets (e.g., [*Geology*]), and encyclopedic data in square brackets — for example, "[*Library symbol*] [*National Library of Canada*] (NLC)" — as in Examples 31 and 32. Translations of foreign abbreviations are consistent in the dictionary.

Example 31

<p>A – Age A – Agglomerate [<i>Geology</i>] (Romaniuk 2006: 5)</p>
--

Example 32

<p>AA – Andrew Public Library, Alberta [<i>Library symbol</i>] [<i>National Library of Canada</i>] (NLC) (Romaniuk 2006: 12)</p>

Attention should also be drawn to an interesting dictionary of abbreviation expansions, *Reverse Acronyms Initialisms and Abbreviations Dictionary* (Gale Research Company 2007). This dictionary is a reversed version of the abbreviation dictionary *Acronyms, Initialisms and Abbreviations Dictionary* (Romaniuk 2006) and allows users to find an abbreviation through an extensive list of expansions. This dictionary has been published regularly for some years as a separate publication.

2.15 *Abbreviations* (King 2014)

The dictionary *Abbreviations* (King 2014) contains over seven thousand abbreviations and has an appendix with chemical elements, European car registration letters, UK postcodes, abbreviations related to the UN, airline acronyms, and other topics. The entry is simple, composed of an entry word in bold followed by expansions divided by semicolons, as in Example 33. The dictionary mainly contains English abbreviation, but some foreign ones are also present. All foreign abbreviations are systematically translated.

Example 33

<p>AA Alcoholics Anonymous; advertising agency; age allowance; Air Attaché; American Airlines; Anglers' Association; anti-aircraft; Associate in Accounting; Associate in Agriculture; Automobile Association (King 2014: 3)</p>

Example 34 includes some additional data; that is, (*proofreading*).

Example 34

<p>aa absolute alcohol; after arrival; attendance allowance; author's alteration (<i>proofreading</i>) (King 2014: 3)</p>
--

The dictionary also includes foreign abbreviations, as in Examples 35 and 36. In Example 35 the German abbreviations *ABS* is translated into English, but no language qualifier is provided. In Example 36 there is a French expansion in italics, followed by an explanation in parentheses, but there is no language qualifier.

Example 35

ABS *anti-blockier* system = anti-lock brakes
(King 2014: 6)

Example 36

AC alternating current; *appellation controlee* (quality control of French wines); Arts Council
(King 2014: 6)

2.16 *Dorland's Dictionary of Medical Acronyms and Abbreviations* (Dorland 2016)

Dorland's Dictionary of Medical Acronyms and Abbreviations (Dorland 2016) has approximately thirty thousand alphabetically ordered entries from medicine. At the front of the dictionary there is a section dedicated to symbols, where the Greek alphabet is found, symbols used in mathematics, statistics and so on. The section is followed by a section called "Abbreviations to avoid," in which the Joint Commission and the Institute for Safe Medication Practice list abbreviations not to be used, mainly due to homonymity; for example, *IN* 'intranasal' could be mistaken for *IM* or *IV*, and for that reason it is advisable to write the entire word (i.e., *intranasal*) or use the abbreviation *NAS*. The dictionary mainly contains English abbreviations, but there are also some foreign ones. All foreign abbreviations are systematically translated. As seen in Example 37, after the entry word in bold there is the expansion or translation, the language qualifier is in square brackets (abbreviated, e.g., *Lat.*) and followed by the foreign expansion in italics and inside the same square brackets (e.g., *contra*). Expansions are divided by semicolons, and synonyms by commas (e.g., *continuation, continue*).

Example 37

con against [*Lat. contra*]; continuation, continue
(Dorland 2016: 90)

As seen in Example 38, the entry is simple: the entry word in bold is followed by expansions divided by semicolons.

Example 38

AACP American Academy of Cerebral Palsy;
American Association of Colleges of Pharmacy;
American Association of Community Psychiatrists
(Dorland 2016: 2)

2.17 *Abbreviations Dictionary* (Stahl and Landen 2019)

Abbreviations Dictionary (Stahl and Landen 2019) is a dictionary started by De Sola and continued by Stahl and Landen, and it includes approximately sixty thousand alphabetically ordered entries and an extensive appendix composed of sixty-three fields, among them airlines, airports, currencies of the world, emoticons, inventions and inventors, national holidays, national parks in Canada and the United States, symbols (e.g., U.S. naval ship), vehicle registrations, weather, and wedding anniversaries. The dictionary mainly contains English abbreviations, but there are also some foreign ones. All foreign abbreviations are systematically translated. The entry is simple, composed of the entry word in bold followed by the expansions divided by semicolons, as in Examples 39 and 40. In Example 39 the foreign expansion is written in italics (e.g., *ana*) and followed by the fully spelled-out language qualifiers in parentheses (e.g., *Greek*) and the translation. The expansions are divided by semicolons.

Example 39

aa *ana* (Greek – of each); *ander andere* (German – among others); *arterias* (Latin – arteries); (Hawaiian – block lava) – pronounced *ah-ah*
(Stahl and Landen 2019: 2)

Example 40

FBI Fast Boats Incorporated; Federal Bureau of Investigation; Federation of British Industries; Food Business Institute; full-blooded Irishman, Icelander, Indian, Indonesian, Iranian, Iraqi, Israelite, Italian, or Ivory Coaster
(Stahl and Landen 2019: 398)

The overview of the macro- and microstructure of the English dictionaries of abbreviations included seventeen dictionaries published from 1942 to 2019.

Except for *Acronyms, Initialisms and Abbreviations Dictionary* (Romaniuk 2006), they were not published online and as such the entries are not frequently updated. In terms of macrostructure, the dictionaries presented differ in the number of entries included. The detailed analysis of the microstructure of the entries shows inconsistency among the entries (e.g., the inclusion of foreign abbreviations and often exclusion of translations). As seen from the dictionaries analyzed, the entries differ in a variety of aspects, from layout (e.g., the inclusion of special symbols, commas, semicolons, and other stylistic elements such as bold, italics, and parentheses) to content, where the entries differ in terms of inclusion of translations, language and field qualifiers, and also additional and encyclopedic data, unified or separate entries, and so on. Common to all the dictionaries presented is the fact that they are not easily accessed and outdated, compared to the freely available online dictionaries of abbreviations presented below. Regarding easy access, some of the dictionaries analyzed offer a limited number of digitalized pages of the dictionary freely available online, but access to the entire book is restricted to specialized libraries. The preparation of this article involved three specialized libraries in three different countries (Belgium, Spain, and Slovenia).

3. Online dictionaries of abbreviations

The following section presents freely available online dictionaries of abbreviations that represent alternatives to paper ones because they could be easily updated. When using a dictionary of abbreviations, it is very important to use a dictionary that is not too old (Kompara Lukančič 2010). As argued by Lan (2005), the days of the predominance of paper dictionaries are gone and, according to his study conducted at The Hong Kong Polytechnic University, online dictionaries or e-dictionaries are more often used compared to traditional paper ones. Similar analyses showing that the majority of students do not use traditional paper dictionaries were also conducted by Collins (2016) and Dziemianko (2010). Even if abbreviations are found in monolingual online dictionaries, such as *The Free Dictionary* (2020), it was decided to observe only online abbreviation dictionaries. *A Dictionary of Abbreviations* (Vance 2020), *The Acronym Finder* (2020), and *The Abbreviations* (2020) were examined and are presented below.

3.1 *A Dictionary of Abbreviations* (Vance 2020)

A Dictionary of Abbreviations (Vance 2020) is an online dictionary that includes over one hundred thousand alphabetically ordered abbreviations. Among the entries are also military and police abbreviations, and abbreviations from medicine, finance, computer science, and other fields. Every individual meaning or expansion is included in a separate entry, as in Figure 1. Figure 1 shows that some entries have a field qualifier (e.g., *Airline* in square brackets) next to

the entry word. The dictionary offers an extensive list of abbreviations, but access to the expansions is time-consuming because the user must click on each individual entry word in order to see the expansion or meaning.

Figure 1

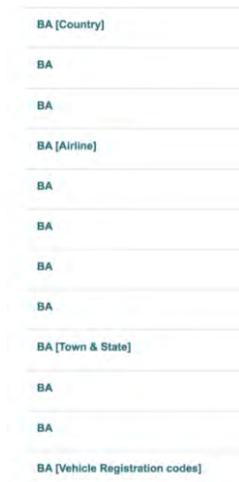


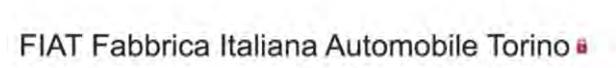
Figure 2 shows the microstructure of the entry. The entry is simple, composed of the entry word, followed by the field qualifier in square brackets and the English expansion.

Figure 2



The dictionary mainly includes English abbreviations, but there are also some foreign abbreviations (e.g., *FIAT*). Unfortunately, the foreign abbreviations are not translated into English and there is also no description, or language or field qualifier. As seen in Figure 3, the dictionary article is composed of the entry word and foreign expansion.

Figure 3



3.2 *The Acronym Finder* (2020)

The Acronym Finder (2020) includes more than a million entries. The entries mainly include English abbreviations, but foreign ones are also present. As seen in Figure 4, every expansion is in an individual entry, but the expansions are visible without clicking on an individual entry. When users click on the entries, they are redirected to another source (e.g., *The Free Dictionary*). The entry is simple, composed of only the entry word and its expansion. There are no language or field qualifiers, or additional or encyclopedic data.

Figure 4



The dictionary also contains foreign abbreviations, as seen in Figure 5 (e.g., *FIAT*). The entry is simple, composed of only the entry word and an expansion. When clicking on the first entry word, the user is redirected to *The Free Dictionary* (2020). In the second and third expansions there is no redirection to a source and there are also no translations. The entry lacks language and field qualifiers, and translations.

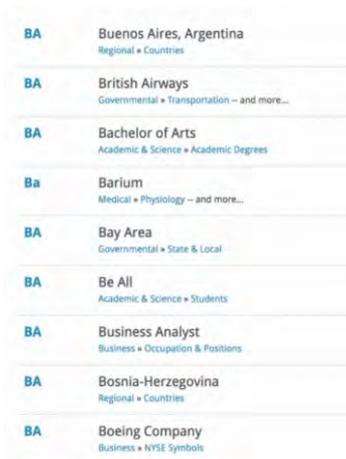
Figure 5



3.3 *The Abbreviations* (2020)

The Abbreviations (2020) has an extensive number of abbreviations. The entries include English and foreign abbreviations. As seen in Figure 6, every individual expansion is in a separate entry. The entry is simple, composed of an entry word, expansion, and field qualifier, which was absent in the previous two dictionaries.

Figure 6



BA	Buenos Aires, Argentina Regional » Countries
BA	British Airways Governmental » Transportation – and more...
BA	Bachelor of Arts Academic & Science » Academic Degrees
Ba	Barium Medical » Physiology – and more...
BA	Bay Area Governmental » State & Local
BA	Be All Academic & Science » Students
BA	Business Analyst Business » Occupation & Positions
BA	Bosnia-Herzegovina Regional » Countries
BA	Boeing Company Business » NYSE Symbols

As seen in Figure 7, foreign abbreviations are also present (e.g., *FIAT*). The entry is simple, composed of an entry word, foreign expansion, and field qualifier. There is no language qualifier and translation, or description of the foreign expansion.

Figure 7



FIAT	Fabbrica Italiana Automobili Torino Business » Companies & Firms
-------------	---

On one hand, the online dictionaries analyzed show more extensive inclusion of entries (e.g., reaching the number of one hundred million entries). On the other hand, they have a poorer entry structure, mainly composed of the entry word and its expansion. In the dictionaries analyzed, among the entries there are inconsistencies in the inclusion of translations, language qualifiers, and field qualifiers. The online accessibility of the dictionary is certainly an advan-

tage, but the poor composition of the entry is not much of a help to users. An attempt to join the online approach to dictionary compilation and the structure of more traditional paper dictionaries of abbreviations is presented also in the compilation of two Slovene dictionaries of abbreviations.

4. Conclusion

As it was explained in detail above, the English dictionaries of abbreviations not published online differ both in terms of micro- and macrostructure. The analyzed paper dictionaries differ in the number of entries included. We encounter inconsistencies among the inclusion of foreign abbreviations, the inclusion/exclusion of translations, the different layouts used, and the content, where differences are seen in the inclusion of language and field qualifiers, additional data, etc. In reference to the paper dictionaries presented, a common issue is the fact that they are of limited access (in specialized libraries) and mainly outdated. Compared to the paper ones, on one hand the analyzed online dictionaries show a more extensive inclusion of entries, periodically updated and freely accessed, but on the other hand they have a poorer entry structure, mainly composed of the entry word and its expansions. The analyzed dictionaries show inconsistencies in the inclusion of translations, language qualifiers and field qualifiers. We agree that compared to paper dictionaries the online accessibility is an advantage, but the poor composition of the entry should be considered in the future of the compilation of contemporary dictionaries of abbreviations. We believe that functional, contemporary, structurally adequate dictionaries of abbreviations will be the result of combining the concepts of traditional and modern lexicography.

References

- Alexander, F. et al. (Eds.). 1992. *The Oxford Dictionary of Abbreviations*. Oxford: Oxford University Press.
- Barnhart, K.R. (Ed.). 1995. *The Barnhart Abbreviations Dictionary*. New York: John Wiley & Sons.
- Buttress, F.A. 1976. *World Guide to Abbreviations of Organizations*. London/Detroit: Gale Research.
- Collins, J.B. 2016. Changes in Electronic Dictionary Usage Patterns in the Age of Free Online Dictionaries: Implications for Vocabulary Acquisition. Korenaga, S. (Ed.). 2016. *APU Journal of Language Research* 1: 36-49.
- Čermák, F. and R. Blatná. 1995. *Manuál lexikografie*. Prague: H&H.
- Dale, R. and S. Puttick. 1999. *The Wordsworth Dictionary of Abbreviations and Acronyms*. Hertfordshire: Wordsworth.
- De Sola, R. 1986. *Abbreviations Dictionary*. New York: Elsevier.
- Dorland, A. 2016. *Dorland's Dictionary of Medical Acronyms and Abbreviations*. Philadelphia: Elsevier.
- Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

- Fergusson, R.** 2000. *The New Penguin Dictionary of Abbreviations*. London: Penguin.
- Gabrovšek, D.** 1994. Kodifikacija angleškega jezika v specializiranih enojezičnih slovarjih: Too much of Everything? Društvo za tuje jezike in književnosti (Ed.). 1994. *Vestnik* 28 (1–2): 150-180. Ljubljana: Društvo za tuje jezike in književnost R Slovenije.
- Gale Research Company.** 2007. *Reverse Acronyms Initialisms and Abbreviations Dictionary*. Detroit: Gale Research Company.
- Geddes & Grosset.** 1999. *Abbreviations and Acronyms*. New Lanark: Geddes & Grosset.
- Heister, R.** 1989. *Dictionary of Abbreviations in Medical Sciences*. Berlin/Heidelberg/New York: Springer.
- Jung, H.** 1991. *The Dictionary of Acronyms and Abbreviations in Applied Linguistics and Language Learning*. Frankfurt am Main: Peter Lang.
- King, G.** 2014. *Abbreviations*. Glasgow: Harper Collins.
- Kompara Lukančič, M.** 2006. *Slovarček krajšav*. Ljubljana: Inštitut za slovenski jezik Fran Ramovš ZRC SAZU.
- Kompara Lukančič, M.** 2009. Prepoznavanje krajšav v besedilih. Weiss, P. (Ed.). 2009. *Jezikoslovní zapiski* 15(1–2): 95-112. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.
- Kompara Lukančič, M.** 2010. Krajšavni slovarji. Weiss, P. (Ed.). 2010. *Jezikoslovní zapiski* 16(2): 111-129. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.
- Kompara Lukančič, M.** 2011. *Slovar krajšav*. Kamnik: Amebis, Termania.
- Kompara Lukančič, M.** 2017. Zasnova novega slovarja krajšav. Weiss, P. (Ed.). 2017. *Jezikoslovní zapiski* 23(1): 77-92. Ljubljana: Inštitut za slovenski jezik Frana Ramovša.
- Lan, L.** 2005. The Growing Prosperity of On-line Dictionaries. *English Today* 21(3): 16-21. doi. <https://doi.org/10.1017/S0266078405003044>.
- Logar, N.** 2005. Norma v slovarju sodobne slovenščine: zloženke in kratice. Družboslovne razprave (Ed.). 2005. *Družboslovne razprave* 21(48): 211-225. Ljubljana: Slovensko sociološko društvo: Fakulteta za družbene vede.
- Mattia, F.B.** 1997. *Elsevier's Dictionary of Acronyms, Initialisms, Abbreviations and Symbols*. Amsterdam: Elsevier Science
- Nielsen, S.** 1990. Lexicographic Macrostructure. *HERMES — Journal of Language and Communication in Business* 3(4): 49-66.
- Partridge, E.** 1942. *A Dictionary of Abbreviations: With Special Attention to War-Time Abbreviations*. London: Routledge.
- Paxton, J.** 1983. *Everyman's Dictionary of Abbreviations*. London: Dent & Sons.
- Paxton, J.** 2003. *Dictionary of Financial Abbreviations*. New York: Routledge.
- Romaniuk, B.** 2006. *Acronyms, Initialisms and Abbreviations Dictionary*. Detroit: Gale Research.
- Stahl, D.A. and K. Landen.** 2019. *Abbreviations Dictionary*. New York: Routledge Revivals.
- Termania*: <http://www.termania.net/>
Date of Access: 7 April 2020.
- The Abbreviations*: <https://www.abbreviations.com/>
Date of Access: 7 April 2020.
- The Acronym Finder*: <https://www.acronymfinder.com/>
Date of Access: 7 April 2020.
- The Free Dictionary*: <https://www.thefreedictionary.com/>
Date of Access: 7 April 2020.

Vance, B. 2020. *A Dictionary of Abbreviations*:

<https://www.oxfordreference.com/view/10.1093/acref/9780199698295.001.0001/acref-9780199698295>

Date of Access: 7 April 2020.

Zidar, J. 1971. *Rečnik jugoslovenskih skraćenica*. Beograd: Međunarodna politika.

Župančič, J. 1948. *Kratice*. Ljubljana: Državna založba Slovenije.

Der LeGeDe-Prototyp: Zur Erstellung eines korpusbasierten Online-Wörterbuchs zu lexikalischen Besonderheiten im gesprochenen Deutsch. Aktueller Stand und Perspektiven

Meike Meliss, *Universidad de Santiago de Compostela,
Spanien (meike.meliss@usc.es)*

und

Christine Möhrs, *Leibniz-Institut für Deutsche Sprache, Mannheim,
Deutschland (moehrs@ids-mannheim.de)*

Zusammenfassung: Im Beitrag steht das LeGeDe-Drittmittelprojekt und der im Laufe der Projektzeit entwickelte korpusbasierte lexikografische Prototyp zu Besonderheiten des gesprochenen Deutsch in der Interaktion im Zentrum der Betrachtung. Die Entwicklung einer lexikografischen Ressource dieser Art knüpft an die vielfältigen Erfahrungen in der Erstellung von korpusbasierten Onlinewörterbüchern (insbesondere am Leibniz-Institut für Deutsche Sprache, Mannheim) und an aktuelle Methoden der korpusbasierten Lexikologie sowie der Interaktionsanalyse an und nimmt als multimedialer Prototyp für die korpusbasierte lexikografische Behandlung von gesprochen-sprachlichen Phänomenen eine innovative Position in der modernen Onlinelexikografie ein. Der Beitrag befasst sich im Abschnitt zur LeGeDe-Projektpräsentation ausführlich mit projektrelevanten Forschungsfragen, Projektzielen, der empirischen Datengrundlage und empirisch erhobenen Erwartungshaltungen an eine Ressource zum gesprochenen Deutsch. Die Darstellung der komplexen Struktur des LeGeDe-Prototyps wird mit zahlreichen Beispielen illustriert. In Verbindung mit der zentralen Information zur Makro- und Mikrostruktur und den lexikografischen Umtexten werden die vielfältigen Vernetzungs- und Zugriffsstrukturen aufgezeigt. Ergänzend zum abschließenden Fazit liefert der Beitrag in einem Ausblick umfangreiche Vorschläge für die zukünftige lexikografische Arbeit mit gesprochensprachlichen Korpusdaten.

Stichwörter: INTERAKTIONSLINGUISTIK, KORPUSLINGUISTIK, LEXIK DES GESPROCHENEN, MAKROSTRUKTUR, MIKROSTRUKTUR, MULTIMEDIALITÄT, ONLINELEXIKOGRAFIE, UMTEXTE, VERNETZUNGSSTRUKTUREN

Abstract: The LeGeDe-prototype: The Creation of an Online Dictionary on Lexical Particularities in Spoken German. Current Status and Perspectives.

The article focuses on the LeGeDe third-party funded project and the corpus-based lexicographic prototype on the particularities of spoken German in interaction developed during the project period. The development of a lexicographic resource of this kind builds on the various experiences in the creation of corpus-based online dictionaries (especially at the Leibniz Institute for the German Language, Mannheim) and on methods of corpus-based lexicology as well as interaction analysis. As a multimedia prototype for the corpus-based lexicographic treatment of spoken language phenomena, it occupies an innovative position in the landscape of online lexicography. In the section on the presentation of the LeGeDe-project, the article deals in detail with research questions relevant to the project, the project goals, the corpus data basis and empirically determined expectations of a resource on spoken German. The presentation of the complex structure of the LeGeDe-prototype is illustrated with numerous examples. Apart from the central information on the macro- and microstructure and the lexicographical outer texts, the various linking and access structures are shown. In addition to the final conclusion, the article concludes with an outlook that provides extensive suggestions for future lexicographic work with spoken language corpus data.

Keywords: INTERACTIONAL LINGUISTICS, CORPUS LINGUISTICS, LEXICOLOGY OF SPOKEN LANGUAGE, MACROSTRUCTURE, MICROSTRUCTURE, MULTIMEDIALITY, ONLINE LEXICOGRAPHY, OUTER TEXTS, LINKING

1. Einleitung

Einsprachige Wörterbücher zum Deutschen gibt es in zahlreichen Varianten: Verlagspublikationen wie der *Duden* (in verschiedenen Reihen und Publikationsformen) oder lexikografische Nachschlagewerke des Verlags von *Langenscheidt* (wie das *Großwörterbuch Deutsch als Fremdsprache*) stehen neben Wörterbüchern, die in wissenschaftlichen Akademien oder Forschungsinstituten erstellt werden, wie z.B. das *DWDS* der Berlin-Brandenburgischen Akademie der Wissenschaften in Berlin oder *lexiko*, das am Leibniz-Institut für Deutsche Sprache (=IDS) in Mannheim konzipiert wurde. Die Bandbreite an Wörterbuchtypen reicht dabei von Standardwörterbüchern (z.B. *Duden-Universalwörterbuch*) bis hin zu ausgewiesenen Spezialwörterbüchern (z.B. *Wörterbuch der deutsch-lothringischen Mundarten=LothWB*). Ebenso vielfältig ist das Angebot für unterschiedliche Zielgruppen (Muttersprachler/-innen, Fachspezialist/-innen, Lerner/-innen des Deutschen als Fremd- oder Zweitsprache etc.). Neben dem Printformat werden die unterschiedlichen Wörterbuchtypen des Deutschen zunehmend online angeboten. Einige der neu konzipierten Wörterbücher, die nicht auf einer vorausgehenden Printversion basieren, sind sogar ausschließlich online zugänglich (vgl. *lexiko*)*.

In dieser Fülle von verschiedenen Nachschlagewerken zur deutschen Sprache gibt es trotz unterschiedlichem inhaltlichen Fokus eine Gemeinsam-

keit: Geht es um ein ausgewiesenes korpusgestütztes Wörterbuch, so handelt es sich bei der empirischen Datengrundlage in der Regel um Sprachkorpora geschriebensprachlicher Daten. Die herangezogene Datenbasis kann insbesondere bei Verlagswörterbüchern und frei zugänglichen lexikografischen Ressourcen (wie z.B. *Linguee*) jedoch nicht immer vollständig transparent nachvollzogen werden (vgl. Tarp 2019: 227). Meist findet sich in den Umtexten eine ungenaue Formulierung wie: „Grundlage für die möglichst exakte Darstellung der Wortbedeutungen sind das umfangreiche Dudenkorpus sowie die nahezu unbegrenzten Recherchemöglichkeiten des Internets“ (aus: Vorwort *Duden-DaF*: 5). Beim Wörterbuch *lexiko* lässt sich hingegen sehr transparent nachvollziehen, mit welchen Quellen gearbeitet wurde, da in den Umtexten detailliert auf die Korpusbasis, die für die Wortartikelarbeit herangezogen wurde, eingegangen wird.¹

Im Gegensatz zur deutschen Schriftsprache als korpusbasierte Datengrundlage wurde Material zum gesprochenen Deutsch insbesondere Daten aus authentischen mündlichen Interaktionssituationen in den meisten Wörterbüchern bislang eher vernachlässigt. Zwar wird in den bislang existierenden Wörterbüchern vereinzelt mit Angaben wie „gesprochen“ oder „umgangssprachlich“ (vgl. *Duden-online* und *DWDS* zu *gucken*) auf Besonderheiten im Gesprochenen hingewiesen, doch bleibt die Datenbasis oft eine unbekannte bzw. unpräzise Variable.

Demzufolge gibt es eine lexikografische Beschreibungslücke hinsichtlich lexikalischer Besonderheiten des gesprochenen Deutsch. Diese Tatsache hängt in sehr starkem Maße mit der Verfügbarkeit geeigneter Korpora zusammen, denn nur eine ausreichend große Datenbasis kann es ermöglichen, korpusgestützte Methoden zur Erstellung einer lexikografischen Ressource zu entwickeln.

Mit dem „Forschungs- und Lehrkorpus Gesprochenes Deutsch“ (=FOLK) wird am IDS seit einigen Jahren ein Korpus aufgebaut, das authentisches Gesprächsmaterial aus natürlicher und spontaner Interaktion überwiegend zum standardnahen Deutsch über eine online verfügbare Datenbank (DGD= Datenbank für Gesprochenes Deutsch) recherchierbar macht. Das Vorhandensein von FOLK als einer adäquaten Datengrundlage zur Dokumentation und Erforschung des gesprochenen Deutsch in der Interaktion in Verbindung mit der am IDS existierenden Expertise im Aufbau von Online-Wörterbüchern erschließt neue Wege. Im von der Leibniz-Gemeinschaft² geförderten Drittmittelprojekt „Lexik des gesprochenen Deutsch“ (=LeGeDe³) wurde in drei Jahren (von September 2016 bis September 2019) ein erster lexikografischer Prototyp zu lexikalischen Besonderheiten im gesprochenen Deutsch entwickelt. Der LeGeDe-Prototyp ist in die Plattform OWID^{plus} eingebettet und über die URL <https://www.owid.de/legede/kostenfrei> seit September 2019 abrufbar. Die erstellte lexikografische Ressource verfolgt zunächst primär das Ziel der Sprachdokumentation und spricht daher als mögliche Zielgruppe Sprachwissenschaftler/-innen verschiedener Disziplinen aber mit Einschränkung auch

Sprachlehrende als Übermittler/-innen an. In diesem Sinne werden auch Sprachlernende als Zielgruppe der Ressource indirekt mitangedacht (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019) (vgl. Abschnitt 2.2 (iii) und (iv)).

Der vorliegende Artikel bietet einen Überblick zum LeGeDe-Projekt (vgl. Abschnitt 2) und eine ausführliche Darstellung der Projektergebnisse (vgl. Abschnitt 3). In Abschnitt 2 werden Informationen zum Projektrahmen, zu Annahmen und Forschungsfragen sowie zu Projektzielen und der Datengrundlage geliefert. Außerdem wird ein Einblick in die Ergebnisse von empirischen Studien zu Erwartungshaltungen an eine zukünftige lexikografische Ressource zu Besonderheiten des Gesprochenen geboten. Im Zentrum von Abschnitt 3 steht insbesondere die Präsentation des LeGeDe-Prototyps. Dabei werden vor allem Aspekte in Verbindung mit der komplexen lexikografischen Struktur (Makrostruktur, Mikrostruktur, Vernetzungs- und Verweisstrukturen, Umtexte) thematisiert. Neben einem Fazit wird in Abschnitt 4 ein Ausblick angeboten, der umfangreiche Vorschläge für die zukünftige lexikografische Arbeit mit gesprochensprachlichen Korpusdaten liefert.

2. Das LeGeDe-Projekt

2.1 Projektrahmen

Das LeGeDe-Projekt hatte zum Hauptziel, einen korpusbasierten, lexikografischen Prototyp zu Besonderheiten des gesprochenen Deutsch in der Interaktion zu erstellen (vgl. Abschnitt 2.3). Die Kooperation zwischen den Abteilungen Pragmatik und Lexik des IDS ermöglichte während der Projektlaufzeit eine Verbindung der entsprechenden interdisziplinären Fachkompetenzen.

Für die Entwicklung des LeGeDe-Prototyps wurden im Projekt sowohl quantitative als auch qualitative Methoden entwickelt, mit denen die Spezifika gesprochensprachlicher Lexik des Deutschen auf der Basis der am IDS erstellten mündlichen Korpora (im Programmbereich „Mündliche Korpora“ der Abteilung Pragmatik) im Vergleich zu Teilkorpora der geschriebenen Sprache identifiziert, analysiert und für die lexikografische Anwendung aufbereitet werden konnten. Neuartige lexikografische Beschreibungsformate für lexikalische Daten aus Gesprächskorpora in audioelektronischer Form, ebenso wie innovative lexikografische Angabetypen, die auf die Funktion lexikalischer Einheiten in interaktionalen Kontexten Bezug nehmen, mussten daher im Laufe der Projektzeit für die Erfüllung der Projektziele entwickelt werden (vgl. dazu insbesondere These 10 aus Engelberg, Klosa-Kückelhaus und Müller-Spitzer 2019: 33). Bei diesem Prozess konnten auch einige der Erwartungen von zukünftigen Nutzer/-innen, die durch Umfragen empirisch erhoben wurden, berücksichtigt werden (vgl. Abschnitt 2.5).

2.2 Annahmen und Forschungsfragen

Das LeGeDe-Projekt basierte auf folgenden Hauptannahmen und Beobachtungen, aus denen sich entsprechende Forschungsfragen ableiten ließen:

(i) Es existieren Unterschiede auf verschiedenen sprachlichen Ebenen zwischen dem gesprochenen und dem geschriebenen Deutsch. Bezüglich der Lexik können sich die Divergenzen sowohl im Bestand als auch in Verbindung mit deren Form, Bedeutung und Verwendung auswirken (vgl. Deppermann, Proske und Zeschel 2017; Fiehler 2016; Imo 2007; Schwitalla 2012). Aus dieser Beobachtung heraus ergab sich die Frage, welche korpusbasierten Methoden geeignet sind, um diese Divergenzen zu identifizieren und für eine lexikografische Ressource nutzbar zu machen.

(ii) Die lexikografische Kodifizierung der interaktionstypischen Besonderheiten der gesprochensprachlichen Lexik des Deutschen ist bis jetzt unzureichend, obwohl eine beachtliche Menge an lexikalischen Studien mit interaktionslinguistischem Ansatz vorliegen (vgl. Meliss 2016; Meliss und Möhrs 2017, 2018, 2019; Moon 1998; Siepmann 2015; Trap-Jensen 2004). Es existieren zurzeit kaum korpusbasierte lexikografische Projekte zur Lexik der gesprochenen Sprache. Lediglich zum Dänischen wurde ein kleines Projekt zu Interjektionen (vgl. Hansen und Hansen 2012) durchgeführt. Die Frage, die sich aus dieser Beobachtung ergab, ist, wie die identifizierten lexikalischen Besonderheiten in einer lexikografischen Ressource auf adäquate Weise beschrieben werden können.

Sowohl die Ergebnisse zweier im Projekt durchgeführter Umfragen zu den Erwartungen und Anforderungen an eine lexikografische Ressource für Spezifika des gesprochenen Deutsch (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019), als auch entsprechende lexikografische Untersuchungen bestätigen, dass die lexikografische Kodifizierung der Merkmale der gesprochenen Sprache in der Interaktion in den aktuellen Wörterbüchern bisher nicht zufriedenstellend berücksichtigt wurde (Meliss 2016: 195; Eichinger 2017: 283). Trotz einiger jüngster Fortschritte in der Erstellung von gesprochensprachlichen Korpora zu verschiedenen Sprachen und Sprachvarietäten (vgl. Barcala Rodríguez et al. 2018; Fandrych, Meißner und Wallner 2017; Schmidt 2014a, 2014b, 2017, 2018; Verdonik und Sepesy Maučec 2017) sind Erfahrungen mit ihrer Nutzung in der Lexikografie bis jetzt kaum bekannt. Daher konnte sich der LeGeDe-Prototyp kaum auf bestehende Modelle stützen, die u.a. bei der Erstellung einer geeigneten Stichwortliste, der Entwicklung von neuen Angabeklassen und deren Darstellung als Orientierungshilfe hätten dienen können.

(iii) Der Informationsbedarf zu typisch gesprochensprachlicher Lexik ist in den letzten Jahren besonders in unterschiedlichen Anwendungsbereichen, wie z.B. in Unterricht und Lehre (speziell im Sekundarbereich und in den Bereichen Deutsch als Fremd- und/oder Zweitsprache) sowie im Verlagswesen in Ver-

bindung mit der Erstellung von geeigneten Unterrichtsmaterialien gestiegen (vgl. Albert und Diao-Klaeger 2018; Handwerker, Bäuerle und Sieberg 2016; Imo und Moraldo 2015; Moraldo und Missaglia 2013; Reeg, Gallo und Moraldo 2012; Sieberg 2013). So wird z.B. im „Gemeinsamen europäischen Referenzrahmen für Sprachen“ (=GeR) beim Beurteilungsraster zur mündlichen Kommunikation und beim Parameter „Interaktion“ für Niveau C1 explizit darauf hingewiesen, dass der Lernende „[...] aus einem ohne weiteres verfügbaren Repertoire von Diskursmitteln eine geeignete Wendung auswählen [kann], um seine/ihre Äußerung angemessen einzuleiten, wenn er/sie das Wort ergreifen oder behalten will, oder um die eigenen Beiträge geschickt mit denen anderer Personen zu verbinden“ (Trim et al. 2001: 37). Sowohl Meliss und Möhrs (2018) als auch Fandrych, Meißner und Wallner (2018) weisen darauf hin, dass Korpora der gesprochenen Sprache von großem Interesse für viele Forschungs- und Anwendungsszenarien sind und gewinnbringend für die Sprachdidaktik eingesetzt werden können, wenn den Lehrenden und Lernenden ein benutzergerechter Zugang zu dem Material angeboten wird. Die sich ergebenden Fragen standen diesbezüglich v.a. in Zusammenhang mit dem Komplexitätsgrad der Informationsdarstellung und dem Nutzen des virtuellen, multimedialen Raums für die Mündlichkeitsdidaktik.

(iv) Außerdem zeigen die Ergebnisse der im LeGeDe-Projekt durchgeführten empirischen Studien zu den Erwartungen künftiger Nutzer/-innen an eine Ressource zum gesprochenen Deutsch in der Interaktion (vgl. Abschnitt 2.5), dass sowohl bei den L1- als auch bei den L2-Sprechenden des Deutschen zu jeweils über 70% Bedarf an einem Wörterbuch zu Spezifika des gesprochenen Deutsch vorhanden ist (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019). Diese Beobachtung bestätigt die grundsätzliche Annahme zum ansteigenden Bedarf sowohl in der Forschung als auch in der Lehre. Die sich daraus ergebenden Fragen betreffen Bereiche der Vermittlung von gesprochener Sprache im Unterricht und die Bereitstellung bzw. Ausarbeitung von geeignetem Material, auf das für Lehrwerke und Unterrichtsstoff zurückgegriffen werden kann (vgl. Meliss und Möhrs 2018).

2.3 Projektziele

Im LeGeDe-Projekt wurden auf Basis der in Abschnitt 2.2 beschriebenen Forschungsfragen und theoretischen Annahmen die folgenden zentralen theoretischen, methodologischen und anwendungsorientierten Ziele verfolgt, die sich bei der Beschäftigung mit dem Thema in der Projektarbeit ergaben:

- Entwicklung von quantitativen und qualitativen Verfahren zur Identifizierung von interaktionstypischen, gesprochensprachlichen lexikalischen Elementen und deren spezifischen Merkmalen im Vergleich zur Lexik der geschriebenen Sprache (vgl. Meliss und Möhrs 2017, 2018, 2019),

- Erstellung einer Stichwortkandidatenliste und Auswahl geeigneter Stichwörter für die Ressource (vgl. Meliss et al. 2018; Möhrs i. Dr.),
- Entwicklung von korpusbasierten quantitativen und qualitativen Verfahren für die lexikalische und interaktional ausgerichtete Datenanalyse und Beschreibung auf verschiedenen sprachlichen Ebenen (vgl. Meliss et al. 2018, 2019; Möhrs und Torres Cajo 2020),
- Bestimmung der Besonderheiten von mündlichem Sprachgebrauch auf verschiedenen Ebenen (Form, Inhalt/Funktion, Situation etc.) im lexikalischen Bereich (vgl. Meliss 2020b; Meliss und Möhrs 2017, 2018, 2019),
- Entwicklung neuartiger lexikografischer Angabeklassen, die u.a. auf die Funktion lexikalischer Einheiten in Interaktionskontexten Bezug nehmen (vgl. Meliss et al. 2019; Meliss i. Dr. [b]),
- Entwicklung innovativer lexikografischer Beschreibungsformate in multi-medialer Form für hochgradig kontextualisierte lexikalische Daten (vgl. Meliss et al. 2019),
- Entwicklung weiterer korpuslinguistischer Methoden und Tools zur Abfrage, Analyse und Strukturierung von automatisch generierten korpusbasierten Daten (vgl. Möhrs, Meliss und Batinić 2017, Lemmenmeier-Batinić 2020). Insbesondere verweisen wir hier auf das Korpusanalysetool „Lexical Explorer“, welches quantitative Analysen ermöglicht und in Abschnitt 2.4 genauer vorgestellt wird.

2.4 Datengrundlage

Die qualitativen Untersuchungen zum Forschungsgegenstand des LeGeDe-Projektes, der gesprochenen Sprache des Deutschen in der Interaktion, erfolgen ausschließlich auf Basis von FOLK (vgl. Schmidt 2014a, Kupietz und Schmidt 2015), das als erstes großes Gesprächskorpus des Deutschen seit 2008 am IDS aufgebaut und ständig erweitert wird. Parallel dazu wird die DGD (Schmidt 2014c, 2018), über die FOLK recherchierbar ist, permanent weiterentwickelt und mit innovativen korpus technologischen Funktionalitäten ausgestattet. Die Gesprächsaufnahmen und Transkripte (teils auch Videoaufnahmen) stammen aus unterschiedlichen privaten, institutionellen und öffentlichen Kontexten aus dem deutschsprachigen Raum. Die Daten zeichnen sich durch die Merkmale medial mündlich, authentisch, spontan (nicht eliziert), größtenteils standardnah und aktuell aus. FOLK ist in seiner Art das größte Korpus für das gesprochene Deutsch in der Interaktion. Zurzeit liegt FOLK in der DGD-Version 2.14 mit rund 285h/2,7 Mio. Tokens und 332 unterschiedlichen Sprechereignissen vor (Stand: 27.04.2020). Die DGD bietet als Korpusanalysetool vielfältige Möglichkeiten der Erschließung mündlicher Daten nach linguistischen und interaktionalen Merkmalen. Über die Anwendungsoberfläche der DGD können strukturierte Tokensuchen realisiert und über vier Annotationsebenen (cGAT-

Transkript, Normalisierung, Lemmatisierung, PoS) recherchiert werden (vgl. Westpfahl 2014; Westpfahl und Schmidt 2016). Außerdem können zu den Gesprächen sowohl Metadaten zu den Sprechenden als auch zu dem Gesprächsereignis abgerufen werden. Ausgehend von der Ausrichtung von FOLK, authentisches, standardnahes Deutsch in der Interaktion zu erfassen, hat sich die Analyse im LeGeDe-Projekt auf diese Sprachvariante fokussiert: „In besonderem Maße sollen jene Phänomene, die als ‚standardnah‘ charakterisiert werden können, behandelt werden. Regionale, soziale, funktionale oder idiolektale Sprachvarietäten werden daher ausgeklammert.“ (Meliss und Möhrs 2017: 44). Über die gesamte Projektlaufzeit wurde die Herausforderung, eine solche Abgrenzung für die konkrete Projektarbeit und nicht zuletzt auch für die Entwicklung der Ressource zu operationalisieren, im Austausch mit Fachkollegen/-innen thematisiert.

Für die im Rahmen der Projektarbeit durchgeführten Analysen wurde einerseits auf die Recherchemöglichkeiten in der DGD zurückgegriffen. Andererseits wurde im LeGeDe-Projekt auch an weitergehenden quantitativen Suchmöglichkeiten gearbeitet, um beispielsweise auch Kollokations- und Kookkurrenzanalysen über gesprochensprachliche Daten durchzuführen. Diese Verfahren sind in der lexikografischen Praxis, in der mit geschriebensprachlichen Daten und Recherchertools gearbeitet wird, inzwischen etabliert (vgl. Geyken und Lemnitzer 2016). Für die FOLK-Daten konnte über die DGD eine Analyse dieser Art bislang nicht durchgeführt werden. Um dem Projektziel „Entwicklung von quantitativen und qualitativen Verfahren zur Identifizierung von interaktionstypischen, gesprochensprachlichen lexikalischen Elementen und deren spezifischen Merkmalen im Vergleich zur Lexik der geschriebenen Sprache“ (vgl. Abschnitt 2.3) näher zu kommen, wurden in der ersten Projektphase Möglichkeiten quantitativer Vergleiche mit DEREKO (dem Deutschen Referenzkorpus; vgl. Kupietz und Keibel 2009; Kupietz et al. 2018) eruiert und ausgearbeitet. Diese Überlegungen und Datenanalysen sind neben weiteren quantitativen Messungen in das Tool „Lexical Explorer“ geflossen, das über die Plattform OWID^{plus} abrufbar ist (vgl. Lemmenmeier-Batinić 2020). Die Anwendung ermöglicht, die quantitativen Daten von zwei Korpora aus der DGD zu durchsuchen und abzufragen: FOLK und GeWiss (Gesprochene Wissenschaftssprache, vgl. Fandrych, Meißner und Wallner 2017). Die hinter dem Tool entwickelten Recherchemethoden wurden für anfängliche Lemmastudien im LeGeDe-Projekt entwickelt und für den Stichwortkandidatenansatz ausgearbeitet. Mit dem „Lexical Explorer“ können quantitative Korpusdaten mit Hilfe von Häufigkeitstabellen bezüglich der Wortverteilung über Wortformen, Kookkurrenzen und Metadaten erforscht werden, was über die Einbindung des Tools in OWID^{plus} auch der interessierten Öffentlichkeit möglich ist. Im LeGeDe-Projekt wurden die Analysen über FOLK maßgeblich für die Definition potenzieller Stichwortkandidaten (vgl. dazu auch Abschnitt 3.2) und die Detektion typischer Wortverbindungen zu einem spezifischen Analysestichwort genutzt. Auch die Daten zu quantitativen Recherchen zu einem Stichwort

und seinem Auftreten in Interaktionstypen (definiert über die Kategorien „private“, „public“, „non-private/non-public“ und „other“) wurden, neben anderen Metadaten, als Analysewerte über den „Lexical Explorer“ herangezogen und in die lexikografische Beschreibung eingebunden. Über ein Verlinkungsangebot können quantitative Angaben aus den lexikografischen Artikeln im LeGeDe-Prototyp direkt nachvollzogen werden (vgl. hierzu auch Abschnitt 3.4).

2.5 Empirische Forschung: Erwartungshaltungen

2.5.1 Hintergrund zu den Studien

In Ergänzung zu den konzeptionellen Arbeiten im LeGeDe-Projekt wurden auch verschiedene Arten empirischer Studien in Verbindung mit den Erwartungshaltungen potenzieller Nutzer/-innen an eine neuartige lexikografische Ressource zu Besonderheiten des Gesprochenen durchgeführt. Dieser Arbeitsbereich wurde in Kooperation mit dem Projekt „Empirische Methoden“⁴ in der Abteilung Lexik des IDS entwickelt und durchgeführt. Die Erwartungen potentieller Nutzer/-innen zu einem recht frühen Zeitpunkt der Projektarbeit abzufragen, sollte ermöglichen, die Konzeption der Ressource speziell daraufhin auszurichten. Zwischen Februar und Mai 2017 wurden dementsprechend zwei Studien durchgeführt, die bezüglich der befragten Zielgruppe und des Formats unterschiedlich gestaltet waren.⁵ Zum einen äußerten sich Expert/-innen in einem Interview gezielt, in geschlossenen und offenen Fragenformaten zu ihren Erwartungen. Und zum anderen wurde eine Online-Befragung konzipiert, die sich an eine breitere Probandengruppe gerichtet hat. Eine der größten Herausforderungen beider empirischer Studien war die Tatsache, dass zu dem Zeitpunkt der Befragungen eine zukünftige Ressource thematisiert wurde, die zum Befragungszeitpunkt noch gar nicht existierte und für die es auch keine Vorbilder gab. Daraus ergab sich die einmalige Chance, geäußerte Erwartungen bei der Erstellung der Ressource zu berücksichtigen.

2.5.2 Ergebnisse der Erwartungsstudien

Zu beiden Studien sind in Meliss, Möhrs und Ribeiro Silveira (2018) die Ergebnisse in umfassender Art dargestellt. Der Abgleich der Antworten bzw. Einschätzungen mit den soziodemografischen Daten ließ bei der Interpretation der Daten auch Rückschlüsse zu bestimmten Probandengruppen zu. So wurde in Meliss, Möhrs und Ribeiro Silveira (2019) besonders die Befragtengruppe der L2-Lerner/-innen in den Blick genommen. Die spezifischen anwendungsorientierten Bedürfnisse in den Bereichen DaF und DaZ in Verbindung mit dem Erwerb und der Anwendung gesprochensprachlicher Lexik in der Interak-

tion, auf die schon in Meliss und Möhrs (2018) hingewiesen wurde, konnten ebenfalls mit den Ergebnissen der Befragungen abgeglichen werden.

Es zeigte sich, dass sowohl bei L1- als auch L2-Sprecher/-innen Bedarf an einem Nachschlagewerk zu gesprochener Sprache in der Interaktion existiert, das Spezifika des gesprochenen Deutsch in Bedeutung, Verwendung, Kombinatorik, Angemessenheit und weiteren Bereichen im Vergleich zum geschriebenen Deutsch erfasst. Dieser Bedarf kann nicht oder nur teilweise durch gängige einsprachige Wörterbücher wie dem *Duden-online* oder auch dem *DWDS* befriedigt werden. Auch zweisprachige Werke wie *Leo*, *Linguee* oder *PONS-online* decken in ihrem lexikografischen Angebot Fragen zu Besonderheiten der Lexik im gesprochenen Deutsch nicht oder nur ungenügend ab. Eine Ressource, in der über den Einbezug authentischer Beispiele zusammen mit Audiomaterial und Transkripten Informationen bereitgestellt werden, stellt laut Meinung der Befragten eine nützliche Quelle für unterschiedliche Rechercheziele dar. Im Laufe der Projektarbeit wurden daher die Erwartungen an „besondere“/„gesprochensprachlich spezifische“ Stichwörter umfassend in den Blick genommen und an deren Bedeutungs- und Funktionsbeschreibung für die lexikografische Umsetzung gearbeitet. Geäußerte Erwartungen nach Verknüpfungen zu unterschiedlichen Ressourcen, wie u.a. zu dem zugrundeliegenden Korpus, wurden ebenfalls umgesetzt (vgl. Abschnitt 3.4).

Nicht alle Erwartungen und Wünsche der Befragten und auch nicht alle im Projekt entwickelten Ideen konnten im Rahmen der Projektlaufzeit im LeGeDe-Prototyp realisiert werden. Das Spektrum von potenziellen Möglichkeiten, das in den unterschiedlichen Frageformaten zum Ausdruck kam, zeigt die Vielfältigkeit der Anknüpfungspunkte für eine Weiterarbeit im Bereich der korpusbasierten Lexikografie gesprochen Sprachlicher Besonderheiten auf (vgl. Abschnitt 4).

3. Der LeGeDe-Prototyp

Im folgenden Abschnitt werden hauptsächlich Informationen zu der lexikografischen Struktur des LeGeDe-Prototyps angeboten. Bezüglich der Makrostruktur wird besonders der Stichwortansatz thematisiert (Abschnitt 3.1). In Verbindung mit der Mikrostruktur (Abschnitt 3.2) wird ausführlich die Methode der Datenanalyse und -strukturierung vorgestellt, um daran anschließend die komplexe lexikografische Angabenstruktur auf den drei Informationsebenen (i) Überblicksartikel, (ii) Modul 1 und (iii) Modul 2 darzustellen. Zur Illustration dienen ausgewählte Einträge aus dem LeGeDe-Prototyp. Außerdem werden die Zugriffsmöglichkeiten und Vernetzungsstrukturen (Abschnitt 3.3) und das Angebot an unterschiedlichen lexikografischen Umtexten (Abschnitt 3.4) vorgestellt.

3.1 Makrostruktur: Stichwortansatz

Eine der zentralen Forschungs- und Methodikfragen, mit denen sich das LeGeDe-Projekt beschäftigt hat, ist die korpusbasierte Identifizierung typischer gesprochensprachlicher Lexik. In einem direkten Bezug zu den distinktiven Merkmalen der Lexik der geschriebenen vs. der gesprochenen Sprache in der Interaktion steht die Erstellung einer Stichwortliste (vgl. Klosa 2013a, 2013b; Klosa und Tiberius 2016; Klosa, Schnörch und Schoolaert 2010; Schnörch 2005; Stadler 2014; Wiegand 1983) mit Kandidaten für den LeGeDe-Prototyp, die möglichst typische Phänomene des Gesprochenen in der natürlichen, spontanen Interaktion aufweisen und sich quantitativ deutlich von der geschriebenen Sprache abgrenzen. Für die Entwicklung des LeGeDe-Prototyps wurden so neben Einwortlemmata, die im Vergleich zur Schriftsprache in der gesprochenen Sprache in der Interaktion eine spezifische Bedeutung oder Verwendung bzw. Funktion einnehmen (z.B. Interjektionen, Diskurspartikeln), auch Mehrwortausdrücke und Konstruktionen mit spezifischen Funktionen in der Interaktion (z.B. *was weiß ich*, *ich weiß nicht*, *keine Ahnung*, *guck mal*) als mehrteilige Stichwortkandidaten integriert (vgl. Bergmann 2017, Günthner 2018, Helmer und Deppermann 2017, Helmer, Deppermann und Reineke 2017, Imo 2007, Möhrs 2020, Torres Cajo 2019, Zeschel 2017).

Für die Erstellung einer Stichwortkandidatenliste wurde daher im LeGeDe-Projekt eine korpusbasierte und interpretative Methode entwickelt (vgl. Meliss et al. 2018), mit der die wichtigsten Kandidaten der typischen gesprochenen Lexik in der Interaktion aufgedeckt werden konnten. Der Vergleich zur Schriftsprache des Deutschen erfolgte über eine Teilmenge aus DEREKO. Das frequenzgesteuerte korpusbasierte Verfahren, bei dem die Häufigkeitsklassen der Lemmata aus FOLK und aus DEREKO ermittelt und diese dann in einem direkten Vergleich betrachtet werden können, ist über den „Lexical Explorer“ nachvollziehbar. Die Stichwortkandidatenliste (Stand: 23.05.2018), die den Ausgangspunkt für die Auswahl der ausführlich bearbeiteten Lemmata im LeGeDe-Prototyp darstellt, kann mit den 322 Kandidaten vollumfänglich über die Umtexte der Online-Ressource abgerufen werden (Anhang im Download-Dokument unter „Über LeGeDe“).⁶

Die Rangliste der 25 in der Abfrage zuerst angezeigten möglichen Stichwortkandidaten (vgl. Abb. 1) vermittelt einen ersten Eindruck von denjenigen Kandidaten, die für den LeGeDe-Prototyp relevant waren. Die hohe Häufigkeitsklassendifferenz bei den Interjektionen (*ah*, *ach*, *oh*), Modalpartikeln (*ja*, *halt*), Gesprächspartikeln (*okay*, *ja*, *na*) und Verben (*gucken*, *kriegen*) weist mit der beschriebenen quantitativen Perspektive auf Besonderheiten im Gesprochenen vs. Geschriebenen hin, die in weiteren Schritten durch qualitative Studien im LeGeDe-Projekt genauer analysiert wurden.

■ **FOLK vs. DeReKo**

Column visibility CSV Show 25 entries

Lemma	FOLK HK	DeReKo HK	HK Diff	Filter	PoS
okay	4	14	10	1	NG
ah	4	14	10	1	NG
ach	4	13	9	1	NG
ja	0	8	8	1	PTK/NG
oh	5	13	8	1	NG
gucken	5	13	8	1	V
halt	4	12	8	1	PTK/NG
du	2	9	7	1	P
nachher	7	14	7	1	ADV
danke	7	14	7	1	NG
irgendetwas	7	14	7	1	P
na	5	12	7	1	NG
irgendwie	5	12	7	1	ADV
kriegen	5	12	7	1	V
nein	3	10	7	1	NG
mal	2	8	6	1	ADV/PTK
eh	7	13	6	1	ADV
Mama	7	13	6	1	N
cool	7	13	6	1	NG/ADJ
drin	6	12	6	1	ADV/PTK
drauf	6	12	6	1	ADV/PTK
dran	6	12	6	1	ADV/PTK
raus	6	12	6	1	PTK/ADV
sozusagen	6	12	6	1	ADV/NG
dein	5	11	6	1	P

Showing 1 to 25 of 322 entries (filtered from 52,966 total entries)

Previous 1 2 3 4 5 ... 13 Next

Abb. 1: Rangliste der 25 in der Abfrage zuerst angezeigten Stichwortkandidaten mit der höchsten Häufigkeitsklassendifferenz zwischen FOLK und DEREKO (Stand 2018)

Die Kombination aus automatisierten Verfahren und manueller Analyse der Korpusdaten hat sich als effektiver Weg erwiesen, um die Stichwortkandidatenauswahl für eine Ressource zu Spezifika der Lexik des gesprochenen Deutsch anzugehen. Mit Hilfe des „Lexical Explorers“ kann die automatische Sortierung und Filterung der Stichwortkandidaten nach Wortarten, Frequenz und anderen Merkmalen sowie die Auswahl weiterer Stichwörter in weiteren Schritten des Projekts nachvollzogen werden.

3.2 Mikrostruktur: Methodik und Informationsangebot

3.2.1 Datenanalyse und Datenstrukturierung

Verschiedene quantitative und qualitative Verfahren zur lexikalischen Bedeu-

tungsdisambiguierung und zur Entwicklung von Wortprofilen, die als methodologische Ansätze in der korpusbasierten Lexikologie und Lexikografie gelten (vgl. u.a. Engelberg 2015, 2018, 2019; Engelberg et al. 2011), wurden mit interaktionslinguistischen Analysen sprachlicher Einheiten und der Beschreibung ihrer Formen und Funktionen in der Interaktion (vgl. u.a. Deppermann 2007; Couper-Kuhlen und Selting 2018) vereint. Unterschiedliche quantitative Informationen, wie u.a. automatisch generierte Daten (Frequenzdaten zum Formenbestand, zu Kollokationen, Kookkurrenzen etc.), wurden nicht nur bei der ersten korpusbasierten Annäherung an die Daten durch entsprechende Hypothesenbildung, sondern auch bei der Interpretation der Analyseergebnisse unterstützend genutzt. Auf der Grundlage von qualitativen lexikalisch-semantischen und interaktionslinguistischen Einzelbeleganalysen wurden für die unterschiedlichen grammatischen Kategorien Kodierschemata entwickelt.

Die Datengrundlage für die Analyse eines Lemmas ist jeweils grundsätzlich eine Zufallsstichprobe aus FOLK, aus der die ersten 100 gültigen Belege manuell kodiert wurden. Der Prozess des Kodierens wurde durch qualitative Analysen gestützt. Neben Metadaten zum Treffer und zum Transkript, die automatisch extrahiert vorliegen, wurden v.a. formale (Person, Numerus, Modus etc.), inhaltliche (Bedeutung, Bedeutungsumschreibung etc.), kombinatorische (Strukturmuster, Kookkurrenzen etc.), morphosyntaktische (Realisierungsmöglichkeiten der Aktanten etc.), funktionale, sequenzbezogene sowie prosodische Aspekte betrachtet und festgehalten. Da sich schon in der Stichprobe aus 100 analysierten Treffern die Notwendigkeit ergab, einerseits die Ebene der Bedeutung und andererseits die funktionale Ebene der ausgewählten Stichwörter mit teilweise unterschiedlichen Analyseparametern und Methoden zu untersuchen, zu strukturieren und lexikografisch darzustellen (Modul 1 vs. Modul 2), wurden für die weiteren Untersuchungen, Strukturierungen und Darstellungen der Ergebnisse zwei unterschiedliche Analyse- und Beschreibungsverfahren angesetzt (vgl. Meliss i. Dr. [b]).

Im Zentrum von Modul 1 steht die semantisch und syntaktisch motivierte Disambiguierung der einzelnen Lesarten, während im Zentrum von Modul 2 die Beschreibung der interaktionalen Funktionen steht. Entsprechend wurden die weiteren Analyse-, Strukturierungs- und Beschreibungsverfahren den Anforderungen des jeweiligen Moduls angepasst.

(a) Modul 1

Im Hinblick auf das Informationsangebot für Modul 1 wurden die Kodierergebnisse der Lemma-Stichprobe in ihren jeweiligen Gesprächssequenzen nach Form, Bedeutung und Kombinatorik (Strukturmuster, Kollokationen, feste Wendungen, interaktionale Einheiten etc.) sowie unter Berücksichtigung von Kontext und Metadaten analysiert (vgl. Abb. 2) und auf der Grundlage einer semantisch-syntaktischen Bedeutungsdisambiguierung einzelne Lesarten identifiziert. Als Orientierungshilfe für die jeweilige Lesartendisambiguierung

dienten dafür entsprechende lexikografische Informationen ausgewählter Wörterbücher (z.B. *E-Valbu*, *LGWB DaF*, *Duden-online*, *DWDS*). Durch den Vergleich der Daten aus der LeGeDe-Stichprobe mit denen aus gängigen lexikografischen Werken, die hauptsächlich auf der geschriebenen Sprache basieren, wurden u.a. folgende Ziele verfolgt:

(i) Identifizierung von typisch gesprochen sprachlichen Lesarten: Mögliche, besonders interessante Beobachtungen beziehen sich auf solche Lesarten, die nur in den FOLK-Daten belegt werden konnten, aber nicht in den konsultierten Wörterbüchern kodifiziert sind (so konnten z.B. für *gucken* in FOLK wesentlich mehr Lesarten identifiziert werden als in den konsultierten Wörterbüchern). Diese Fälle erlauben den Schluss, dass es sich zumindest um Lesarten handelt, die vor allem in der gesprochenen Interaktion auf der Grundlage der LeGeDe-Stichprobe aufgedeckt werden konnten.

(ii) Identifizierung von eher nicht typisch gesprochen sprachlichen Lesarten: Die Identifizierung von solchen Lesarten, die nicht in den FOLK-Daten belegt werden konnten, aber in den konsultierten Wörterbüchern kodifiziert sind, erlaubt den Schluss, dass es sich zumindest um Lesarten handelt, die in der gesprochenen Interaktion auf der Grundlage der LeGeDe-Stichprobe eher selten sind. Da sie nicht in der LeGeDe-Stichprobe belegt werden konnten, wurden sie für den LeGeDe-Prototyp ausgeschlossen. Allerdings bieten wir in den Überblicksartikeln (vgl. Abschnitt 3.2.2) Links zu Wörterbuchquellen an, sodass ein/-e Nutzer/-in zu weiteren Lesarten, die in anderen Wörterbüchern verzeichnet sind, hier nachschlagen kann (vgl. auch Abschnitt 3.3).

(iii) Information zu formalen, morphosyntaktischen Eigenschaften bzw. Distributionsbeschränkungen (z.B. Präferenz für Modus Imperativ bei *gucken* in der Bedeutung: ‚richtungsorientiert visuell wahrnehmen‘).

(iv) Identifizierung von typisch gesprochen sprachlichen, kombinatorischen Eigenschaften: Mögliche Fälle sind solche mit unterschiedlichen kombinatorischen Eigenschaften, die explizit in den FOLK-Daten belegt werden konnten, aber in den konsultierten Wörterbüchern nicht oder nur unzureichend explizit kodifiziert sind. Diese Fälle umfassen verschiedene Bereiche:

- Informationen zu den Strukturmustern und nicht realisierten Aktanten (z.B. für *kriegen* in der Bedeutung ‚bekommen‘ konnte die Rolle „Sender“ nicht belegt werden),
- Informationen zu den Strukturmustern und der Verwendung von zusätzlichen deiktischen Elementen (vgl. *gucken* und *schauen* in der Bedeutung ‚richtungsorientiert visuell wahrnehmen‘),
- Informationen zu typischen Verbindungen/Kollokationen (So zeigt z.B. *wissen* in der Bedeutung ‚informiert sein‘ eine sehr häufige Verbindung mit den Partikeln *ja* und *aber* auf. Das Adjektiv *gut* in der Bedeutung von ‚positiv bewertet‘ tritt häufig mit *echt* und das Adverb *eben* mit *gerade* auf.),

- Informationen zu festen Wendungen (z.B. *keine Ahnung haben, gucken wo jemand bleibt, frei kriegen, Besuch kriegen, die Kurve kriegen*) oder den Routineformeln (z.B. *guten Appetit, danke schön*).

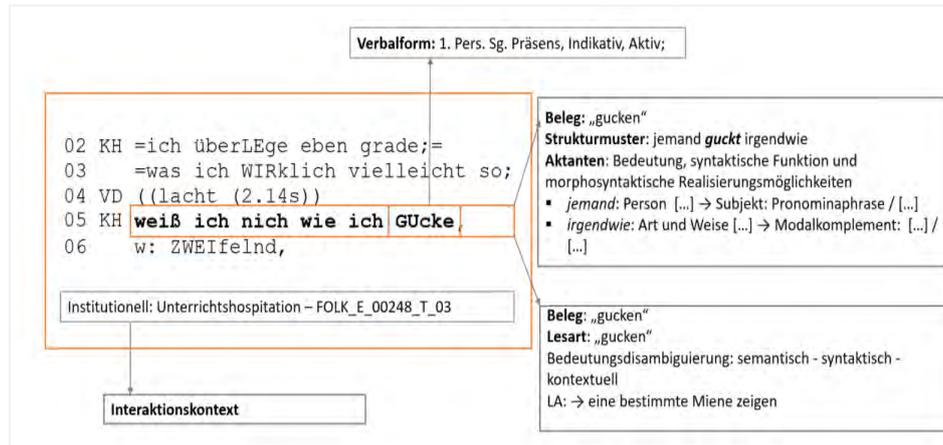


Abb. 2: Illustration ausgewählter Analyseparameter für die Kodierung (Modul 1)

(b) Modul 2

Die Analysen für Modul 2 basieren auf der Lemma-Stichprobe von 100 gültigen Belegen, die Hinweise auf relevante Phänomene liefert (vgl. Abb. 3). Die Belege wurden daraufhin geprüft, ob und welche Kandidaten sich für eine interaktional ausgerichtete Beschreibung, so wie sie in Modul 2 erfolgt, eignen, d.h. Einheiten mit einer spezifischen kommunikativen Funktion im Gespräch wie z.B. die verschiedenen Muster zum Verb *wissen*: *ich weiß nicht, weißt du, ich weiß, wer weiß, was weiß ich*. In den Fällen, in denen die Stichprobe zwar Hinweise auf verfestigte Muster gibt, diese jedoch in nicht ausreichender Anzahl vorliegen, um belastbare Aussagen treffen zu können, wurden separate, gezielte Abfragen in FOLK generiert (z.B. *was weiß ich*: 2 Belege in der Stichprobe; eine Suche in der DGD ergibt insgesamt weit mehr Belege). Für die Analysen für Modul 2 wurde das grundlegende Kodierschema für die qualitativen Einzelfallanalysen angepasst und um weitere, für das jeweilige Phänomen relevante formale und funktionale Parameter ergänzt. Anschließend wurden die ermittelten typischen, rekurrenten und verfestigten Form-Funktions-Zusammenhänge unter Berücksichtigung der Sequenz und Interaktionskontexte beschrieben. Dabei wurde versucht, das Informationsangebot sprachlich, inhaltlich und gestalterisch an die Erfordernisse einer lexikografischen Umgebung anzupassen.

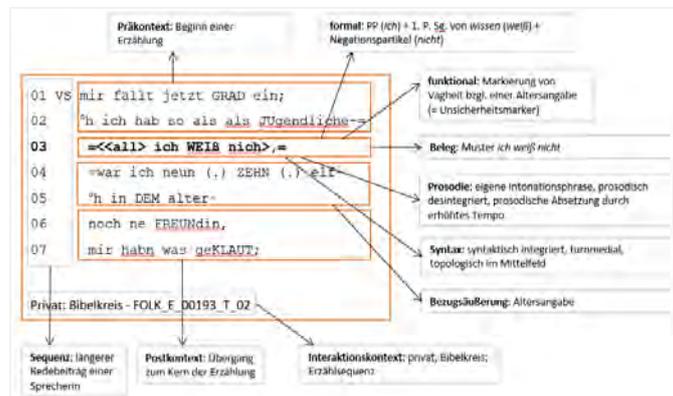


Abb. 3: Illustration ausgewählter Parameter für die Kodierung (Modul 2)

Da sich die Analysen beider Module zunächst auf die jeweilige Lemma-Stichprobe (100 gültige Belege) beziehen, wurden auch nur solche Phänomene (Bedeutungen, kommunikative Muster, Funktionen) beschrieben, die hier belegt werden konnten. Solche Phänomene, die ggf. Gegenstand von Beschreibungen in anderen wissenschaftlichen Einzelstudien und/oder lexikografischen Werken sind, aber nicht in der Stichprobe dokumentiert werden konnten, sind daher nicht Teil des LeGeDe-Informationsangebots.

3.2.2 Überblicksartikel

schauen

Das Vollverb **schauen** konnte in der LeGeDe-Stichprobe in acht verschiedenen Lesarten und in zwei verschiedenen interaktionalen Mustern belegt werden. Die Grundbedeutung von **schauen** steht in Zusammenhang mit der visuellen richtungsorientierten Wahrnehmung. Diese konnte auch in anderen Bedeutungsvarianten von **schauen** in Zusammenhang mit den Lesarten 'überprüfen', 'ansehen' und 'suchen' nachgewiesen werden. Die visuelle Wahrnehmungsbedeutung verliert deutlich in den Bedeutungsvarianten, in denen **schauen** in den Lesarten 'abwarten', 'abwägen', 'sich bemühen' und 'achten auf' auftritt. In diesen Fällen dominiert teilweise eine kognitiv-mentale Bedeutungskomponente. Neben dem Muster **schau mal** als Diskursmarker tritt das Muster **mal schauen** zur Verschiebung einer Handlung in die Zukunft teilweise als Beendigungssignal auf.

Frequenz
FOLK 2.11 (absolut): 669

FOLK-Zugriff
schauen (in FOLK)
schauen (im Lexikal Explorer)

Bedeutungen	Funktionen im Gespräch
<p>schauen [VOLLVERB]</p> <ul style="list-style-type: none"> ▷ richtungsorientiert visuell wahrnehmen ▷ überprüfen ▷ ansehen ▷ abwarten ▷ abwägen ▷ sich bemühen ▷ suchen ▷ achten auf 	<p>schau mal [VERBALSYNTAGMA]</p> <ul style="list-style-type: none"> ▷ Diskursmarker <p>mal schauen [VERBALSYNTAGMA]</p> <ul style="list-style-type: none"> ▷ Verschiebung einer Handlung in die Zukunft

<p>Information in anderen lexikografischen Ressourcen</p> <ul style="list-style-type: none"> • Online-Wortschatz-Informationssystem Deutsch (GWDI) • Elektronisches Wörterbuch zur Verbalenz (eWlbu) • Digitales Wörterbuch der Deutschen Sprache (DWDS) 	<p>Forschungsliteratur</p> <ul style="list-style-type: none"> • Günthner (2017) • Imo (2007) • Wegner (2015)
--	--

Abb. 4: Überblicksartikel zum Lemma *schauen*

Für jedes Stichwort stehen allgemeine Übersichtsinformationen zur Verfügung und bieten in beschreibender Form bedeutungs- und funktionsorientierte Informationen (z.B. *schauen*, vgl. Abb. 4). Eine klare modulare Aufteilung der Informationen ermöglicht einerseits die Darstellung lexikalisch-semantischer Informationen, die sich an der jeweiligen Bedeutung der entsprechenden Lesarten eines Lemmas (=Modul 1) orientieren, und andererseits das Angebot funktionspezifisch interaktiv orientierter Informationen (=Modul 2). Für beide Module wurden spezifische lexikografische Angabeklassen verwendet oder neu entwickelt, die neben herkömmlichen Wörterbuchinformationen völlig neue Erkenntnisse und Formate bieten. Unterschiedliche Querverbindungen zwischen den beiden Modulen stehen durch eine interne Verknüpfung explizit zur Verfügung (vgl. Abschnitt 3.4). Ein erweitertes externes Informationsangebot wird durch Links bereitgestellt, die einerseits zu weiteren lexikografischen Ressourcen und andererseits zu FOLK und dem „Lexical Explorer“ führen (vgl. Abschnitte 2.4, 3.4). Zusätzlich wird die berechnete korpusbasierte Häufigkeitsklassendifferenz zwischen den Stichwörtern in den jeweiligen Korpora (geschrieben: DEREKO, gesprochen: FOLK) visualisiert (vgl. Abb. 5).

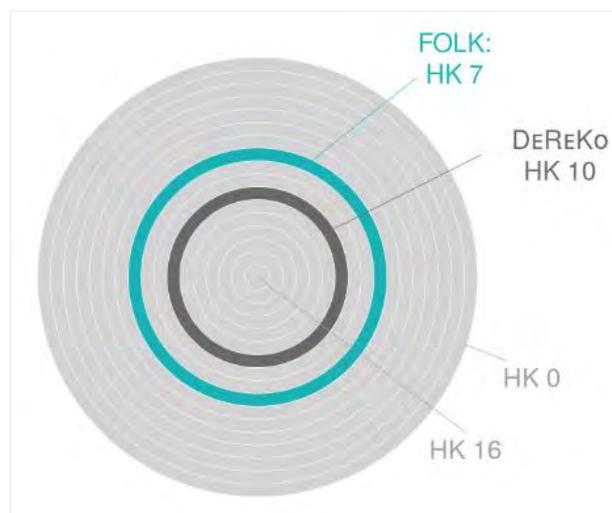


Abb. 5: Frequenzvisualisierung (FOLK/DEREKO) zum Lemma *schauen*

3.2.3 Modul 1

Bewegen sich Nachschlagende aus der alphabetisch angeordneten Stichwortliste (siehe linker Menübalken) oder aus dem Überblicksartikel (vgl. Abschnitt 3.2.2) zum Hauptlemma in Modul 1 (Bsp. *schauen*), so werden im Wortartikelkopf zunächst lesartübergreifende Informationen (zu Form, Varianz und sehr zen-

trale, lemmabezogene Forschungsliteratur) angezeigt. Darunter sind die unterschiedlichen Bedeutungen mit einem kurzen Label (z.B. ‚abwarten‘) und einer kurzen Bedeutungsbeschreibung aufgelistet (vgl. Abb. 6). Über das Pfeilsymbol neben dem Label klappt sich auf Klick die detaillierte, lesartspezifische Beschreibung auf.

The screenshot shows a digital dictionary entry for the verb 'schauen'. At the top, it is labeled 'schauen [VOLLVERB]'. Below this, there are three columns: 'Form' with '---', 'Varianz' with '---', and 'Forschungsliteratur' with a list of references: 'Günthner (2017)', 'Imo (2007)', and 'Wegner (2015)'. The main body of the entry lists four meanings, each with a red square icon and a right-pointing arrow:

- richtungsorientiert visuell wahrnehmen** ▶
Mit der Grundbedeutung von **schauen** wird ausgedrückt, dass jemand etwas durch einen visuellen Eindruck wahrnimmt, indem er oder sie den Blick dabei bewusst auf etwas oder jemanden richtet.
- überprüfen** ▶
Mit **schauen** wird ausgedrückt, dass jemand etwas durch eigene visuelle Wahrnehmung in Verbindung mit einer bewussten Rechercheaktivität überprüft bzw. feststellt oder beobachtet.
- ansehen** ▶
Bei **schauen** in der Bedeutung 'sich etwas ansehen/anschauen' richtet sich die visuelle Wahrnehmung darauf, sich etwas (z.B. einen Film) bewusst in einem entsprechenden Medium anzusehen. Wenn konkret das Medium 'Fernseher' thematisiert wird, dann erhält das Verb die spezifische Bedeutung 'fernsehen'.
- abwarten** ▶
Bei **schauen** zum Ausdruck von einem zukunftsorientierten Abwarten in Bezug auf das ungewisse Eintreten eines Sachverhaltes oder der Durchführung einer zukünftigen Handlung verblasst die visuelle Wahrnehmungsbedeutung deutlich.

Abb. 6: Ausschnitt aus Modul 1 zu *schauen* mit übergreifenden Informationen („Kopf“) und einer Kurzübersicht zu den einzelnen Lesarten

Eine weitere Möglichkeit des Zugriffs direkt auf eine bestimmte Lesart kann über den Überblicksartikel erfolgen. Hier können Nachschlagende auf das Label einer Lesart klicken und gelangen zum ausgeklappten kompletten Wortartikelinhalt.

Der Wortartikel beinhaltet neben Label und einer Bedeutungsbeschreibung die folgenden Bereiche:

- **Belegblock:** Ein illustrierender Belegausschnitt aus FOLK mit Belegtitel, Transkriptausschnitt als Minimaltranskript aus der DGD, Belegkontext und optional einer weiteren Erläuterung zum Belegausschnitt ergänzen die Bedeutungsbeschreibung. Unterhalb des Transkripts steht die FOLK-ID, über die angemeldete DGD-Nutzer/-innen direkt in die DGD zum entsprechenden Transkriptausschnitt gelangen, darin vertieft recherchieren und die Audiodatei anhören können (vgl. Abb. 7).

[1] **schauen** in der Bedeutung 'abwarten'

MI 0103 **AZ** weiß
MI 0104 (0,84)
MI 0105 **DZ** ja weißwein auch
MI 0106 (5,23)
MI 0107 **DZ** packst du die für mich ein
MI 0108 (1,8)
MI 0109 **AZ** ja
MI 0110 (1,04)
MI 0111 **AZ** kannst hier einfach reinstecken
MI 0112 (1,46)
MI 0113 **AZ** **schauen** wie lang das hier dauert
MI 0114 (1,54)
MI 0115 **DZ** die ham
MI 0116 (10,15)
MI 0117 **AZ** nja ja
MI 0118 (1,89)
MI 0119 **AZ** na des is h^o genau is kein gemütliches thema is gut formuliert
MI 0120 (3,96)
MI 0121 ((Ein Mann bezahlt seine Getränke, 7,16 Sek.))
MI 0122 (1,49)
MI 0123 **AZ** hätt gern_n weißwein

Belegkontext
Während einer Theaterpause warten Mutter (AZ) und Tochter (DZ) darauf, ihre Getränkebestellung aufgeben zu können, und unterhalten sich dabei.

Privat: Pausenkommunikation im Theater - FOLK_E_00246_T_01

Abb. 7: Belegblock im Artikel *schauen*, Bedeutung ‚abwarten‘

- **Formbesonderheiten:** Ein optionaler lexikografischer Hinweistext erläutert Besonderheiten auf formaler und kombinatorischer Ebene (vgl. Abb. 8).

Der auffordernde Gebrauch in der 1. Person Plural (**schau(e)n wir mal**) verweist auf eine Art zuständigkeitsorientierte Einladung an alle Interagierenden, abzuwarten, was in Zukunft passieren wird. Gleichzeitig konnte in allen Belegen eine Kombination mit der Modalpartikel **mal** dokumentiert werden, die der Äußerung eine gewisse Unverbindlichkeit verleiht.

Abb. 8: Formbesonderheiten im Artikel *schauen*, Bedeutung ‚abwarten‘

- **Kombinatorik:** Diese Information wird durch die Beschreibung der Strukturmuster, der festen Verbindungen (im weiten Sinne) sowie der Kollokationen und der interaktionalen Einheiten differenziert. Kurze Belegblöcke illustrieren in allen Fällen die aufgeführten Verwendungsweisen.

(i) Die **Strukturmuster** (vgl. Abb. 9a) werden durch eine abstrahierte, formelhafte Information angeboten. Die einzelnen Aktanten, aus denen die Strukturmuster zusammengesetzt sind, werden bezüglich ihrer semantischen Rolle, ihrer syntaktischen Funktion und den morphosyntaktischen Realisierungsmöglichkeiten genauer erfasst, wobei Besonderheiten angegeben werden.

jemand schaut, (was sein wird) [STRUKTURMUSTER]

- **jemand**: Person, die durch ihre kognitiv-mentale Tätigkeit abwartet, ob/wie etwas bzw. was sein wird
SYNTAKTISCHE POSITION: Subjekt
- **was sein wird**: ungewisser zukünftiger Sachverhalt, auf den sich eine Person in abwartender Haltung bezieht
SYNTAKTISCHE POSITION: satzförmiges Komplement: fakultativ
REALISIERUNGSFORMEN: **ob** oder **was/wie**-Satz
ANMERKUNG: Wenn dieses satzförmige Komplement nicht realisiert wird, wird das Ende einer thematischen Sequenz markiert.

mal **schauen**, was dann passiert

MI 1169 SK mal **schauen** was dann pa[ssiert]

Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00021_T_11

mal **schauen**, wie er es macht

MI 1248 BP könnt man ja so linien machen und dann fotos damit verbinden zum beispiel

MI 1249 (0.29)

MI 1250 BP hab ich ihm geschrieben

MI 1251 (0.29)

MI 1252 BP *hh ja s hat er

MI 1253 (0.46)

MI 1254 BP dankend angenommen un (.) mal **schauen** wie er_s macht

Privat: Tischgespräch - FOLK_E_00271_T_01

Abb. 9a: Strukturmuster zu *schauen*, Bedeutung ‚abwarten‘

(ii) Unter dem Oberbegriff **Feste Wendungen/Kollokationen** (vgl. Abb. 9b) werden im LeGeDe-Prototyp verschiedene Arten von mehr oder weniger festen (idiomatischen) lexikalischen Einheiten, ohne eine weitere Spezifizierung bzw. terminologische Präzisierung vorzunehmen, zusammengefasst. Die entsprechenden Kommentare umfassen Aspekte der Bedeutung und des Gebrauchs.

schön wär's [FESTE WENDUNGEN / KOLLOKATIONEN]

Die Verbindung **schön wär's** kommt in der LeGeDe-Stichprobe als alleinstehende Äußerung vor. Sie wird von einem Sprechenden A als Reaktion auf eine davorstehende Äußerung eines Sprechenden B eingesetzt. Der/Die Sprechende A kommentiert die zuvor gebrachte Äußerung des Sprechenden B in einer Weise, dass es aus seiner/ihrer Sicht so ist, dass der Sachverhalt, auf den referiert wird, so nicht zutrifft oder umgesetzt werden kann. Die Konjunktivform von **sein** in der Verbindung **schön wär's** deutet an, dass es der Sprechende A theoretisch als erfreulich einstufen würde, wenn die Äußerung doch zutreffen würde oder umgesetzt werden könnte. Mit der Verbindung **schön wär's** schließt er/sie dies aber im Sinne von 'es wäre positiv, wenn der Fall einträfe, dies scheint aber nicht realistisch' und damit eines hypothetischen Wunsches aus.

schön wär's

MI 0044 KA zum zweiten mal schon das gleiche falsch gedacht (.) vor allm

MI 0045 (0.25)

MI 0046 PA wenn das zusammen nimmt hast dann richtig gedacht oder

MI 0047 KA ((Lachansatz)) schön wär_s ((lacht))

MI 0048 AM minu[s] mal minus macht plus

MI 0049 KA [ʔh ich glaub das w]ird nix

Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00132_T_16

Abb. 9b: Feste Wendungen/Kollokationen zu *schön* (Adj.), Bedeutung ‚erfreulich‘

(iii) Die Auflistung von festen Verwendungsmustern, die als **interaktionale Einheiten** identifiziert werden konnten (vgl. Abb. 9c), ermöglicht die Einrichtung einer internen Verlinkung (siehe mal schauen unterstrichen in Abb. 9c) zu interaktionalen Funktionalitäten, die in Modul 2 ausführlich beschrieben werden.

mal schauen : Verschiebung einer Handlung in die Zukunft [INTERAKTIONALE EINHEIT]

Mit **mal schauen** schieben Sprechende eine Handlung, Entscheidung, Einschätzung, Bewertung, Planung o.Ä. zeitlich auf. Damit dient **mal schauen** häufig als (vorläufiges) Beendigungssignal.

Abb. 9c: Ausschnitt zu interaktionalen Einheiten zu *schauen*, Bedeutung ‚abwarten‘

— **weitere Auffälligkeiten:** In einem zweiten optionalen Textfeld werden auffällige Analyseergebnisse zu Metadaten und Frequenzen kommentiert (vgl. Abb. 10).



Über 90% der Belege der LeGeDe-Stichprobe konnten dieser Lesart zugeordnet werden. In dieser Bedeutungsvariante wird **wissen** nur selten in öffentlichen Interaktionssituationen belegt. Außerdem konnte **wissen** als Teil eines interaktionalen Musters in einem Viertel der Belege in Verbindung mit dieser Lesart dokumentiert werden. In diesen Fällen werden die Argumente des Strukturmusters nicht immer alle realisiert und es treten zusätzliche interaktionale Funktionen auf.

Abb. 10: Weitere Auffälligkeiten zu *wissen*, Bedeutung ‚informiert sein‘

3.2.4 Modul 2

In Modul 2 werden die Funktionen sowohl von ein- als auch von mehrteiligen Lemmata beschrieben. In einem Artikelkopf sind funktionsübergreifende Informationen zu Form, Varianz, Kombinatorik und Forschungsliteratur für Nachschlagende aufbereitet. Direkt angeschlossen werden die einzelnen Funktionalitäten mittels eines kurzen Labels und einer kurzen Funktionsbeschreibung angeboten (vgl. Abb. 11).

schön [DISKURSPARTIKEL]			
Form	Varianz	Kombinatorik	Forschungsliteratur
---	---	sehr schön, voll schön, ja schön, ah schön, oh schön, ach schön, na schön	• Willkop (1988)
<i>Die Partikel schön kann um vorangestellte Intensitätspartikeln (sehr, voll), Interjektionen (ah, oh, ach) oder Diskurspartikeln (ja) ergänzt werden. Diese können jeweils eigenständige Funktionen übernehmen (ja schön, ah schön, oh schön, ach schön) oder die Bedeutung/Funktion von schön modifizieren (ja sehr schön, voll schön, na schön).</i>			
Bewertung ▶ Mit schön nehmen Sprechende Bezug auf eine vorausgegangene Äußerung und bewerten den Inhalt dieser Äußerung oder eine Handlung als positiv (vgl. dazu auch schön in den Bedeutungen passend und erfreulich).			
Abschlussmarker ▶ Mit schön zeigen Sprechende an, dass eine vorausgegangene kommunikative Aufgabe ausreichend bearbeitet wurde, um diese abzuschließen und in der Interaktion fortzufahren.			

Abb. 11: Artikelkopf und Funktionsbeschreibungen im Überblick zu schön als Diskurspartikel

Die funktionsspezifischen Wortartikel in Modul 2 beinhalten das im Folgenden erläuterte Informationsangebot:

- **Belegblock:** Die Funktionsbeschreibung wird durch einen oder mehrere Belege illustriert (vgl. Abb. 12). Im Unterschied zu Modul 1 wird bei den Belegen zur Funktionsbeschreibung eine obligatorische sequenzanalytische Kommentierung angeboten. Für interaktionslinguistische, formfunktionsbasierte Analysen sind prosodische Parameter hochrelevant. Aus diesem Grund wurden die Transkripte, die in Modul 2 angeboten werden, um prosodische Informationen ergänzt und somit zu Basistranskripten ausgebaut. Wenn sich Nutzer/-innen auch mit dieser Transkriptnotation noch näher vertraut machen wollen, finden sie dazu in den Umtexten (vgl. Abschnitt 3.4) vertiefendes Material.

[1] schön zur Bewertung des Inhalts eines Berichts

MI 001 MD ^h aber [heut mo]rgen WAR er dann mal im gang,
 MI 002 MP [^h]
 MI 003 MD hat dann sein BETT auch frisch be[zogen gl]eisch heut morgen,
 MI 004 MP [h^öi,]
 MI 005 MD ^h ähm WA:R,
 MI 006 (θ.ɜ)
 MI 007 MD nicht (.) g der GANZ erste in der visite,
 MI 008 ^h aber der ZWEite heut in der visite,]
 MI 009 ME [((räuspert sich))]
 MI 010 MD un nach der visite gleich nach HAUSE ab-
 MI 011 MD ^h
 MI 012 MP sch[ön];]
 MI 013 MD [un kom]mt morgen abend um ACHT wieder;
 MI 014 MD der is auch über NACHT weg.]
 MI 015 MP [sch^ön];
 MI 016 (6.θɜ)

Belegkontext
 In einer Schichtübergabe im Krankenhaus berichtet die Krankenschwester MD den anderen Mitarbeitenden von einem Patienten, der aufgrund seiner Medikation immer Mühe hat, morgens früh aufzustehen.

Analyse
 MD schildert detailliert das veränderte, positive Verhalten des Patienten. Während ihres Berichts gibt die Kollegin MP Rückmeldungen und bewertet in Überlappung mit den noch laufenden Darlegungen von MD die außergewöhnliche Tatsache, dass der Patient heute Morgen bereits früh aktiv war und erst am nächsten Tag wiederkommt, jeweils mit "schön;" (Z. 012, 015) als positiv.

Institutionell: Schichtübergabe in einem Krankenhaus - FOLK_E_00114_T_01

Abb. 12: Belegblock im Artikel *schön* (Diskurspartikel), Funktion ‚Bewertung‘

- **Funktionsabstrahierung:** Dieser lexikografische Kommentar fasst die jeweils beschriebene Funktion auf einer allgemeineren Ebene und in Rückgriff nicht nur auf die illustrierenden Belege, sondern auf zahlreiche analysierte relevante Belege, zusammen und bildet eine zentrale Informationseinheit in Modul 2 (vgl. Abb. 13).

Mit **schön** nehmen Sprechende Bezug auf eine vorausgegangene Äußerung und bewerten den Inhalt dieser Äußerung oder eine Handlung als positiv bzw. in Lehr-Lern-Interaktionen als korrekt. Mit dieser Bewertung auf Sachverhaltsebene drücken sie ihre Haltung dem jeweiligen Bewertungsgegenstand gegenüber aus, die vor allem durch eine entsprechende prosodische Realisierung unterschiedlich stark emphatisch aufgeladen werden kann. Dabei wird **schön** als positive, emphatische Bewertung häufig dazu genutzt, Anteilnahme sowie subjektive Stellungnahmen und Empfindungen expressiv auszudrücken, und trägt als solche zur Diskursorganisation bei. Dies geschieht am häufigsten als Reaktion auf längere Äußerungen wie Erzählungen, Berichte oder Beschreibungen sowie in dritter Position meist nach einer Frage-Antwort-Sequenz. Eine spezifische Verwendung tritt dabei in schulischen Interaktionen auf, wenn mit **schön** eine Schülerantwort innerhalb einer dreischrittigen Sequenz aus Frage-Antwort-Evaluierung als korrekt bewertet wird. Auffällig ist, dass **schön** sehr häufig als Bewertung verwendet wird. Dabei wird **schön** auch mit weiteren Partikeln wie **oh**, **ah** oder **och** und Intensitätspartikeln wie **sehr** oder **voll** kombiniert sowie häufig durch eine emphatische prosodische Gestaltung hervorgehoben.

Abb. 13: Funktionsabstrahierung im Artikel *schön* (Diskurspartikel), Funktion ‚Bewertung‘

- **Syntax-Sequenz-Realisierung** und **Prosodie:** Kurze Informationen zur Syntax- und Sequenz-Realisierung (vgl. Abb. 14) und zur Prosodie (vgl. Abb. 15), die wie bei *ja* und *eben* als Modalpartikel vs. Diskurspartikel auch distinktiv sein können, stehen zusammen mit illustrativen kurzen Belegausschnitten den Nachschlagenden zur Verfügung.

The screenshot shows a Lexikos entry for the particle 'mal schauen'. The title is 'Syntax-Sequenz-Realisierung'. The main text states: 'mal schauen wird meistens ohne weitere Komplemente realisiert. Es kann turnfinal stehen und den Sprecher beenden.' Below this is a sub-section with a speaker icon and the text 'mal schauen turnfinal, ohne Komplement'. Two example utterances are provided: 'MI 001 JA ich nehm erst mal ne KARTE;' and 'MI 002 JA un dann mal SCHAUen;'. A teal bar at the bottom indicates the source: 'Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00132_T_12'.

Abb. 14: Syntax-Sequenz-Realisierung im Artikel *mal schauen*, Funktion ‚Verschiebung einer Handlung in die Zukunft‘

The screenshot shows a Lexikos entry for the particle 'ja'. The title is 'Prosodie'. The main text states: 'ja wird immer prosodisch integriert und unbetont realisiert.' Below this is a sub-section with a speaker icon and the text 'ja prosodisch integriert, unbetont'. Two example utterances are provided: 'MI 001 AC ich kann ja mal FRAGn;' and 'MI 002 AC aber eigentlich geht sie ned da HIN.'. A teal bar at the bottom indicates the source: 'Privat: Kochinteraktion - FOLK_E_00332_T_01'.

Abb. 15: Prosodie im Artikel *ja* (Modalpartikel), Funktion ‚Abschwächung‘

- **weitere Auffälligkeiten:** Optionale Informationen zu weiteren Auffälligkeiten, die hauptsächlich in Verbindung mit den Metadaten stehen, reflektieren in Form eines lexikografischen Kommentares Beobachtungen, die aus der Analyse der Korpusbelege hervorgegangen sind (vgl. Abb. 16).

schön als Abschlussmarker tritt überwiegend in handlungsschematischen Interaktionen wie Unterrichtsstunden, Planungen oder Spielinteraktionen auf, in denen einzelne Themen und Handlungen aufeinander aufbauen und systematisch nacheinander bearbeitet werden.

Abb. 16: Weitere Auffälligkeiten im Artikel *schön* (Diskurspartikel), Funktion ‚Abschlussmarker‘

3.3 Zugriffsmöglichkeiten und Vernetzungsstrukturen

Die Zugriffsmöglichkeiten auf die vielfältige multimediale Information im LeGeDe-Prototyp sind unterschiedlich und orientieren sich jeweils an der Art der Information. Die Umsetzung berücksichtigt dabei die aktuellen Anforderungen an elektronische Wörterbücher und wird auch den Erwartungen der Befragten aus den LeGeDe-Studien gerecht (vgl. Dziemianko 2018; vgl. insbesondere These 7 in Engelberg, Klosa-Kückelhaus und Müller-Spitzer 2019: 33; Engelberg, Müller-Spitzer und Schmidt 2016; Meliss, Möhrs und Ribeiro Silveira 2018). Zunächst lassen sich wörterbuchinterne von externen Zugriffsmöglichkeiten und Vernetzungsstrukturen unterscheiden.

(a) Die Nachschlagenden können sich über eine Vielzahl von **wörterbuchinterne** Zugriffs- und Vernetzungsstrukturen innerhalb des LeGeDe-Prototyps bewegen. Dadurch ist die Möglichkeit gegeben, Wörterbuchinhalte über verschiedene Zugriffspunkte und Verknüpfungen zwischen dem Bedeutungs- (Modul 1) und dem Funktionsmodul (Modul 2) anzusteuern sowie Glossar- und Literatureinträge aus den Umtexten (vgl. Abschnitt 3.4) direkt abzurufen (vgl. Tab. 1). So können Nachschlagende beispielsweise direkt von der Seite „Übersicht: Index“ auf einen Überblicksartikel oder Modul 1 bzw. 2 eines Lemmas gelangen. Die Lemmaliste ist beim Navigieren zwischen verschiedenen Wortartikeln immer sichtbar, sodass hierdurch bequem zwischen verschiedenen Wortartikeln gewechselt werden kann.

Beschreibung	Verlinkung/Verweis
<p>Übersicht Index Überblick zu allen LeGeDe-Lemmata, den jeweiligen Kurzlabeln der Module 1 und 2 zu Bedeutungen/Funktionen sowie kurze Auszüge aus der Mikrostruktur (z.B. Phonetik, Kombinatorik)</p>	<p><u>Ort:</u> LeGeDe-Startseite und immer sichtbar im linken Menüband <u>Ziel (nach Klick):</u> der jeweils ausgewählte Linkort</p>
<p>Modul 1 → Modul 2 Wörterbuchinterner Verweis zu einer Funktion in Modul 2; der Verweis deutet darauf hin, dass eine inhaltliche Verknüpfung zwischen den Modulen vorliegt.</p>	<p><u>Ort:</u> Kombinatorik → Interaktive Einheit als Form ist verlinkt <u>Ziel (nach Klick):</u> Wörterbuchartikelteil zur relevanten Funktion in Modul 2</p>
<p>Modul 2 → Modul 1 Wörterbuchinterner Verweis zu einer Bedeutung in Modul 1; der Verweis deutet darauf hin, dass eine inhaltliche Verknüpfung zwischen den Modulen vorliegt.</p>	<p><u>Ort:</u> Funktionsbeschreibung; am Ende des Satzes ist das relevante → Bedeutungslabel aus Modul 1 verlinkt <u>Ziel (nach Klick):</u> Wörterbuchartikelteil zur relevanten Bedeutung in Modul 1</p>

<p>Glossar-Stichwörter Erläuterungen eines Fachbegriffes</p>	<p><u>Ort</u>: innerhalb des Wörterbuchartikels (Modul 1, Modul 2); im Fließtext weist eine gestrichelte Linie auf den Link hin <u>Ziel (nach Klick)</u>: Pop-up Fenster mit der Erläuterung, die auch im Umtext „Glossar“ zu finden ist</p>
<p>Kurztitel: Forschungsliteratur Zentrale Forschungsliteratur, die für das jeweilige Lemma relevant ist</p>	<p><u>Ort</u>: 1. Überblicksartikel, 2. Artikelkopf in Modul 1 und 2 <u>Ziel (nach Klick)</u>: Pop-up Fenster mit dem vollständigen bibliografischen Hinweis, der auch im Umtext „Literatur“ zu finden ist</p>

Tab. 1: Wörterbuchinterne Zugriffs- und Vernetzungsstrukturen

(b) Die **externen** Zugriffs- und Vernetzungsstrukturen im LeGeDe-Prototyp führen zu weiteren Ressourcen und lassen sich in drei große Bereiche aufteilen (vgl. Tab. 2⁷): (i) Einerseits geht es um Ressourcen, die eine direkte Verbindung zu der Datengrundlage des LeGeDe-Prototyps herstellen (FOLK, die DGD und der „Lexical Explorer“). (ii) Andererseits handelt es sich um externe Onlinewörterbücher, über die die lexikologische Information zu den einzelnen Stichwörtern erweitert werden kann. (iii) Neben diesen Verknüpfungen wird außerdem ein externer Link zu einer fachterminologischen Datenbank in dem lexikografischen Umtext „Glossar“ angeboten. Die Tabelle 2 resümiert die Information in Verbindung mit dem Ort, an dem die Verlinkung angesteuert werden kann und den möglichen Erreichbarkeitsbedingungen.

Ressource	Verlinkung	Erreichbarkeit
<p>DGD (Datenbank)</p>	<p><u>Ort</u>: LeGeDe-Menüleiste <u>Ziel (nach Klick)</u>: Anmeldeseite der DGD</p>	<p>mit Anmeldung über persönlichen DGD-Account</p>
<p>FOLK in DGD (Korpus abrufbar über Datenbank)</p>	<p><u>Ort</u>: Überblicksartikel <u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis in FOLK über die DGD zum jeweiligen Lemma</p>	

	<p><u>Ort</u>: Belegblock innerhalb der einzelnen Bedeutungen (Modul 1) oder Funktionen (Modul 2)</p> <p><u>Ziel (nach Klick)</u>: belegbezogen: Suchergebnis in FOLK über die DGD zum jeweiligen Transkriptausschnitt</p>	
Lexical Explorer (Tool)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im "Lexical Explorer" zum jeweiligen Lemma</p>	ohne Anmeldung
OWID (Wörterbuchportal)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis in OWID zum jeweiligen Lemma</p>	
eValbu (elekt. Valenzwörterbuch)	<p><u>Ort</u>: Überblicksartikel bei Verben, wenn das Stichwort im eValbu verzeichnet ist</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im eValbu zum jeweiligen Lemma</p>	
DWDS (lexikografisches Informationssystem)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im DWDS zum jeweiligen Lemma</p>	
grammis (grammatisches Informationssystem)	<p><u>Ort</u>: Glossar</p> <p><u>Ziel (nach Klick)</u>: stichwortbezogen: Suchergebnis in grammis zum jeweiligen Glossarstichwort</p>	

Tab. 2: Verlinkungen aus dem LeGeDe-Prototyp zu externen Ressourcen

3.4 Umtexte

Die verschiedenen Umtexte im LeGeDe-Prototyp bieten ein umfassendes Informationsangebot an, welches sich an unterschiedliche Nutzergruppen wendet und die verschiedensten Möglichkeiten einer internetbasierten Ressource ausnutzt (vgl. Klosa und Gouws 2015).

Der Informationstext „Über LeGeDe“ stellt in ausführlicher Form die theoretischen und methodologischen Grundlagen des LeGeDe-Projekts dar und bietet detailliert Information zu der lexikografischen Mikrostruktur und den einzelnen Angabeklassen. In seiner Ausführlichkeit und Komplexität ist dieser

Text besonders an ein Fachpublikum gerichtet, das neben dem Interesse an der Ressource an sich, Interesse an metalexikografischen Fragestellungen hat.

Im Unterschied dazu sind die „Benutzungshinweise“ so konzipiert, dass sie dem Nachschlagenden anschaulich und zusammenfassend den Zugang zu der Ressource und zu den einzelnen Strukturelementen durch eine „Guided Tour“ ermöglichen. Ein separates Dokument führt zudem in die wichtigsten Transkriptionskonventionen ein.

Das „Glossar“ beinhaltet in der Ressource relevante linguistische Fachbegriffe, die zum einen mit einer Kurzdefinition erläutert sind und zum anderen mit einem Verlinkungsangebot zu dem Modul „Wissenschaftliche Terminologie“, das in *grammis* integriert ist, versehen sind. Im Vordergrund steht dabei hauptsächlich das Ziel, den für den LeGeDe-Prototyp verfolgten terminologischen Ansatz zu skizzieren. Der Zugriff auf den entsprechenden Glossareintrag ist dabei entweder direkt über einen entsprechenden Begriff in der Ressource (vgl. Abschnitt 3.3) oder über das alphabetisch angeordnete vollständige Glossar möglich.

Der Umtext „Literatur“ bietet einen Einblick in für die Lemmata im LeGeDe-Prototyp relevante Forschungsliteratur. Zugriffsoptionen existieren entweder über Kurzverweise in den einzelnen Wörterbuchartikeln (vgl. Abschnitt 3.3) oder alphabetisch über eine komplette Liste.

4. Fazit und Ausblick

Abschließend sollen einerseits in Form eines Fazits die Leistungen des LeGeDe-Prototyps aufgezeigt werden. Andererseits wird ein umfangreicher Ausblick geliefert, der auf verschiedene und relevante Aspekte verweist, die bei der Ausarbeitung des LeGeDe-Prototyps nicht berücksichtigt werden konnten, die aber bei zukünftigen lexikografischen Projekten unter Einbindung von korpusbasierten gesprochensprachlichen Daten von Interesse sein können.

4.1 Fazit

Es konnte verdeutlicht werden, dass die neuartigen Aspekte des LeGeDe-Prototyps vielfältig sind und sowohl methodologische als auch inhaltliche, strukturelle und darstellungsorientierte Perspektiven umfassen. Die konkreten innovativen Merkmale der Ressource lassen sich in folgenden Bereichen ausmachen:

(i) **Datengrundlage:** Die dargebotene Information im LeGeDe-Prototyp basiert ausschließlich auf korpusbasierten gesprochensprachlichen Daten und stellt in diesem Sinne ein Novum dar (vgl. Abschnitt 2.4).

(ii) **Stichwortkandidatenliste:** Die Erstellung einer Stichwortkandidatenliste auf der Basis von Häufigkeitsklassenvergleichen zwischen zwei Korpora

(DEREKO als Referenzkorpus für das geschriebene Deutsch und FOLK als Korpus des gesprochenen Deutsch in der Interaktion) bietet eine korpusbasierte Grundlage zur Identifizierung von relevanten standardnahen, gesprochen sprachlichen lexikalischen Elementen (vgl. Abschnitt 3.1).

(iii) **Erwartungshaltungen:** Die Neukonzipierung einer lexikografischen Ressource zur Darstellung von gesprochensprachlichen Spezifika, für die kaum auf Vorbilder zurückgegriffen werden konnte, ermöglichte die konkrete Berücksichtigung von bestimmten empirisch erhobenen Erwartungshaltungen der zukünftigen Benutzer/-innen (vgl. Abschnitt 2.5).

(iv) **Lexikografische Angabeklassen:** Das Informationsangebot zu den angebotenen lexikografischen Angabeklassen kombiniert klassische und neuartige Formen. Die entwickelten Vorschläge für die Darstellung lexikalischer Phänomene des Gesprochenen in der Interaktion ermöglichen, diese für lexikografische Zwecke adäquat zu strukturieren und zu beschreiben. Die korpusbasierten Daten wurden durch eine spezifische Methode quantitativ ermittelt und qualitativ analysiert und strukturiert (vgl. Abschnitt 3.2.1).

(v) **Verknüpfung mit Korpusdaten:** Das Angebot an authentischen Korpusbelegen erfolgt mittels ausgewählter Transkriptausschnitte, über die eine Schnittstelle zu den Audiodateien und den ausführlichen Informationen zu den Metadaten der DGD hergestellt wird. Damit wird der LeGeDe-Prototyp den technischen Herausforderungen gerecht (vgl. Abschnitte 3.2.3, 3.2.4 und 3.3).

(vi) **Multimedialität:** Der multimediale Charakter der Ressource zeichnet sich v.a. dadurch aus, dass für die Korpusdaten neben den Transkripten Audio dateien und teilweise auch entsprechende Videodateien über den Zugriff auf die DGD zur Verfügung stehen. Die Verknüpfung zu dem Analysetool „Lexical Explorer“ ermöglicht den Benutzer/-innen zusätzlich einen direkten Zugriff auf die der Ressource zugrundeliegenden Korpusdaten und eigene erweiterte Analysemöglichkeiten (vgl. u.a. Abschnitte 2.4 und 3.3).

(vii) **Komplexe Struktur der Informationsvernetzung:** Durch diverse interne und externe Verlinkungsmöglichkeiten wird eine komplexe Informationsvernetzung und -strukturierung angeboten. Der LeGeDe-Prototyp wird auf diese Weise dem Internetmedium mit all seinen Vorteilen gerecht (vgl. Abschnitt 3.3).

4.2 Ausblick

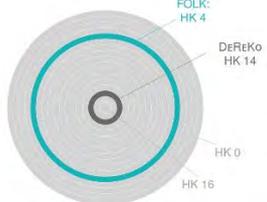
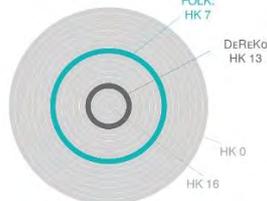
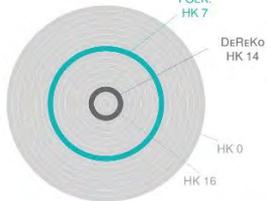
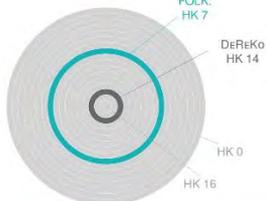
Die vorgestellte lexikografische Ressource liegt zurzeit als Prototyp vor. Die theoretischen und methodologischen Annahmen für die Konzipierung der Ressource sowie die entsprechende technische Umsetzung erlauben jederzeit eine Erweiterung des lexikografischen Angebots. Als Ausblick sollen daher einige Ausbaumöglichkeiten aufgezeigt werden, die nicht nur für eine Erwei-

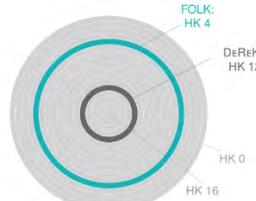
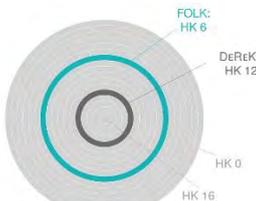
terung des LeGeDe-Prototyps von Interesse sind, sondern auch bei der lexikografischen Aufbereitung von anderen mündlichen Korpusdaten für zukünftige Projekte relevant sein könnten.

Diese potenziellen Ausbaumöglichkeiten betreffen verschiedene Ebenen, aus denen folgende zwei für eine genauere Betrachtung ausgewählt wurden: (i) die Makrostruktur und konkrete Vorschläge für die Erweiterung der Anzahl der (nicht) redaktionell bearbeiteten Stichwörter und (ii) unterschiedliche Recherche- und Zugriffsmöglichkeiten auf die zahlreichen Informationsmodule. Im Folgenden sollen diese Möglichkeiten eingehender beschrieben werden.

4.2.1 Art und Anzahl der Stichwörter

Über den korpusbasierten Stichwortansatz, der für den LeGeDe-Prototyp entwickelt wurde (vgl. Abschnitt 3.1), konnten 322 für die gesprochene Sprache relevante Stichwortkandidaten ermittelt werden, die verschiedenen grammatischen Kategorien angehören (z.B. Partikeln, Interjektionen, Adverbien) und/oder verschiedene Bereiche abdecken (z.B. Routineformeln, Vagheitsausdrücke, deiktische Ausdrücke, Lemmata mit Besonderheiten hinsichtlich Kombinatorik, Stil und Register). Da die LeGeDe-Ressource zunächst als lexikografischer Prototyp konzipiert wurde, erfolgte die Auswahl der behandelten Stichwörter aus der Stichwortkandidatenliste nach Kriterien in Zusammenhang mit dem Anspruch, eine Bandbreite von unterschiedlichen Phänomenen bzw. Erscheinungen abzudecken und dabei gleichzeitig auch die Ergebnisse wissenschaftlicher Teilstudien zu berücksichtigen. Dennoch konnten in der Projektlaufzeit nicht alle Besonderheiten erfasst werden. Bereits die Liste der 25 Stichwortkandidaten aus der Rangliste (vgl. Abb. 1) verweist auf unterschiedliche Ausschnitte, wie z.B. Interjektionen, die im LeGeDe-Prototyp bislang nicht beschrieben worden sind. Die Bearbeitung der Bereiche, die in den LeGeDe-Prototyp der ersten Arbeitsphase nicht aufgenommen werden konnten, bildet zusammen mit der quantitativen Erweiterung der Lemmaanahl zu den schon behandelten Phänomenen ein breites zukünftiges lexikografisches Arbeitsfeld. Daher wäre es durchaus denkbar, neben den redaktionell bearbeiteten Stichwörtern auch eine Auswahl relevanter Stichwörter anzubieten, die redaktionell nicht ausführlich bearbeitet wurden, für die jedoch entsprechende Verlinkungsangebote in der Ressource bereitgestellt werden könnten. Ein solches Angebot würde es interessierten Nutzer/-innen ermöglichen, eine eigenständige Recherche durchzuführen. In Tab. 3 wird exemplarisch zu Stichwörtern der 25-er Rangliste, die unterschiedlichen Phänomenbereichen zugeordnet wurden, ein Vorschlag für ein Informationsangebot mit verschiedenen Verlinkungsangeboten und der Visualisierung des Häufigkeitsklassenvergleichs präsentiert, das den Recherchierenden ohne größeren redaktionellen Aufwand zur Verfügung gestellt werden könnte, um schnell und gezielt quantitative Informationen zu verschiedenen grammatikalischen und metalinguistischen Daten zu erhalten.

Weitere Lemmakandidaten: nicht redaktionell ausgearbeitetes Informationsangebot	
<p>ah</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Interjektionen (<i>ah, ach, oh</i> etc.)</p>	 <p>FOLK-Zugriff ah (in FOLK) ah (im Lexical Explorer)</p>
<p>cool</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Diskurspartikel/ Adjektive// Entlehnungen (<i>okay, cool</i> etc.)</p>	 <p>FOLK-Zugriff cool (in FOLK) cool (im Lexical Explorer)</p>
<p>nachher</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Adverbien//temporale und lokale Deixis (<i>nachher, jetzt; hier, da</i>, etc.)</p>	 <p>FOLK-Zugriff nachher (in FOLK) nachher (im Lexical Explorer)</p>
<p>danke</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Partikeln//Routineformeln (<i>danke, bitte</i> etc.)</p>	 <p>FOLK-Zugriff danke (in FOLK) danke (im Lexical Explorer)</p>

<p>irgendwie</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Partikeln//Vagheitsmarker (<i>irgendwie, irgendwas</i> etc.)</p>	 <p>FOLK-Zugriff irgendwie (in FOLK) irgendwie (im Lexical Explorer)</p>
<p>raus</p> <p>Information in anderen lexikografischen Ressourcen OWID</p> <p>Adverbien/Trennbare Verbpartikeln// Formreduktion (<i>raus, rein, rauf</i> etc.)</p>	 <p>FOLK-Zugriff raus (in FOLK) raus (im Lexical Explorer)</p>

Tab. 3: Mögliches Informationsangebot für nicht redaktionell bearbeitete Artikel⁸

4.2.2 Zugriffs- und Suchfunktionen

Bislang erfolgt der Zugriff auf die Wortartikel im LeGeDe-Prototyp alphabetisch-semasiologisch. Die Datenanalysen und -modellierungen wurden jedoch so angelegt, dass auch Zugriffsmöglichkeiten über formale und funktionale Kategorien technisch realisierbar wären. Ein formal-kategoriales, syntaxsequenz- und prosodieorientiertes sowie funktional-strukturiertes Informationsangebot und eine entsprechende Suchanfrage für diese erweiterten Zugriffsarten wurden von den Probanden unserer Umfragen erwünscht und stellen v.a. für anwendungsorientierte Zwecke im DaF- und DaZ-Bereich in Zusammenhang mit dem Nachschlageverhalten für Produktionszwecke einen interessanten Mehrwert dar (vgl. Meliss, Möhrs und Ribeiro Silveira 2019: 111-112; Sieberg 2013), wobei außerdem auch den Anforderungen des GeR zur Kompetenzentwicklung in der mündlichen Interaktion Rechnung getragen würde.

Hierfür schlagen wir u.a. folgende Möglichkeiten vor, die (fast) alle bereits in den aufbereiteten LeGeDe-Daten entsprechend ausgezeichnet sind:

- (i) **rekurrente Muster:** z.B. <mal + visuelles Verb> (*mal schauen, mal gucken, mal sehen*), <Imperativform eines visuellen Verbs + mal> (*guck mal, schau mal*) oder <weiß X> (*weiß der Geier / der Kuckuck*)
- (ii) **grammatische Kategorien:** z.B. formal-kategoriale Auszeichnungen der Wortarten (Nomen, Verb, Diskurspartikel etc.)
- (iii) **Syntax-Sequenz-Merkmale:** z.B. Merkmale zur syntaktischen Integration (desintegriert, integriert etc.), Sequenzposition (initial, final etc.)
- (iv) **Prosodie-Merkmale:** z.B. prosodische Merkmale wie Betonung und prosodische Integration
- (v) **Funktionen in der Interaktion:** Im LeGeDe-Prototyp sind über die einzelnen Funktionslabel die Ansätze für eine funktionale Suchanfrage gelegt. In Meliss (2020a) wird ein erweiterter Vorschlag diesbezüglich vorgelegt, der die Funktionskategorien im LeGeDe-Prototyp mit Informationen aus einschlägiger Fachliteratur verbindet. In Tab. 4 wird exemplarisch eine Auflistung der möglichen sprachlichen Mittel zum „Ausdruck von Vagheit“ zusammen mit weiteren Subfunktionalitäten aufgeführt.

Funktion	Subfunktionalitäten	Beispiele (fett = in LeGeDe-Prototyp beschrieben)
Ausdruck von Vagheit	Unschärfemarker: epistemic hedge	<ul style="list-style-type: none"> • <i>(ich) weiß nicht</i> • <i>was weiß ich</i> • <i>keine Ahnung</i> • <i>weiß der Teufel was</i> • <i>was weiß ich was</i> • <i>irgendwie</i> • <i>oder so</i> • <i>ungefähr</i> • <i>oder so (was)</i> • <i>glaub ich</i>
	Anzeige von Unschärfe eines Beispiels	<ul style="list-style-type: none"> • <i>was weiß ich</i> • <i>keine Ahnung</i> • <i>zum Beispiel</i> • <i>beispielsweise</i> • <i>sagen wir mal</i>
	Anzeige von Unschärfe einer Listung: ‚etcetera-Formeln‘	<ul style="list-style-type: none"> • <i>(und) was weiß ich</i> • <i>keine Ahnung</i> • <i>weiß der Geier (was)</i> • <i>weiß der Kuckuck (was)</i> • <i>und so (weiter) (und so fort)</i> • <i>und und und</i> • <i>bla bla bla</i> • <i>etcetera pp</i>

Tab. 4: Sprachliche Mittel zum „Ausdruck von Vagheit“ (Auswahl)

Die Liste (vgl. Tab 4) setzt sich aus den Lemmata zusammen, die im LeGeDe-Prototyp beschrieben sind, und aus weiteren, die u.a. in Fachstudien angeführt werden. Über ein Informationsangebot dieser Art könnten die Benutzer/-innen aus einer Palette von mehreren Ausdrucksmöglichkeiten auswählen.

Abschließend kann festgehalten werden, dass der LeGeDe-Prototyp dank der unterschiedlichen Zugriffe auf die Korpusdaten und dem strukturierten und selektierten multimedialen Informationsangebot neben seinem primären Ziel der sprachwissenschaftlichen Dokumentation vielfältige Möglichkeiten für anwendungsorientierte Szenarien bietet. Das Glossar und die Benutzungshinweise, die als Umtexte zur Verfügung stehen, leisten dazu u.a. eine adäquate Hilfe. Bedingung dafür ist allerdings, dass die Sprachdidaktiker/-innen diese Nutzungsmöglichkeiten kennen, um eine entsprechende Vermittlerrolle einnehmen zu können (vgl. Meliss und Möhrs 2018, Meliss i. Dr. [a]). Daher soll an dieser Stelle auch für eine intensivere Beschäftigung mit der anwendungsorientierten Korpuslinguistik und der Lexikografie im Studium und für Weiterbildungsangebote plädiert werden. Darüber hinaus zeichnen sich auch für die Lexikografie des Gesprochenen Ansatzpunkte für neue technologische Herausforderungen in Zusammenhang mit der künstlichen Intelligenzforschung, der Spracherkennung und Sprachproduktion ab, die für die künftige Lexikografie neue Wege bahnen und u.a. eine mobile, multimediale und interaktive Anwendung zum Ziel haben können. In diesem Sinne schließen wir uns der folgenden Einschätzung von Tarp und Gouws (2019: 266) an:

Modern-day lexicographers are in a position to make some of the unfulfilled dreams of the past a reality. The challenge of the future is to make the impossible possible. We have work to do.

Endnoten

- * Wir danken den anonymen Gutachter/-innen für die vielen wertvollen Hinweise und Anmerkungen.
- 1. Vgl. *elexiko*-Glossareintrag zum *elexiko*-Korpus:
<https://www.owid.de/wb/elexiko/glossar/elexiko-Korpus.html>
[zuletzt abgerufen: 14.08.2020].
- 2. Leibniz-Wettbewerb 2016, Förderlinie „Innovative Vorhaben“:
<https://www.leibniz-gemeinschaft.de/forschung/leibniz-wettbewerb/gefoidertevorhaben.html>
[zuletzt abgerufen: 14.08.2020].
- 3. LeGeDe-Projektwebseite:
<https://www1.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch.html>
[zuletzt abgerufen: 14.08.2020].
- 4. Empirische Methoden-Projektwebseite:
URL: <https://www1.ids-mannheim.de/lexik/empirische-methoden.html>
[zuletzt abgerufen: 14.08.2020].

5. Die Eckdaten und Rahmenbedingungen der Studien können in Meliss, Möhrs und Ribeiro Silveira (2018: 107-110) genauer nachgelesen werden. Vgl. auch die Informationen zu den Studien auf der LeGeDe-Projektwebseite:
<https://www1.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch/projektbeschreibung/empirische-forschung.html>
[zuletzt abgerufen: 14.08.2020].
6. Umtext „Über LeGeDe“:
URL: <https://www.owid.de/legede/about.jsp>
[zuletzt abgerufen: 14.08.2020].
7. Die Links zu allen in den Tabellen aufgeführten Ressourcen können in der Bibliographie eingesehen werden.
8. Bei den Links zu den in der Tabelle angegebenen Lemmata zu FOLK ist (genauso wie im LeGeDe-Prototyp) zu beachten, dass das Abfrageergebnis in der DGD nur nach vorheriger kostenloser Anmeldung eingesehen werden kann (vgl. dazu auch die Information in Tab. 2 in Abschnitt 3.3).

Bibliografie

- Albert, G. und S. Diao-Klaeger (Hrsg.).** 2018. *Mündlicher Sprachgebrauch zwischen Normorientierung und pragmatischen Spielräumen*. Stauffenburg Linguistik 101. Tübingen: Stauffenburg.
- Barcala Rodríguez, F.M. et al.** 2018. El corpus ESLORA de español oral: diseño, desarrollo y explotación. *CHIMERA: Romance Corpora and Linguistic Studies* 5(2): 217-237.
- Bergmann, P.** 2017. Gebrauchsprofile von *weiß nich* und *keine Ahnung* im Gespräch. Ein Blick auf nicht-responsive Vorkommen. Blühdorn, H. et al. (Hrsg.). 2017. *Diskursmarker im Deutschen. Reflexionen und Analysen*: 157-182. Göttingen: Verlag für Gesprächsforschung.
- Couper-Kuhlen, E. und M. Selting.** 2018. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- Deppermann, A.** 2007. *Grammatik und Semantik aus gesprächsanalytischer Sicht*. Linguistik — Impulse und Tendenzen 14. Berlin/New York: De Gruyter.
- Deppermann, A., N. Proske und A. Zeschel (Hrsg.).** 2017. *Verben im interaktiven Kontext. Bewegungs- verben und mentale Verben im gesprochenen Deutsch*. Studien zur deutschen Sprache 74. Tübingen: Narr.
- Dziemianko, A.** 2018. Electronic Dictionaries. Fuertes-Olivera, P.A. (Hrsg.). 2018. *The Routledge Handbook of Lexicography*: 663-683. London/New York: Routledge.
- Eichinger, L.M.** 2017. Gesprochene Alltagssprache. Deutsche Akademie für Sprache und Dichtung/ Union der deutschen Akademien der Wissenschaften (Hrsg.). 2017. *Vielfalt und Einheit der deutschen Sprache. Zweiter Bericht zur Lage der deutschen Sprache*: 283-331. Tübingen: Stauffenburg.
- Engelberg, S. et al.** 2011. Argumentstrukturmuster als Konstruktionen? Identität — Verwandtschaft — Idiosynkrasien. Engelberg, S., A. Holler und K. Proost (Hrsg.). 2011. *Sprachliches Wissen zwischen Lexikon und Grammatik*: 71-112. Jahrbuch des Instituts für Deutsche Sprache 2010. Berlin: De Gruyter.

- Engelberg, S.** 2015. Gespaltene Stimuli bei Psych-Verben. Kombinatorische Mustersuchen in Korpora zur Ermittlung von Argumentstrukturverteilungen. Engelberg, S. et al. (Hrsg.). 2015. *Argumentstruktur zwischen Valenz und Konstruktion*: 469-491. Studien zur deutschen Sprache 68. Tübingen: Narr.
- Engelberg, S.** 2018. The Argument Structure of Psych-verbs: A Quantitative Corpus Study on Cognitive Entrenchment. Boas, H. und A. Ziem (Hrsg.). 2018. *Constructional Approaches to Syntactic Structures in German*: 47-84. Trends in Linguistics Studies and Monographs 322. Berlin: De Gruyter.
- Engelberg, S.** 2019. Argumentstrukturmuster. Ein elektronisches Handbuch zu verbalen Argumentstrukturen im Deutschen. Czicza, D., V. Dekalo und G. Diewald (Hrsg.). 2019. *Konstruktionsgrammatik VI. Varianz in der konstruktionalen Schematizität*: 13-38. Tübingen: Stauffenburg.
- Engelberg, S., A. Klosa-Kückelhaus und C. Müller-Spitzer.** 2019. Lexikographie zwischen Grimm und Google? *Sprachreport* 35(2): 30-34.
- Engelberg, S., C. Müller-Spitzer und T. Schmidt.** 2016. Vernetzungs- und Zugriffsstrukturen. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 153-196. Berlin: De Gruyter.
- Fandrych, C., C. Meißner und F. Wallner (Hrsg.).** 2017. *Gesprochene Wissenschaftssprache — digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg.
- Fandrych, C., C. Meißner und F. Wallner.** 2018. Das Potenzial mündlicher Korpora für die Sprachdidaktik. Das Beispiel GeWiss. *Deutsch als Fremdsprache* 55(1): 3-13.
- Fiehler, R.** 2016. Gesprochene Sprache. Wöllstein, A. (Hrsg.). 2016. *Duden — Die Grammatik. Unentbehrlich für richtiges Deutsch*: 1181-1260. Berlin: Dudenverlag.
- Geyken, A. und L. Lemnitzer.** 2016. Automatische Gewinnung von lexikografischen Angaben. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 197-248. Berlin: De Gruyter.
- Günthner, S.** 2018. Routinisierte Muster in der Interaktion. Deppermann, A. und S. Reineke (Hrsg.). 2018. *Sprache im kommunikativen, interaktiven und kulturellen Kontext*: 29-50. Berlin/New York: De Gruyter.
- Handwerker, B., R. Bäuerle und B. Sieberg (Hrsg.).** 2016. *Gesprochene Fremdsprache Deutsch. Perspektiven Deutsch als Fremdsprache* 32. Baltmannsweiler: Schneider.
- Hansen, C. und M.H. Hansen.** 2012. A Dictionary of Spoken Danish. Fjeld, R.V. und J.M. Torjusen (Hrsg.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August 2012, Oslo, Norway*: 929-935. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Helmer, H. und A. Deppermann.** 2017. ICH WEIß NICHT zwischen Assertion und Diskursmarker: Verwendungsspektren eines Ausdrucks und Überlegungen zu Kriterien für Diskursmarker. Blühdorn, H. et al. (Hrsg.). 2017. *Diskursmarker im Deutschen. Reflexionen und Analysen*: 131-156. Göttingen: Verlag für Gesprächsforschung.
- Helmer, H., A. Deppermann und S. Reineke.** 2017. Antwort, epistemischer Marker oder Widerspruch? Sequenzielle, semantische und pragmatische Eigenschaften von *ich weiss nicht*. Deppermann, A., N. Proske und A. Zeschel (Hrsg.). 2017: 377-406.
- Imo, W.** 2007. *Construction Grammar und Gesprochene-Sprache-Forschung. Konstruktionen mit zehn matrixsatzfähigen Verben im gesprochenen Deutsch*. Germanistische Linguistik 275. Berlin: De Gruyter.

- Imo, W. und S.M. Moraldo (Hrsg.).** 2015. *Interaktionale Sprache und ihre Didaktisierung im DaF-Unterricht*. Deutschdidaktik 4. Tübingen: Stauffenburg.
- Klosa, A.** 2013a. Aktuelle Tendenzen in der deutschen Lexikographie der Gegenwart. Stickel, G. und T. Váradi (Hrsg.). 2013. *Lexical Challenges in a Multilingual Europe. Contributions to the Annual Conference 2012 of EFNIL in Budapest*: 75-93. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 99. Frankfurt am Main: Peter Lang.
- Klosa, A.** 2013b. The Lexicographical Process (with Special Focus on Online Dictionaries). Gouws, R.H. et al. (Hrsg.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 517-524. Handbücher zur Sprach- und Kommunikationswissenschaft 5.4. Berlin: De Gruyter.
- Klosa, A. und R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.
- Klosa, A., U. Schnörch und S. Schoolaert.** 2010. Stichwort, Stichwortliste und Eigennamen in *lexiko*. Dykstra, A. und T. Schoonheim (Hrsg.). 2010. *Proceedings of the XIV EURALEX International Congress, 6–10 July 2010, Leeuwarden, the Netherlands*: 653-663. Leeuwarden: Afûk.
- Klosa, A. und C. Tiberius.** 2016. Der lexikografische Prozess. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 65-110. Berlin: De Gruyter.
- Kupietz, M. et al.** 2018. The German Reference Corpus DEREKO: New Developments — New Opportunities. Calzolari, N. et al. (Hrsg.). 2018. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, May 7–12, 2018, Miyazaki, Japan*: 4353-4360. Miyazaki: European Language Resources Association.
- Kupietz, M. und H. Keibel.** 2009. The Mannheim German Reference Corpus (DEREKO) as a Basis for Empirical Linguistic Research. Minegishi, M. und Y. Kawaguchi (Hrsg.). 2009. *Working Papers in Corpus-based Linguistics and Language Education*: 53-59. Tokyo: Tokyo University of Foreign Studies (TUFSS).
- Kupietz, M. und T. Schmidt.** 2015. Schriftliche und mündliche Korpora am IDS als Grundlage für die empirische Forschung. Eichinger, L.M. (Hrsg.). 2015. *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*: 297-322. Jahrbuch des Instituts für Deutsche Sprache 2014. Berlin: De Gruyter.
- Lemmenmeier-Batinić, D.** 2020. Lexical Explorer: Extending Access to the Database for Spoken German for User-specific Purposes. *Corpora* 15: 55-76.
- Meliss, M. i. Dr. [a].** Die LeGeDe-Ressource: korpusbasierte lexikografische Einblicke und anwendungsorientierte Ausblicke. *Zeitschrift DaF* 1/2021.
- Meliss, M. i. Dr. [b].** Das Spektrum der Angabeklassen in der lexikografischen Ressource zum gesprochenen Deutsch: LeGeDe. Piosik, M., J. Taborek und M. Woźnicka (Hrsg.). i. Dr. 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven, 19.–20. Oktober 2018, Adam-Mickiewicz-Universität in Poznań, Polen*. *Lexicographica: Series Maior*. Berlin/New York: De Gruyter.
- Meliss, M.** 2016. Gesprochene Sprache in DaF-Lernerwörterbüchern. Handwerker, B., R. Bäuerle und B. Sieberg (Hrsg.). 2016. *Gesprochene Fremdsprache Deutsch*: 179-199. *Perspektiven Deutsch als Fremdsprache* 32. Baltmannsweiler: Schneider.
- Meliss, M.** 2020a. Musterhaftigkeit in der Lexik des gesprochenen Deutsch: eine handlungs- und funktionsorientierte Annäherung mit lexikografischer Anwendungsperspektive. Mellado Blanco, C. et al. (Hrsg.). 2020. *Produktive Muster in der Phraseologie*: 181-202. *Studia Phraseologica et Paroemiologica* 4. Hamburg: Verlag Dr. Kovač.

- Meliss, M.** 2020b. Lexik des gesprochenen Deutsch in der Interaktion: Partikelverben und ihre Besonderheiten. García Adánez, I. et al. (Hrsg.). 2020. Dr. *Das Leben in einem Rosa Licht sehen. Festschrift für Rosa Piñel/Homenaje a Rosa Piñel*: 77-97. Perspektiven der Germanistik und Komparatistik in Spanien 16. Bern: Peter Lang.
- Meliss, M. et al.** 2018. Creating a List of Headwords for a Lexical Resource of Spoken German. Krek, S. et al. (Hrsg.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia*: 1009-1016. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Meliss, M. et al.** 2019. A Corpus-based Lexical Resource of Spoken German in Interaction. Kosem, I. et al. (Hrsg.). 2019. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 783-804. Brno: Lexical Computing CZ, s.r.o.
- Meliss, M. und C. Möhrs.** 2017. Die Entwicklung einer lexikografischen Ressource im Rahmen des Projektes LeGeDe. *Sprachreport* 33: 42-52.
- Meliss, M. und C. Möhrs.** 2018. Lexik in der spontanen, gesprochensprachlichen Interaktion: Eine anwendungsorientierte Annäherung aus der DaF-Perspektive. *German as a Foreign Language* 3: 79-110.
- Meliss, M. und C. Möhrs.** 2019. Herausforderungen eines innovativen lexikografischen Projekts zu Besonderheiten des gesprochenen Deutsch in der Interaktion. Doval, I. und E. Liste Lamas (Hrsg.). 2019. *Germanistik im Umbruch — Linguistik, Übersetzung und DaF*: 13-26. Berlin: Frank & Timme.
- Meliss, M., C. Möhrs und M. Ribeiro Silveira.** 2018. Erwartungen an eine korpusbasierte lexikografische Ressource zur ‚Lexik des gesprochenen Deutsch in der Interaktion‘: Ergebnisse aus zwei empirischen Studien. *Zeitschrift für Angewandte Linguistik* 68: 103-138.
- Meliss, M., C. Möhrs und M. Ribeiro Silveira.** 2019. Anforderungen und Erwartungen an eine lexikografische Ressource des gesprochenen Deutsch aus der L2-Lernerperspektive. *Lexicographica* 34: 89-121.
- Möhrs, C.** i. Dr. Quantitative und qualitative Ansätze zu Stichwortkandidaten für die lexikografische Ressource zum gesprochenen Deutsch: LeGeDe. Piosik, M., J. Taborek und M. Woźnicka (Hrsg.). i. Dr. 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven, 19.–20. Oktober 2018, Adam-Mickiewicz-Universität in Poznań, Polen*. *Lexicographica: Series Maior*. Berlin/New York: De Gruyter.
- Möhrs, C.** 2020. „Hast du eine Ahnung, ...?“ Eine lexikografische und korpusbasierte Untersuchung am Beispiel des Lexems *Ahnung*. Behr, J. et al. (Hrsg.). 2020. *Schnittstellen der Germanistik. Festschrift für Hans Bickes*: 185-214. Berlin: Peter Lang.
- Möhrs, C. und S. Torres Cajo.** 2020. The Microstructure of a Lexicographical Resource of Spoken German: Meanings and Functions of the Lemma *eben*. *Rasprave: Journal of the Institute of Croatian Language and Linguistics* 46(2): 921-939.
- Möhrs, C., M. Meliss und D. Batinić.** 2017. LeGeDe — Towards a Corpus-based Lexical Resource of Spoken German. Kosem, I. et al. (Hrsg.). 2017. *Electronic Lexicography in the 21st Century, Proceedings of eLex 2017 Conference, Leiden, the Netherlands, 19–21 September 2017*: 281-298. Brno: Lexical Computing CZ s.r.o.
- Moon, R.** 1998. On Using Spoken Data in Corpus Lexicography. Fontenelle, T. et al. (Hrsg.). 1998. *Proceedings of the 8th EURALEX International Congress, 4–8 August 1998, Liège, Belgium*: 347-355. Liège: English and Dutch Departments, University of Liège.

- Moraldo, S. M. und F. Missaglia (Hrsg.).** 2013. *Gesprochene Sprache im DaF-Unterricht. Grundlagen — Ansätze — Praxis. Sprache — Literatur und Geschichte* 43. Heidelberg: Winter.
- Reeg, U., P. Gallo und S. M. Moraldo (Hrsg.).** 2012. *Gesprochene Sprache im DaF-Unterricht. Zur Theorie und Praxis eines Lerngegenstandes. Interkulturelle Perspektiven in der Sprachwissenschaft und ihrer Didaktik* 3. Berlin: Waxmann.
- Schmidt, T.** 2014a. The Research and Teaching Corpus of Spoken German — FOLK. Calzolari, N. et al. (Hrsg.). 2014. *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, May 26–31, 2014, Reykjavik, Iceland*: 383-387. Reykjavik: European Language Resources Association.
- Schmidt, T.** 2014b. Gesprächskorpora und Gesprächsdatenbanken am Beispiel von FOLK und DGD. *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion* 15: 196-233.
- Schmidt, T.** 2014c. The Database for Spoken German — DGD2. Calzolari, N. et al. (Hrsg.). 2014. *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, May 26–31, 2014, Reykjavik, Iceland*: 1451-1457. Reykjavik: European Language Resources Association.
- Schmidt, T.** 2017. DGD — Die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. *Zeitschrift für germanistische Linguistik. Deutsche Sprache in Gegenwart und Geschichte* 45(3): 451-463.
- Schmidt, T.** 2018. Gesprächskorpora. Kupietz, M. und T. Schmidt (Hrsg.). 2018. *Korpuslinguistik: 209-230. Germanistische Sprachwissenschaft um 2020, Bd. 5. Berlin/Boston: De Gruyter.*
- Schnörch, U.** 2005. Die *lexiko*-Stichwortliste. Haß, U. (Hrsg.). 2005. *Grundfragen der elektronischen Lexikographie. lexiko — das Online-Informationssystem zum deutschen Wortschatz*: 71-90. Schriften des Instituts für Deutsche 12. Berlin/New York: De Gruyter.
- Schwittalla, J.** 2012. *Gesprochenes Deutsch. Eine Einführung*. 4. Aufl. Grundlagen der Germanistik 33. Berlin: Schmidt.
- Sieberg, B.** 2013. *Sprechen lehren, lernen und verstehen. Stufenübergreifendes Studien- und Übungsbuch für den DaF-Bereich*. Tübingen: Julius Groos.
- Siepmann, D.** 2015. Dictionaries and Spoken Language: A Corpus-based Review of French Dictionaries. *International Journal of Lexicography* 28(2): 139-168.
- Stadler, H.** 2014. *Die Erstellung der Basislemmaliste der neuhochdeutschen Standardsprache aus mehrfach linguistisch annotierten Korpora*. OPAL — Online publizierte Arbeiten zur Linguistik 5(2014). Mannheim: Institut für Deutsche Sprache.
- Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249. URL: <https://lexikos.journals.ac.za/pub/issue/view/87> [zuletzt abgerufen: 14.08.2020].
- Tarp, S. und R. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268. URL: <https://lexikos.journals.ac.za/pub/issue/view/87> [zuletzt abgerufen: 14.08.2020].
- Torres Cajo, S.** 2019. Zwischen Strukturierung, Wissensmanagement und Argumentation im Gespräch — Interaktionale Verwendungsweisen der Modalpartikeln *halt* und *eben* im gesprochenen Deutsch. *Deutsche Sprache* 47(4): 289-310.
- Trap-Jensen, L.** 2004. Spoken Language in Dictionaries: Does it really matter? Williams, G. und S. Vessier (Hrsg.). 2004. *Proceedings of the 11th EURALEX International Congress, 6–10 July 2004, Lorient, France*: 311-318. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

- Trim, J. et al.** 2001. *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Verdonik, D. und M. Sepesy Maučec.** 2017. A Speech Corpus as a Source of Lexical Information. *International Journal of Lexicography* 30: 143-166.
- Westpfahl, S.** 2014. STTS 2.0? Improving the Tagset for the Part-of-Speech-Tagging of German Spoken Data. Levin, L. und M. Stede (Hrsg.). 2014. *LAW VIII. The 8th Linguistic Annotation Workshop in conjunction with COLING 2014. Proceedings of the Workshop, August 23–24, 2014, Dublin, Ireland*: 1-10. Dublin: Association for Computational Linguistics.
- Westpfahl, S. und T. Schmidt.** 2016. FOLK-Gold — A GOLD Standard for Part-of-Speech Tagging of Spoken German. Calzolari, N. et al. (Hrsg.). 2016. *Proceedings of the Tenth International Conference on Language Resources and Evaluation. May 23-28, 2016. Portorož, Slovenia*: 1493-1499. Portorož: European Language Resources Association.
- Wiegand, H.E.** 1983. Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. Wiegand, H.E. (Hrsg.). 1983. *Studien zur neuhochdeutschen Lexikographie III*: 401-474. Germanistische Linguistik 1–4. Hildesheim: Georg Olms.
- Zeschel, A.** 2017. *Denken und wissen im gesprochenen Deutsch*. Deppermann, A., N. Proske und A. Zeschel (Hrsg.). 2017. *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*: 249-336. Studien zur deutschen Sprache 74. Tübingen: Narr.

Wörterbücher und Online-Ressourcen

(Online-Ressourcen: alle zuletzt abgerufen am 14.08.2020)

DEREKO = Deutsches Referenzkorpus.

URL: <http://www.ids-mannheim.de/kl/projekte/korpora/>

DGD = Datenbank Gesprochenes Deutsch.

URL: <http://dgd.ids-mannheim.de/>

Duden-DaF = *Duden, Standardwörterbuch — Deutsch als Fremdsprache* / hrsg. von der Dudenredaktion. 2., neu bearbeitete und erweiterte Auflage. Berlin: Dudenverlag, 2010.

Duden-online. URL: <https://www.duden.de/>

Duden-Universalwörterbuch = *Duden, deutsches Universalwörterbuch* / hrsg. von der Dudenredaktion; red. Bearbeitung Melanie Kunkel u.a.; unter Mitarbeit von Dr. Laura Sturm und Lena Wagner. 9., vollständig überarbeitete und erweiterte Auflage. Berlin: Dudenverlag, 2019.

DWDS = Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart.

URL: <https://www.dwds.de/>

elexiko = Online Wörterbuch zur deutschen Gegenwartssprache im Verbund OWID (2003ff.).

Leibniz-Institut für Deutsche Sprache, Mannheim.

URL: <https://www.owid.de/docs/elex/start.jsp>

E-Valbu = Elektronisches Valenzwörterbuch deutscher Verben.

URL: <https://grammis.ids-mannheim.de/verbvalenz>

FOLK = Forschungs- und Lehrkorpus Gesprochenes Deutsch.

URL: <http://agd.ids-mannheim.de/folk.shtml>

GeWiss = Gesprochenes Wissenschaftsdeutsch.

URL: <https://gewiss.uni-leipzig.de/>

grammis = Grammatisches Informationssystem.

URL: <https://grammis.ids-mannheim.de/>

LEO. URL: <https://www.leo.org/>

LGWB DaF = Götz, Dieter (Hg.). 2015. *Langenscheidt Großwörterbuch Deutsch als Fremdsprache*. Neubearbeitung. München: Langenscheidt.

LeGeDe = Lexik des gesprochenen Deutsch.

URL (Projektwebseite): <http://www.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch.html>

LeGeDe-Ressource. URL: <https://www.owid.de/legede/>

Lexical Explorer. URL: <https://www.owid.de/lexex/>

Linguee. URL: <https://www.linguee.de/>

LothWB = Wörterbuch der deutsch-lothringischen Mundarten. Bearb. von Michael Ferdinand Follmann. Leipzig 1909. (Quellen zur lothringischen Geschichte — Documents de l'Histoire de la Lorraine 12). Onlinefassung über das Wörterbuchnetz.

URL: http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=LothWB

OWID = Online-Wortschatz-Informationssystem Deutsch.

URL: <https://www.owid.de>

OWID^{plus} = Online-Wortschatz-Informationssystem Deutsch ^{plus}.

URL: <https://www.owid.de/plus/index.html>

PONS-online = PONS Online-Wörterbuch. Einfach nachschlagen und übersetzen.

URL: <https://de.pons.com/%C3%BCbersetzung>

Users and Contexts of Use of Romanian Multilingual Dictionaries

Mihaela Mocanu, *Social Sciences and Humanities Research Department,
Institute for Interdisciplinary Research, Alexandru Ioan Cuza
University of Iași, Romania (mocanu.mihaela@uaic.ro)*

and

Alina-Mihaela Bursuc, *A. Philippide Institute of Romanian Philology,
Romanian Academy, Iași Branch, Romania
(alina.bursuc@hotmail.com)*

Abstract: A significant number of multilingual dictionaries (in at least three languages) were published in Romania after the Revolution of December 1989. In the context of an unprecedented evolution of multilingualism, experts in various fields, as well as higher education professors identify specific communication needs and potential users and foresee preferential contexts of dictionary use. This article presents the outcomes of a study on the main categories of users and contexts of use of multilingual dictionaries edited in Romania after 1989. This analytical approach targets a representative sample of dictionaries selected on the basis of three criteria imposed by the research limits: temporal (dictionaries edited in the period 1990–2010), idiomatic (works presenting the list of entries in Romanian) and formal (the multilingual character stated explicitly in the title of the dictionary). The first part of the article accounts for the stage of research related to dictionary users and a short presentation of Romanian lexicography in general and of multilingual lexicography in particular. The second part of the study describes the methodology and proposes a systematization of the categories of potential users and the contexts of dictionary use as envisaged by the authors of the dictionaries under analysis.

Keywords: ROMANIAN LEXICOGRAPHY, MULTILINGUAL DICTIONARIES, USERS, CONTEXTS OF DICTIONARY USE, PRAGMATICS

Résumé: Utilisateurs et Contextes d'Emploi des Dictionnaires Multilingues Roumains. En Roumanie, après la Révolution de décembre 1989, un nombre impressionnant de dictionnaires multilingues (dans au moins trois langues) fut publié. Dans les conditions de l'évolution sans précédent du multilinguisme, les spécialistes des différents domaines et les professeurs universitaires de langues étrangères captent des besoins de communication spécifiques, identifient des potentiels utilisateurs et préconisent des contextes préférés d'emploi des dictionnaires. Cet article présente les résultats d'une étude sur les principales catégories des utilisateurs et des contextes d'emploi des dictionnaires multilingues édités en Roumanie après 1989. La démarche analytique vise un échantillon représentatif de dictionnaires, choisi à partir de trois critères imposés par les limites de la recherche: le critère temporel (les dictionnaires édités dans la période 1990–2010), le critère idiomatique (les travaux qui présentent la liste des entrées en roumain) et le critère formel (le caractère multilingue clairement présenté dans le titre des dictionnaires). La première partie de l'article contient l'état de la recherche relatif aux utilisateurs des dictionnaires et une brève présen-

tation de la lexicographie roumaine, en général, et de celle multilingue, en particulier. La deuxième partie de l'étude présente la méthodologie utilisée et propose une systématisation des catégories des utilisateurs potentiels et des contextes d'emploi préconisés par les auteurs des dictionnaires analysés.

Mots-clés: LEXICOGRAPHIE ROUMAINE, DICTIONNAIRES MULTILINGUES, UTILISATEURS, CONTEXTES D'EMPLOI, PRAGMATIQUE

1. Introduction

The main purpose of lexicography has always been to provide an inventory of and to explain the lexis of a language, be it a general lexis or a specialized one, for the speakers of that specific language or for the speakers of other languages. With the development of pragmatics, there occurs a shift of focus from the study of language in terms of the relationships between linguistic signs towards the relationships between verbal signs and their users. The development of pragmatics impacts on both linguistic studies, as the research on any type of text (dictionaries included) involves concepts specific to pragmatics, and the manner in which texts are elaborated. The authors and/or editors of dictionaries (regardless of them being lexicographers or not) place increased emphasis on the 'user's perspective' when elaborating dictionaries. In the texts that constitute the prefaces of dictionaries generally three main strategies that prove a transition of focus from the lexis to the users are employed: (i) the identification of the users' needs in relation to actual social realities (e.g. the emergence and development of new socio-professional categories) and the manner in which the content is organized depending on the aforementioned needs, marked by an ever increasing influence of the publishing houses upon the authors of dictionaries; (ii) the analysis of the profile and the assignment of the target readers/users; (iii) an invitation to the users to participate in the lexicographic approach (not only in view of the elaboration and publication of the dictionary, but also with regard to the impact and the improved quality of lexicographic products). Each of these strategies will be analyzed from the speech acts theory perspective (locutionary, illocutionary, perlocutionary), in order to emphasize the manner in which dictionary authors adapt to the users' needs and socio-cultural profiles.

Taking into account the fact that the value of a dictionary also resides in the manner in which it manages to relate the lexicographic data to the specific needs of the target audience, this paper proposes an analysis of Romanian multilingual dictionaries issued in Romania after the Revolution of 1989, from the perspective of the categories of users targeted by these dictionaries and the contexts of dictionary use anticipated by the authors of the respective dictionaries. The study is organized as follows: the first part includes a review of the meta-lexicographic research directions, with a special focus on user-oriented studies, as well as a review of the history of Romanian lexicography, whereas

the second part presents the methodology and the outcomes of the analysis of multilingual dictionaries.

2. State of research

The elaboration of dictionaries is motivated by the identification of practical necessities among potential users, while the normative character and the authority of such works raise the issue of the responsibility of lexicographic approaches, since: the dictionary was a resource of what was considered to be the best or correct language (Hartmann and James 1998: VII). However, the authors of dictionaries started only in the 20th century to shift the focus of their interest towards a 'user-oriented perspective'. 'One of the principal advances in lexicography in recent years has been the focus on the user perspective, that is the realisation that different users have different reasons for using a dictionary, and that the dictionary can, and should, respond to these' (ibid. VIII).

The digitization of lexicography challenges authors to resort to a deep restructuring of the process involved in the elaboration of dictionaries. According to Sue Atkins, the real challenge is not 'how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new' (Atkins 1996: 516). The storage and consulting of dictionaries by the users are also subject to important changes following the digitization of lexicography, which allows users to build up 'digital libraries' of resources. From this perspective 'the future dictionary is rather an integrated tool or a number of tools in a professional user's toolbox where it coexists with other language technology products such as encyclopaedic sources of reference, different types of corpora, corpus analysis tools [...], as well as corpus compilation software, translation memory systems, etc.' (Varantola 2002: 35).

The users themselves can become authors of dictionaries, leading to the rapid expansion of 'collaborative lexicography' where dictionary information is created and edited by users (Nesi 2016: 580).

Dictionary user research 'takes account of the profiles of dictionary users; the various contexts of dictionary use; the functions of the dictionary in its many situations of use; and the skills necessary for, and brought to bear on, dictionary use' (Hartmann and James 1998: VIII).

Two of the aspects analyzed in user-oriented studies relate to the issues the users have to cope with, and the systematization of user categories and contexts of use. Dictionary users face numerous challenges for which various solutions are proposed. On the one hand, there are authors who, despite targeting a larger audience, end up producing a dictionary destined to a specific segment of audience. In such cases, the following aspect needs to be considered: 'the dictionary maker naturally has to take into account all potential users and pay special attention to the largest groups of users and buyers' (Varantola 2002: 30-44). On the other hand, there are dictionaries with structures that render the task of consulting them rather difficult for the user, who fails to access and interpret

the data that is provided (Pricop and Mocanu 2019: 130). The suggested solution is to provide 'a model of dictionary usage or a model for the decoding of lexicographic definitions' (Bidu-Vrănceanu 1993: 6). It is thus necessary to 'place more responsibility on the dictionary makers who, as professionals, should have a holistic view of what their dictionary offers to its users' (Kalliokuusi and Varantola 1998: 601-610).

Literature distinguishes between three main categories of users: language learners, non-professional users and professional users (Varantola 2002: 30-44). Another classification of users, this time only in relation to explanatory monolingual dictionaries, distinguishes between: '(i) readers referred to as 'cultivated', whose mother tongue is the described language; these readers are proficient users of the language; (ii) readers with imperfect skills in their mother tongue, whose aim is to enhance their competence and language skills and (iii) users of a pedagogical profile (the users are always teachers/pupils/students) who are only supposed to focus on the information that is pertinent for systematic learning from a determined perspective' (Bidu-Vrănceanu 1993: 6-7). Literature on the topic also emphasizes the variety of contexts of dictionary use: 'Dictionaries can thus be used in various ways in language teaching and language learning but also consulted as tools in "non-learning" activities, for example in reading comprehension, text production and in professional translation' (Varantola 1997: 30-31). According to the functional theory of lexicography, there are two categories of contexts of dictionary use: 'there are two main groups of use situations: cognition and communication-oriented situations' (León Araúz et al. 2008: 998).

3. Multilingual Romanian lexicography

In Romania, the production of multilingual dictionaries occurs in two significant stages: in the 17th–19th centuries dictionaries are produced to fulfill communication needs, while in the 20th–21st centuries multilingual terminological dictionaries are issued to serve professional needs. During the first stage, a traditional line becomes visible, which gradually fades by the end of the 19th century:

The real practical needs that have led to the emergence and development of national lexicography explain to a large extent the geographic distribution of the issued works, as well as their internal composition (first of all, the constituent languages) (Seche 1966: 181).

In the last decades of the 19th century, the Romanian traditional plurilingual lexicography was on the verge of irremediable decline, due to the obvious lack of practical usefulness of the works it had produced. As they cumulated several bilingual dictionaries and consequently became difficult to consult, polyglot dictionaries were eventually abandoned in favor of bilingual dictionaries proper (ibid. 1969: 317).

Starting with the second half of the 20th century, the number of Romanian multilingual dictionaries 'increases constantly in order to meet professional needs' (ibid. 318).

The first known multilingual dictionary in Romania is a lexicon in Latin (the source language), Romanian and Hungarian (the target languages), dating from 1687–1701. Subsequently, in the second half of the 18th century, a series of multilingual lists of names of plants are issued (for the first one, dating from 1783, the source language is still Latin), along with manuscript vocabularies in which Romanian is placed in relation with Greek, French, German and Russian (languages whose culture influenced the Romanian culture at the time), as well as a series of multilingual dictionaries elaborated by the representatives of the Transylvanian School. The ever-increasing interest for the elaboration of a general dictionary of the Romanian language becomes, at the end of the 18th century, one of the main desiderates of the Transylvanian School. However, since the communication needs of the Transylvanian territory at the time, between different ethnic groups (Romanian, Hungarian, German) made it practically impossible to achieve such desiderate, multilingual dictionaries are produced instead. The first explanatory and etymological dictionary of the Romanian language ever printed is *Lexiconul de la Buda* [*The Buda Lexicon*], issued in 1825. However, the plant names glossaries remain dominant up to the end of the century, representing a constant feature of Romanian multilingual lexicography. In the 20th century, due to the progressive development of science and various technical fields, a significant number of multilingual terminological dictionaries are issued, especially by the Technical Publishing House. The best such dictionary from a lexicographic point of view was *The Romanian Technical Lexicon*, issued in the mid-20th century. At the beginning of the 21st century, of the terminological multilingual dictionaries we should mention at least almost 50 multilingual dictionaries of a normative character belonging to an academic series: *Explanatory Dictionaries for Science and Technology* and *Explanatory Dictionaries for Exact Sciences*.

We should also emphasize the fact that Romanian multilingual dictionaries have a perceived documentary value rather than a linguistic, scientific one. Some works stand out indeed (*The Buda Lexicon*, according to Mocanu 2019: 170), yet not due to their multilingual component, but rather due to their etymological and/or explanatory component, representing different stages or attempts aimed at the elaboration of The Romanian Language Dictionary. Although during the 20th century the Romanian lexicographic technique improves, multilingual terminologies are still not dealt with from a linguistic perspective, but rather as inventories and parallel term lists. Very few exceptions can be noted along with a few dictionaries aimed at the learning of foreign languages and which are treated somehow from a linguistic perspective, as they were elaborated by philologists specialized in foreign language teaching. The development of digital technologies led, in Romania as well, to the development of electronic multilingual dictionaries and terminological databases. Hence the

decrease of the authors' and editors' interest in the elaboration and publication of terminological dictionaries, multilingual dictionaries included, a phenomenon that started around the year 2010.

3.1 Multilingual dictionaries after the Revolution of December 1989

For Central and Eastern European countries, the year 1989 was a year of crucial changes, the transition from various types of communism towards democracy occurring differently for each different country. The Revolution of 1989 triggered substantial political, social, economic, cultural, and scientific changes. From a social perspective, an increasing mobility of the population from Romania towards Western Europe and from Europe towards Romania can be observed from one year to another, generating linguistic contacts and a constant need for an accelerated acquisition of foreign languages. From an economic and cultural viewpoint, important changes can be noted with regard to the editorial market, influenced by the increasing number of private publishing houses. The products recording the biggest sales are now dictionaries, regardless of their type. Some of the publishing houses turn into private businesses, forming editorial groups with their own distribution services, while others stick to the structure of the National Culture Press and Publishing House and subsequently become self-financed public publishing houses, supported by the relevant ministries.

The universe of books is undergoing profound changes [...] in all its aspects, editorial life is complex and diverse, the book offer is adapted according to the readers' demands; however, based on governmental policies and subsidies, cultural programmes of a national interest that could not be supported without the state's intervention are also targeted (Buluță 1996: 77).

However, the private initiative dominates the Romanian editorial market, where the market economy laws and the Western editorial model are already in force. As far as the scientific field is concerned, information from various fields are absorbed rapidly and massively, contributing thus to the formation and development of various terminologies in the Romanian language.

With regard to dictionary editing, the first decade following the Revolution of 1989 is marked by a gap between the dynamics of lexicographic production (the number of lexicographic works increases significantly every year) and the absence of coherent editorial policies at a national level (according to Mocanu 2017). During the first decade (1990–1999) after the Revolution, 86 multilingual dictionaries are published in Romania, as compared to only 27 multilingual dictionaries in the previous decade. The quality of lexicographic editions is also subject to many changes: whereas before 1989 most dictionaries were edited under the patronage of state institutions (various Ministries, the Romanian Academy, higher education institutions, and the like), in the following years the editing of books is liberalized and the access to printing becomes

readily available. This aspect is reflected in the heterogeneous character of Romanian lexicography after 1989: academic, professional editions co-exist with dictionaries drafted by specialists in various fields with no specialized lexicography training, an aspect which impacted on the quality of the lexicographic works issued during that period.

4. Methodology

4.1 Object of study and sources

The object of our study refers to the categories of users and the contexts of dictionary use considered by multilingual dictionary (with at least three component languages) authors who published their works in Romania after the socio-political events in 1989 known as 'the Revolution of December 1989'. The framework of our research requires the analysis of relevant data from a limited, but representative sample of multilingual dictionaries issued after 1990. The research aim refers, on the one hand, to the identification and systematization of the main categories of users and on the other hand, to the identification and (brief) description of contexts of dictionary use. According to the data provided by the first and only inventory of Romanian multilingual dictionaries (Pricop et al. 2017: 175-339), in the period 1990–2016, 303 such works were drafted and published, of which 86 titles were issued in the period 1990–1999, 202 titles in the following decade and 25 titles in the period 2011–2016. Taking into account the fact that for the decade that ended with the events in 1989, namely the period between 1980 and 1989, there are 27 titles (Pricop et al. 2017: 160–174), one can note a lexicographic 'boom' of a multilingual type in the period 1990–2010, with an obvious peak between 2001–2006 and a visible decline after the year 2010 (see Figure 1).



Figure 1: An outlook on Romanian multilingual dictionaries

A pertinent explanation for the impressive number of dictionaries (no less than 288) issued during the most prolific decades for the Romanian multilingual lexicography (1990–2010), is to be offered by the outcomes of the analysis carried out in this study. The selection of a representative sample for the analysis is made according to two (meta)linguistic criteria: (1) Romanian as the source language, respectively (2) the multilingual character rendered explicitly in the titles of the dictionaries. The status of the Romanian language varies within multilingual dictionaries: in the vast majority of these works, Romanian is the source language, yet there are some dictionaries in which Romanian is one of the target languages — the main word list being rendered into a foreign language (this was mainly the case at the beginnings of Romanian multilingual lexicography, when the main language was either Latin or Hungarian, while in recent times this is the case of the dictionaries of informatics, where English is the source language). The selection of a dictionary word list consists in a complex, often original process in which the user's perspective is of utmost importance. Consequently, we have regarded the selection of the list of entries at the Romanian language level as a core criterion in the selection of dictionaries used in our analysis. With regard to the 288 dictionaries published between 1990 and 2010, we have noted that for 221 works the list of entries is drafted in Romanian, whereas for the remaining 67 titles, the source language is a foreign language (Latin, English, Russian, French or Greek), in a ratio of approximately 3:1.

The monolingual or bilingual character of a dictionary is almost never rendered explicitly in the title, since the language or languages used are usually specified (e.g. *Romanian–French Dictionary*). In the case of multilingual dictionaries, a brief analysis of the titles (Pricop et al. 2017: 9-30) reveals two modalities used for indicating the multilingual character: the use of an explicit determinant (*multilingual*, *polyglot*, *plurilingual* or *trilingual*) or the enumeration of the component languages (to which the determinant *polyglot* is sometimes added). These specifications with regard to the type of dictionary and its polyglot nature reflect the clear intention of the author and inform the user about the content of the work, thus facilitating the selection of a certain dictionary according to the user's needs. As far as the 288 dictionaries issued in the period 1990–2010 are concerned, 73 titles render the multilingual character in an explicit manner, while in the remaining 215 titles this is either implicit (the component languages are listed) or is not mentioned at all (the field is the only indicated element). Out of the 288 dictionaries issued in the period 1990–2010, just 36 multilingual dictionaries display the two selection markers concomitantly (Romanian as the source language and the polyglot nature rendered explicitly, see Figure 2).

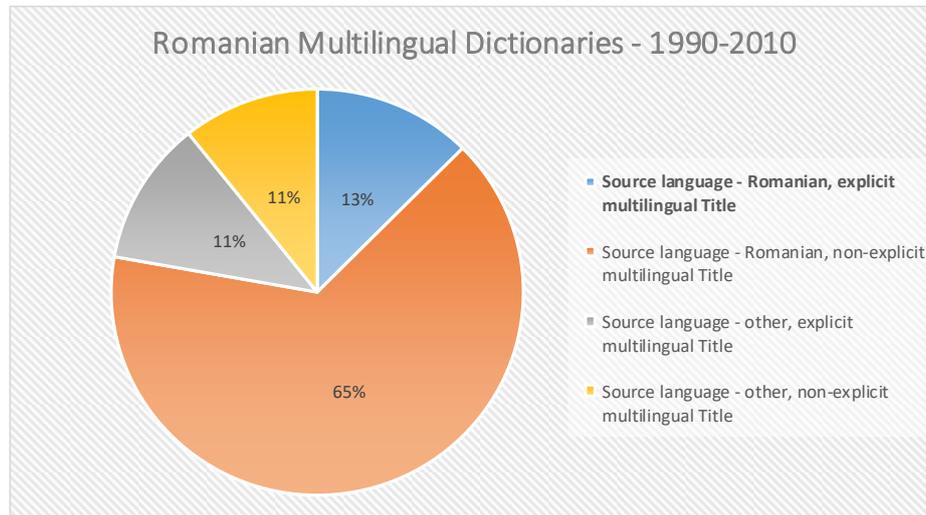


Figure 2: Romanian multilingual dictionaries between 1990–2010 (source language and title)

4.2 Documentation and filing

The multilingual dictionary, with its specific frame structure (body of the dictionary and indices) is usually preceded, in pragmatic terms, by a 'discourse of the emitter' made of various introductory texts (prefaces, indications for use, etc.) signed by the authors and/or the editors (according to Van Campenhoudt 1997). The introductory texts usually specify the target readers — the users, the actual argument of the elaboration — and their communication and/or professional needs, namely the contexts of usage, that usually represent the answers to the following questions: For whom?, Why?, To what purpose?. In order to identify, systematize and describe the users and the contexts of dictionary use we have extracted from the introductory texts of the 36 selected dictionaries various types of information regarding: the way in which 'receivers are designated' (in pragmatic terms), users and contexts of dictionary use, editorial needs and offer, 'requests for answers', acknowledgements addressed to collaborators in different stages, clarifications regarding previous works or editions. The basis of our analysis consists mainly of a corpus of introductory paragraphs and fragments that concentrate information regarding users and contexts.

4.3 Analysis and results

The analysis of the data provided by authors and/or editors is conducted in

several stages: the identification of the categories of users and contexts of dictionary use, the extraction of systematization criteria, the classification of categories according to the extracted criteria, and the setting of (proto)types. In pragmatic terms, dictionary authors and editors can be assimilated to the sender, while the users can be assimilated to the receiver. The analysis of the 'sender's discourse', rendered in the introductory texts (prefaces, forewords, instructions for use and the like) of multilingual dictionaries identifies elements that are similar to the three components of speech acts: locutionary, illocutionary and perlocutionary (according to Ionescu-Ruxăndoiu 2003: 22-23). The locutionary dimension of the discourse involves, also in the case of dictionaries, two basic operations: 'the designation of the receiver' and the 'request for an answer'. The designation of the receiver, identified in most of the works we analyzed, is achieved by means of various formulae that contain intentions, representative for the 'illocutionary' dimension (offers, promises, recommendations). Thus, such a work 'addresses' (Biriş 2000, Iliescu et al. 2001, Frâncu 2002, Frâncu et al. 2003, Vlad et al. 2004, Surdu and Surdu Soreanu 2009), 'is destined to' (Dinu 2000, Tetean and Vințeler 2002, Becea 2003, Hanga et al. 2008) or 'is made available for' (Ioniță and Marin 2002) a certain user. The dictionary 'is a useful and very important instrument for' (Țarcă 2009), 'can be extremely helpful for' (Becea 2003), 'can be used by' (Frâncu et al. 2003, Șelaru et al. 2004), 'aims at offering assistance to' (Frâncu et al. 2003), 'answers the needs of' (Dincă et al. 2004), 'aims at assisting' (Grozavu and Kocsis 2005). The authors formulate claims such as: 'we recommend' (Plahteanu 2005) or 'we hope it to be useful especially to' (Corniță, 2007). The introductory texts of a few analyzed works do not contain any such formulae (Butnaru et al. 2002, Dvoracek 2002, Manoliu et al. 2002, Cosma and Cosma 2005, Lungu et al. 2006, Lupchian and Lazăr 2010). Our analysis indicated the existence of two categories of multilingual dictionaries: dictionaries in which the target users are explicitly designated in various introductory sections and dictionaries in which the target users are not designated. In most introductory sections, the receiver is required to accept an active part, contributing with improvement suggestions for prospective future editions (Firuță and Popa 1992, Sopa 1993, Chira and Iliescu 1995, Chira 1999). In the prefaces of their works, the authors of future editions thank all those who offered suggestions for improvement, a clear expression of the performance of a perlocutionary act, when the effect coincides with the authors' intention. Moreover, the authors of most dictionaries we analyzed also address the people who already contributed, in one way or another (translation, revision, editing, printing, etc.) to the issue of the respective work.

The categories of users regarded as target audiences by the authors of the dictionaries we analyzed can be systematized based on four criteria: the numerical criterion (one/several categories aimed at by a dictionary), the lexicological criterion (the nature of the lexis recorded in the dictionary), the professional criterion (the measure in which users are conditioned by their professional background) and the idiomatic criterion (the degree of proficiency in the

component languages). The main contexts of dictionary use anticipated by dictionary authors can be systematized as such: the translation of specialized texts, research using foreign specialized literature, learning a foreign language, communication in certain situations, and recording and listing of Romanian terminology in a specific field (the main lexicological and lexicographic approach). The results of the analysis: the systematization of the categories of users and the description of the main contexts of dictionary use, rendered in two distinct chapters, with a specific emphasis on the results that are relevant for most dictionaries.

5. Categories of users

The Romanian literature in the field does not include any studies on the users of multilingual dictionaries or any classification of these users, except for a typology of monolingual dictionary users (Bidu-Vrăncianu 1993: 6-7).

5.1 The numerical criterion

Our analysis has revealed that whereas the authors of a few dictionaries are restrained to a single category of users, most authors designate several categories of users, while some others choose not to mention any category of users whatsoever (Butnaru et al. 2002, Manoliu et al. 2002, Lupchian and Lazăr 2010). Thus, the authors of a few dictionaries state that they consider a single category of users: students (Nagy 1991), professionals working in a certain field (Chira and Iliescu 1995, Savin 1997), specialists in a certain field (Mihalciuc et al. 1995) or tourists (Firuță and Popa 1992). Most dictionaries target at least two categories of users: specialists and translators (Dinu 2000, Vlad et al. 2004); specialists and students (Hanga et al. 2008); specialists and people with an interest in the field (Frâncu 2002, Surdu and Surdu Soreanu 2009); students and people with an interest in the field (Mănăilă et al. 1995); specialists and people interested in related fields (Șelaru et al. 2004: 3); specialists and professionals in a certain field, students (Biriș 2003); specialists, students and translators (Plahteanu 2005); specialists, students, translators, and foreign languages learners (Frâncu et al. 2003); specialists, university professors, students and people with an interest in the field (Dincă et al. 2004); specialists, university professors, students and decision-makers (Grozavu and Kocsis 2005). According to the systematization based on the numerical criterion, the dominant type of multilingual dictionary is the one targeting several categories of users.

5.2 The lexicological criterion

A smaller number of multilingual dictionaries record either the basic vocabulary of the Romanian language (Iliescu et al. 2001), or a nucleus of the basic vocabu-

laries belonging to several languages (Firuță and Popa 1992, Cuza 1994, Tetean and Vințeler 2002, Corniță 2007). The targeted users are foreign language learners: either foreigners who learn Romanian or Romanians who learn a foreign language. The users can be professionally determined (pupils, students, teachers) or not.

Most of the dictionaries we analyzed list terminologies of some specific fields. Being tools that aim at facilitating the access of the audience to the corresponding terms in a foreign language, most multilingual dictionaries we analyzed target, first and foremost, the specialists, indicating the respective fields: medical sciences (Nagy 1991), car engineering (Sopa 1993), Physics and Mathematics (Mănăilă et al. 1995), Economy (Mihalciuc et al. 1995, Frâncu et al. 2003), wood industry (Stoian 1995), hydrology, water management, meteorology, the protection of the environment (Savin 1997), agricultural technique and food industry (Biriș 2000), zootechnics (Dinu 2000), skiing (Becea 2003), forestry (Dincă et al. 2004), hunting (Șelaru et al. 2004), ethology and animal physiology (Vlad et al. 2004), the environment and environmental sciences (Grozavu and Kocsis 2005), machine construction (Plahteanu 2005), poultry industry and the industry of poultry-based products (Surdu and Surdu Soreanu 2009). Some of the works indicate the profession as such: the staff of the Ministry of Domestic Affairs (Ioniță and Marin 2002), musicians (Cosma and Cosma 2005), veterinarians (Lungu et al. 2006), legal experts (Hanga et al. 2008). According to the systematization based on the lexicological criterion, the dominant type of multilingual dictionary is the dictionary addressed to specialists in specific fields.

5.3 Degree of professional determination

The category of professionally determined users (see supra 5.2.) is targeted by terminological dictionaries and falls into two sub-categories: users involved in the educational system (pupils, students, teaching staff) and users who have various degrees of specialization in a certain field, with the specification of the specializations and/or fields of activity. Student users fall, in turn, into two sub-categories: students from a certain specialization, who acquire specialized multilingual terminology and students who, regardless of their specialization, are interested in enhancing their idiomatic skills in various foreign languages. The mainly didactic purpose of multilingual dictionaries is emphasized by the authors in the prefaces to their works. Thus, some dictionaries are conceived as auxiliaries for certain courses, guidelines for practical works, manuals or other didactic/teaching materials (Nagy 1991). We should note that the term 'specialist' is used both with a broader meaning and a specific one, as the authors state that in their view the specialists are people who work in research, people who work in research, planning, development, production, sales or people who work in research, translation, interpreting and education. Another professional category is represented by translators who mainly have a philological formation and training and who translate and interpret texts to and from foreign lan-

guages (Mănăilă et al. 1995, Mihalciuc et al. 1995, Stoian 1995). Besides the categories determined according to their educational or professional background, there are also undetermined categories of users who have particular interests: people who are interested in a specific field or people who are interested in a related field (Frâncu 2002, Dincă et al. 2004, Şelaru et al. 2004), people who want to carry out a conversation on topics related to a specific field (Constantinescu et al. 2003), people who are interested in learning foreign languages (Constantinescu et al. 2003), self-taught people (*Manual* 1992). According to the systematization based on the degree of professional determination, the dominant type of multilingual dictionary is the one addressed to those users whose level of studies and professional status are determined.

5.4 The language proficiency in the component languages

In principle, the multilingual dictionaries elaborated in Romania mainly target people who know the language, the Romanians, as (native) speakers of Romanian (regardless of their professional determination) (Mihalciuc et al. 1995). The Romanian audience is considered from a geo-political or touristic perspective: (future) members of the European community (Dvoracek 2002, Țarcă 2009), tourists (Firuță and Popa 1992, Stănciulescu et al. 1998) or as communicators, a recent profession listed in the official classification of professions (Corniță 2007). There are also the users who are familiar with one or several target languages, the foreigners who do not speak Romanian, but intend to learn it while learning a particular terminology (Nagy 1991). A small number of dictionaries target both Romanian and foreign users with regard to elementary skills (beginner level) in either a foreign language or Romanian (Firuță and Popa 1992, Iliescu et al. 2001, Tetean and Vințeler 2002). The authors of some other dictionaries address explicitly the translators who interpret and translate texts, and native users of Romanian who can speak at least two of the component languages (the source language and a target language) of the respective works (Frâncu et al. 2003, Vlad et al. 2004). According to the systematization based on the criterion regarding the language proficiency in the component languages, the dominant type of multilingual dictionary is addressed to the users who are familiar at least with the source language, Romanian in this specific case.

6. Contexts of multilingual dictionaries use

The Romanian literature in the field does not include any studies on multilingual dictionaries use or any attempt at systematizing these dictionaries.

6.1 The translation of specialized texts

One of the main categories of users targeted by multilingual dictionaries is rep-

resented by translators. The use of multilingual dictionaries is mandatory for the accurate translation of the technical and scientific documentation accompanying machinery and installations in agriculture and the food industry (Biriş 2000: 3), economic texts (Frâncu et al. 2003), terms belonging to biotechnology (Manoliu et al. 2007: 3) or the Romanian legal terminology in the European context (Hanga et al. 2008).

6.2 Documenting based on foreign literature and the proper understanding of multilingual terminology

The monolingual explicative dictionary is necessary in the development stage of a terminology, according to the authors of a zootechnics dictionary (Dinu 2000: 6). Considering the tendency of 'internationalization' of the terminologies, a stage where there might occur confusions or inappropriate uses of zootechnical notions (ibid.), animal ethology and physiology notions (Vlad et al. 2004: 5) or musical notions (Cosma and Cosma 2005: 5), making use of multilingual dictionaries becomes a necessity. The use of a multilingual dictionary is necessary to access and use foreign literature, in order to train specialists from particular fields (medical biophysics — Nagy 1991, skiing — Becea 2003, forestry — Stoian 1995, Dincă et al. 2004). The development of a multilingual dictionary in an international project, with native editors of several European languages, aims at:

... introducing a common language that will facilitate the educational process, the research activity, and the legislation, etc. in the environmental field, contributing to supporting the policies and initiatives of the EU Member States and the candidate countries (Grozavu and Kocsis 2005: 6).

In a multilingual illustrated dictionary, the illustrations are not only explanatory (the image renders the proper meaning of a term), but rather play the part of a 'universal intermediary between different languages' (Plahteanu 2005: 3), contributing to the dissemination and wider use of information and multilingual terminology in the field. The selection of terms in a dictionary of aviculture and poultry product industry is dictated by practical needs: 'to help identify quickly and precisely the terms used in specialized periodicals and books published in the three important foreign languages' (Surdu and Surdu Soreanu 2009: ix).

6.3 Acquiring a foreign language

Acquiring a foreign language represents a didactic goal that can be achieved by including multilingual dictionaries in the learning process (with bilingual ones regarded as more appropriate). Two types of dictionaries are noted in the analyzed corpus sample: those that contain the basic vocabulary of some languages and are addressed directly to the Romanian (or foreign) learners of foreign lan-

guages, respectively a series of terminology dictionaries, aimed at foreign language learners, although their main category of users is represented by specialists in the field. The multilingual dictionary for foreign language learning may contain the basic vocabulary of the Romanian language being primarily dedicated to foreigners who learn Romanian (and secondly to Romanians who are thus provided in a single volume the translation of the basic vocabulary into a foreign language, Iliescu et al. 2001), or it may contain a balanced and representative basic fund for both Romanian and foreign languages, addressing a broad category of users (determined according to their educational or professional background: Romanian language learners and students who learn foreign languages according to the Romanian curriculum, teachers teaching either foreign languages or Romanian as a foreign language at all levels, Tetean and Vințeler 2002, or indefinite: Romanian tourists, Firuță and Popa 1992).

6.4 Acquiring multilingual terminology in a specific field

A multilingual dictionary of medical biophysics terms is intended for limited use: namely the students from the biophysical specialization of the University of Medicine and Pharmacy of Timișoara (Nagy 1991). Another dictionary of pathological anatomy is conceived as an individual learning instrument for students at the Faculty of Veterinary Medicine to 'rapidly acquire an adequate medical vocabulary' that is at the same time multilingual (Lungu et al. 2006). The use of a dictionary of legal terms is necessary for the students of the Faculty of Law to acquire specialized terms in other languages (Hanga et al. 2008).

6.5 The development of communicative skills

Generally, dictionaries are tools that support communication between speakers of different languages; in some cases, this is the main objective of a dictionary. Certain multilingual dictionaries, though apparently terminological, compile inventories of a varied lexis encompassing both common terms specific to a certain field or to related fields and words used in ordinary conversational exchanges. Such dictionaries explicitly support the communication purpose, the dialogue between specialists or between specialists and the general public, on topics related to a specialized field. This is the case of a professional lexis addressed to the staff of the Ministry of Domestic Affairs, meant to support the dialogue with the general public or colleagues from similar structures belonging to the European and North American countries (Ioniță and Marin 2002: 2). It is also the purpose of a medical dictionary that 'responds to the needs of an immediate medical conversation', a useful tool 'not only for medical specialists, but also for anyone who wants to discuss medical matters in French or in English' (Constantinescu et al. 2003: 5).

Other such communicative situations are: exchange of information between the participants to road traffic, during motor shows and auto fairs, respectively dialogues involving tourists or Romanian specialists engaged in international relations (Sopa 1993: 5; Firuță and Popa 1992: 5; Chira and Iliescu 1995: 7; Mihalciuc et al. 1995: 5; Şelaru et al. 2004: 3; Cosma and Cosma 2005: 5). A special dictionary is the one dedicated to people working in different institutions (educational institutions included) and who are permanently or occasionally invested with the role of communicators (Corniță 2007: 15).

6.6 The 'thesaurization' of Romanian terminology

This aspect refers to a series of dictionaries with a secondary multilingual character, which focus on the delimitation and explicitation of the Romanian terminology in the field. In this respect, a dictionary of genetics is meant to provide its users with explanation for specific terms, so that these terms can be used with their real meaning, the definitions of the entries (some concise, others detailed) being rendered exclusively in Romanian (Butnaru et al. 2002). In this case, the dictionary is subject to 'a different type of integration in communication, as terms are provided synonyms in the main languages used at the international level' (Butnaru et al. 2002: i). The use of a library science-related thesaurus responds to the needs of indexing and retrieving the information contained in the documents belonging to Romanian small and medium libraries with encyclopaedic funds. In this particular case:

... the presence of terms in the other two participating languages, namely English and French, adds value to the instrument and widens the scope of its use (Frâncu 2002: i).

One must add that other dictionaries, while emphasizing the multilingual terminology they contain, equally preserve and store the Romanian terminology in the field (the Romanian lexical inventory of some previous editions: economic, tourism or foreign trade terminology — Frâncu et al. 2003, skiing terminology — Becea 2003, forestry terminology — Dincă et al. 2004).

7. Conclusive remarks

The purpose of practical lexicography consists in the recording and explicitation of various types of lexis for various users, yet the user's perspective becomes the focus of dictionary authors' interest only starting with the 20th century. User-oriented studies propose various classifications of the categories of users and contexts of dictionary use. Romanian lexicography has a history that spans over five centuries, being dominated by bilingual and multilingual dictionaries; we should mention that Romanian monolingual lexicography, on the other hand, is only one century and a half old. This study represents the

first analytical approach based on information selected from the introductory texts of a representative sample of 36 multilingual dictionaries. The outcomes of the study are presented in two distinct chapters that deal with the systematization of the categories of users and the description of contexts of dictionaries use. Our analysis has indicated that the dominant type of multilingual dictionary targets several categories of users, determined from a professional viewpoint, proficient in Romanian and interested, due to their professions, in multilingual terminologies. The dictionaries that made the subject of our analysis are elaborated in order to serve various contexts: the accurate translation of foreign terminologies and of specialized texts, the acquisition and accurate use of a certain terminology, research using foreign specialized literature, learning a foreign language, the elaboration of dictionaries of specialized terms. Most dictionaries aim at facilitating communication, between Romanian and foreign specialists in various communication situations (information exchange, common interests, business partnerships, and others). This study emphasizes the huge need for rapid knowledge of foreign languages (as many as possible) that serve an immediate professional interest which the authors of multilingual dictionaries (including the ones we analyzed) were able to identify. The dictionaries we included in our sample represent the lexicographic expression of the social phenomenon triggered by the Revolution of December 1989 in Romania, which led to the opening of the country's borders towards Europe. This study represents a contribution to the field of Romanian metalexicography. The originality of the analytical approach derives from: (1) the nature of the corpus we analyzed (multilingual dictionaries issued in the first two decades after the Revolution of 1989) and (2) the pragmatic approach perspective, focused on categories of users and contexts of dictionary use. The analysis of the lexicographic corpus has emphasized, on the one hand, the authors' interest in the profile and more and more diverse needs of the categories of users and on the other hand, the need to adjust the lexicographic approach according to the specificity and particularities of the target audience, in order to produce high-quality lexicographic works.

References

Dictionaries

- Becea, L.** 2003. *Dicționar poliglot de schi*. București: Editura Printech.
- Biriș, S.-Ș.** 2000. *Dicționar tehnic poliglot de mașini și instalații pentru agricultură și industrie alimentară: român, englez, francez, german, maghiar*. București: Editura Printech.
- Butnaru, G., M. Bodogai, A. Baci and R. Zehan.** 2002. *Dicționar de genetică poliglot*. Timișoara: Editura Mirton.
- Chira, V.** 1999. *Dicționar poliglot explicativ de cardiologie*. Cluj-Napoca: Editura medicală universitară 'Iuliu Hațieganu'.
- Chira, C. and M. Iliescu.** 1995. *Dicționar tehnic rutier poliglot*. Cluj-Napoca: Editura Albastră.

- Constantinescu, I., G. Godeanu and M. Botorog. 2003. *Dicționar medical poliglot: român-francez- englez*. București: Editura Garamond.
- Corniță, G. (Coord.). 2007. *Mimica. Dicționar multilingv (român, francez, englez, german, maghiar) / Mimicry. Multilingual Dictionary (Romanian, French, English, German, Hungarian)*. Baia Mare: Editura Universității de Nord.
- Cosma, M. and R. Cosma. 2005. *Dicționar multilingv de termeni muzicali*. București: Editura Universității Naționale de Muzică.
- Cuza, V. 1994. *Dicționar poliglot (engleză, franceză, germană, italiană,spaniolă, română)*. București: Editura Gramar.
- Dincă, L., D. Drăghiciu and D. Chira. 2004. *Dicționar forestier poliglot. Român- englez-francez-german. Englez-Român. Francez-Român. German-Român*. Brașov: Editura Silvodel.
- Dinu, I. (Coord.). 2000. *Dicționar poliglot de zootehnie: română, franceză, engleză, germană, spaniolă*. București: Editura Coral Sanivet.
- Dvoracek, A. 2002. *Europa în cuvinte-cheie: dictionar explicativ multilingv, român, francez, italian, spaniol*. Vol. I (A-C). Iași: Editura Vasiliana'98.
- Firuță, C. and A. Popa (Eds.). 1992. *Dicționar euro-turistic poliglot retroversiv*. București: Editura Alcor S.R.L, Editura Arta Grafică.
- Frâncu, N. (Coord.), S. Drăghici, A. Hârlăoanu and L. Șoptoreanu. 2003. *Dicționar poliglot economic, de comerț exterior și turism: română, franceză, engleză, germană, spaniolă, rusă*. București: Editura Garamond.
- Grozavu, A. and L.Ș. Kocsis (Coord.). 2005. *Dicționar poliglot multimedia de mediu și de științe ale pământului*. Iași: Editura Azimuth.
- Hanga, V., R. Hanga-Calciu and I. Trifa. 2008. *Dicționar juridic polyglot. Român- englez-francez- german-italian*. Cluj-Napoca: Editura Cordial LEX.
- Iliescu, M., C. Nedelcu, G. Scurtu, V. Neagu and A. Costăchescu. 2001. *Dicționar poliglot. Română. Engleză. Franceză. Germană. Italiană. Spaniolă*. București: Teora.
- Ioniță, T. and R. Marin. 2002. *Lexic profesional poliglot: română, franceză, germană, engleză, spaniolă, italiană*. București: Editura Ministerului de Interne.
- Lungu, A., M. Parodi, V. Grieco and G.L. Fiore. 2006. *Dicționar explicativ poliglot de anatomie patologică veterinară*. București: Editura Printech.
- Lupchian, G. and C.A. Lazăr. 2010. *Finanțe, bănci și contabilitate. Mic dicționar poliglot*. București: Pro Universitaria.
- Mănăilă, D., C. Popa, D. Popa, I.M. Popescu and V.I. Vlad. 1995. *Mic dicționar poliglot de fizică, tehnică, matematică, în limbile română, engleză, germană, franceză*. București: Editura Agora.
- Manoliu, A., L.A. Oprică and S. Diaconeasa. 2002. *Dictionar poliglot de biotehnologie: român- englez- rus-francez*. Iași: Editura Corson.
- Manoliu, A., M.E. Sidoroff, L. Oprică, S. Diaconeasa and T. Tofan-Burac. 2007. *Dicționar poliglot de biotehnologie: român, englez, rus, francez, finlandez*. Second Edition. București: Editura Tehnică.
- Manual. 1992. *Manual — dicționar poliglot în 8 limbi: română, esperanto, franceză, engleză, germană, olandeză, spaniolă, italiană*. Cluj-Napoca: S.C. Diamant de Arieș.
- Mihalciuc, M., L. Mureșan, G. Stănculescu and S.V. Stan. 1995. *Dicționar poliglot explicativ. Termeni uzuali în economia de piață*. București: Editura Enciclopedică.
- Mihalciuc, M., L. Mureșan, G. Stănculescu, S.V. Stan and L.E. Csaba. 1996. *Dicționar poliglot explicativ. 1000 termeni uzuali în economia de piață*. București: Editura Enciclopedică.

- Nagy, I.I. (Coord.).** 1991. *Dicționar poliglot cu termeni de biofizică medicală*. Timișoara: Editura LITO I.M.T.
- Plahteanu, B. (Coord.).** 2005. *Dicționar ilustrat poliglot în construcția de mașini*. Iași: Editura Performantica.
- Savin, C.** 1997. *Dicționar științific poliglot pentru domeniile: hidrologie, gospodărirea apelor, meteorologie, protecția mediului și alți termeni necesari din discipline conexe-complementare*. Engleză, franceză, germană, spaniolă, rusă. (2 volumes 1996, 1997). București: Editura Tipored.
- Șelaru, N. (Coord.), M. Georgescu, I. Svoronos, B. Ioriatti and P. Moise.** 2004. *Dicționar poliglot de vânătoare*. București: Editura Cynegis.
- Sopa, S.** 1993. *Dicționar tehnic poliglot. Motoare, automobile, tractoare*. Română. Maghiară. Engleză. Germană. Franceză. Spaniolă. Cluj-Napoca: Editura Microinformatica.
- Stănciulescu, G., N. Lupu and G. Țigu.** 1998. *Dicționar poliglot explicativ de termeni utilizați în turism*. București: All Educational.
- Stoian, O. (Coord.).** 1995. *Dicționar forestier poliglot: român, englez, german, francez în domeniile: silvicultură, exploatarea pădurilor, prelucrarea lemnului*. (2 volumes). București: S.C. Forest Infodoc S.A.
- Surdu, A.A. and R.V. Surdu Soreanu.** 2009. *Dicționar poliglot de avicultură și industria produselor avicole*. București: Editura Academiei Române.
- Tetean, D. and O. Vințeler.** 2002. *Dicționar tematic poliglot: română, franceză, spaniolă, engleză, germană, rusă*. Cluj-Napoca: Risoprint.
- Vlad, I., V. Băcilă and I. Oprea.** 2004. *Dicționar poliglot de etologie și fiziologie animală*. Sibiu: Editura Universității Lucian Blaga din Sibiu.

Other literature

- Atkins, B.T.S.** 1996. Bilingual Dictionaries. Past, Present and Future. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. Rödger Pappmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Göteborg: Göteborg University.
- Bidu-Vrâncănu, A.** 1993. *Lectura dicționarilor*. București: Editura și Atelierele Tipografice Metropol.
- Buluță, G.** 1996. *Scurtă istorie a editurii românești*. București: Editura Enciclopedică.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Ionescu-Ruxăndoiu, L.** 2003. *Limbaj și comunicare. Elemente de pragmatic lingvistică*. București: Editura All.
- Kalliokuisi, V. and K. Varantola.** 1998. From General Dictionaries to Terminological Glossaries. Fontenelle, T., P. Hiligsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *Actes EURALEX '98 Proceedings. Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 601-610. Liège: English and Dutch Departments, University of Liège.
- León Araúz, P., P. Faber Benitez and C. Pérez Hernández.** 2008. LSP Dictionaries and their Genuine Purpose: A Frame-based Example from MARCOCOSTA. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 997-1006. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Mocanu, M.** 2017. Dicționarele multilingve românești. O analiză comparativă. *Studii de Știință și Cultură* 13(3): 51-60.

- Mocanu, M.** 2019. The Dictionary as Signifying Practice. A Semiotic Approach of the Buda Lexicon. *Hermeneia. Journal of Hermeneutics, Art Theory and Criticism* 22: 157-171.
- Nesi, H.** 2016. The Demands of Users and the Publishing World: Printed or Online, Free or Paid for? Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*: 579-589. Oxford: Oxford University Press.
- Pricop, A. (Coord.), A.-G. Marcu, M. Mocanu and O. Zamfirescu.** 2017. *Lexicografia poliglotă în secolele al XVII-lea–al XXI-lea: liste, vocabulare, glosare, dicționare, lexicoane, tezaure care conțin limba română. Bibliografie*. Iași: Editura Universității 'Alexandru Ioan Cuza' din Iași.
- Pricop, Alina-Mihaela and Mihaela Mocanu.** 2019. The Design of Multilingual Dictionaries According to Users' Needs. *Philologica Jassyensia* 15: 119-135.
- Seche, M.** 1966–1969. *Schiță de istorie a lexicografiei române* [Brief History of the Romanian Lexicography]. (2 volumes). București: Editura Științifică.
- Van Campenhoudt, M.** 1997. Évaluation des terminographies multilingues: le dictionnaire nautique du capitaine Paasch face au dictionnaire aéronautique de l'ingénieur Schlomann. Hermans, A. (Ed.). 1997. *Les dictionnaires spécialisés et l'analyse de la valeur, actes du colloque organisé en avril 1995 par le Centre de terminologie de Bruxelles (Institut libre Marie Haps)*: 75-115. Louvain-la-Neuve: Peeters.
- Varantola, K.** 1997. On the Information Needs of Dictionary Users. Karpova, O. (Ed.). 1997. *Aktual'nye problemy teoretiaeskoj prikladnoj leksikografii*: 98-110. Ivanovo: Ivan.
- Varantola, K.** 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 30-44. Grenoble: EURALEX.

Sethantšo sa Sesotho and *Sesuto–English Dictionary*: A Comparative Analysis of their Designs and Entries*

T.L. Motjope-Mokhali, *Department of African Languages and Literature,
University of eSwatini, eSwatini*
(Corresponding Author, tlmotjope@gmail.com)

I.M. Kosch, *Department of African Languages, University of South Africa,
Pretoria, South Africa* (ikosch7@gmail.com)
and

Munzhedzi James Mafela, *Department of African Languages, University of
South Africa, Pretoria, South Africa* (munzh.mafela@gmail.com)

Abstract: This article seeks to establish the relationship and extent of similarity between two Sesotho dictionaries, published in the 1800s and 2005 respectively. The two dictionaries under discussion are the *Sesuto–English Dictionary* by Mabile and Dieterlen and *Sethantšo sa Sesotho* by Hlalele. The former dictionary, like most dictionaries of other African languages pioneered by the missionaries, is bilingual. The latter dictionary is the first monolingual dictionary for Sesotho and it was compiled by a mother tongue-speaker of the language. The closeness of the content of the two dictionaries is established by applying the user-perspective approach as the framework of analysis. Through an analysis of the designs and entries in the two dictionaries, the study discovers similarities and differences in terms of the use of non-standard symbols and atypical sound patterning, illustrative phrases/sentences and obsolete or archaic words. Given the amount of obsolete items in *Sethantšo sa Sesotho*, one of the recommendations emanating from this study is that *Sethantšo sa Sesotho* be revised or that a new monolingual dictionary be produced which will include more modern words that will meet the needs of contemporary users.

Keywords: BILINGUAL DICTIONARY, MONOLINGUAL DICTIONARY, LEXICAL ITEMS, COMPARISON, THEORY OF ADAPTATION, SESOTHO DICTIONARIES, SIMILARITIES, DIFFERENCES, OBSOLETE AND NEW WORDS

Opsomming: *Sethantšo sa Sesotho en Sesuto–English Dictionary: 'n Vergeelykende analise van die ontwerp en inskrywings.* Hierdie artikel poog om die verhouding en omvang van ooreenkoms tussen twee Sesotho woordeboeke, wat onderskeidelik in die

* This study is based on T.L. Motjope-Mokhali's doctoral thesis *A Comparative Analysis of Sesuto–English Dictionary and Sethantšo sa Sesotho with Reference to Lexical Entries and Dictionary Design* which was submitted to and granted by UNISA, Pretoria in 2016.

1800s en 2005 gepubliseer is, te bepaal. Die twee woordeboeke onder bespreking is die *Sesuto–English Dictionary* deur Mabile en Dieterlen en *Sethantšho sa Sesotho* deur Hlalele. Eersgenoemde woordeboek is tweetalig, soos die meeste woordeboeke van ander Afrikatale wat die pionierswerk van sendelinge was. Laasgenoemde woordeboek is die eerste eentalige woordeboek vir Sesotho en is saamgestel deur 'n moedertaalspreker van die taal. Die ooreenkoms in die inhoud van die twee woordeboeke word bepaal deur die toepassing van die gebruikersperspektiefbenadering as raamwerk van ontleding. Deur middel van 'n analise van die ontwerp en inskrywings in die twee woordeboeke stel die studie ooreenkomste en verskille vas ten opsigte van die gebruik van nie-standaard simbole en atipiese klankkombinasies, verduidelikende frases/sinne en die gebruik van ouderwetse woorde. In die lig van die aantal ouderwetse inskrywings in albei woordeboeke, is een van die aanbevelings van hierdie studie dat *Sethantšho sa Sesotho* hersien moet word of dat 'n nuwe eentalige woordeboek opgestel moet word wat meer hedendaagse woorde insluit wat aan die behoeftes van huidige gebruikers sal voorsien.

Sleutelwoorde: TWEETALIGE WOORDEBOEK, EENTALIGE WOORDEBOEK, LEKSIKALE ITEMS, VERGELYKING, TEORIE VAN AANPASSING, SESOTHO WOORDEBOEKE, OOREENKOMSTE, VERSKILLE, OUDERWETSE EN NUWE WOORDE

1. Background information

Sesotho is a language spoken in Lesotho and in the Republic of South Africa (RSA). However, there is a Lesotho orthography and a South African orthography that is used respectively by the Basotho residing in these two countries. Although the language in these countries is the same, each country retains its identity through its orthography. As a result, most prescribed and recommended Sesotho texts used in Lesotho schools are written in the Lesotho orthography. This study focuses only on the dictionaries written in Lesotho orthography and the word 'Sesotho' refers to the language used strictly in Lesotho.

Sesotho is one of the first Southern African languages to have documents penned down in writing compared to the other indigenous languages. Sesotho's strong literary traditions are seen in works such as Thomas Mofolo's novels *Moeti oa Bochabela* (The Traveller to the East) (1907), and *Chaka* (1925), as well as Mangoela's *Lithoko tsa Marena a Basotho* (A collection of praises of Basotho chiefs) (1921) (<http://www.kwintessential.co.uk/lang>). Although Lesotho orthography is older than the South African one, the development of dictionaries in Lesotho has been very slow.

The Sesotho lexicography was pioneered by the missionaries just like the lexicography of other African languages. Literature reveals that the dictionaries of African languages were particularly meant for second-language speakers and were utilised as instrumental tools for the acquisition of vocabulary. Scholars such as Awak (1990), Busane (1990), Gouws (2005), Makoni and Mashiri (2007), Nkomo (2008), Prinsloo (2013) and Otlogetswe (2013) argue that the missionaries' priority was not to develop African languages but rather to create tools

enabling them to fulfil their goals in Africa. Awak (1990: 17) states that the early vocabularies were not intended to be used by Africans but were aimed at guiding the missionaries and other Europeans who wanted to learn African languages for evangelisation purposes. Many dictionaries produced around the late nineteenth and the early twentieth centuries were therefore bilingual.

In Sesotho, like in other African languages, such as isiXhosa, the dictionaries compiled by the missionaries are still used as reliable and accessible sources (Mtuze 1992). However, these dictionaries contain several words that have become obsolete and their vocabulary is limited because many words which are currently used do not occur in such dictionaries. Africans have recently begun engaging in producing dictionaries that are geared towards the needs of their fellow Africans. It is assumed that (monolingual) dictionaries produced by mother-tongue speakers are expected to meet the needs of the mother-tongue speakers. Dictionary production has recently developed considerably in African communities, however, Sesotho dictionaries have lagged behind. The rate at which Sesotho dictionaries are produced is very slow despite the fact that Sesotho was one of the first languages to have written documents. The first Sesotho monolingual dictionary was published in 2005. When one looks at the gap between the prominent dictionary published by the missionaries in the 1800s, the *Sesuto–English Vocabulary* (1878), later revised and titled the *Sesuto–English Dictionary*, which was last edited in 1937 (i.e. the last edition of the old *Sesuto–English Dictionary*), and a new dictionary, the *Sethantšo sa Sesotho* (2005), one learns that several changes have occurred in the language. The changes were motivated by various factors such as time, technological advances, language changes, and the borrowing and creation of new words (Rundell 2008). It is therefore necessary for Basotho scholars to come together and compile Sesotho dictionaries as a group and not as individuals. The current study intends to establish the nature of the divergence or overlap between *Sethantšo sa Sesotho* and *Sesuto–English Dictionary* by analysing their designs and entries. The research seeks to determine, amongst others, if the new Sesotho monolingual dictionary has moved beyond the older one by incorporating the new words that have entered Sesotho and to establish if the dictionary meets the needs of the contemporary users.

2. Dictionaries under scrutiny

This section provides the history of the two dictionaries under scrutiny: *Sesuto–English Dictionary* and *Sethantšo sa Sesotho*.

2.1 *Sesotho–English Dictionary*

The first Sesotho dictionary had its beginning on a sailing ship from England to South Africa around 1859. Paroz (1950) records that Adolph Mabile started the

Sesotho vocabulary list during that long journey to South Africa with the assistance of his wife born at Thaba-Bosiu as the daughter of Eugene Casalis. Mabile began the Sesotho vocabulary list initially for his personal use but on his arrival at Morija, he established a printing press and published his first Sesotho dictionary in 1878 under the title *Sesuto–English Vocabulary*. Ambrose (2006: 20) argues that the dictionary was published in 1876 and not 1878. He stresses that even though most sources give the date as 1878, it looks like they confuse this year with Mabile's *Helps for to learn the Sesuto language* [sic], which was published in 1878. Mabile edited the dictionary in 1893, and after his death in 1894, Dieterlen took over. Paroz (1950) highlights that the Dieterlens (Mrs. Dieterlen included) added the names of plants to the vocabulary and were responsible for the third edition in 1904, the fourth edition in 1911 (when Dieterlen changed the title to the *Sesuto–English Dictionary*) and the fifth edition in 1917. Since then, the dictionary has had no additions to the word list. In 1937, the words in the addendum of the fifth edition were fused with the main text in the sixth edition of the dictionary. According to Ambrose (2006: 4-5):

The 7th, 8th, and 9th editions were effectively reprints of the 6th edition and should have been indicated as such by the publishers (the Morija Sesuto Book Depot) and not as new editions.

Ambrose further posits that there was, however, a true seventh edition of the dictionary by a new missionary called R.A. Paroz who observed that Sesotho is an inflected language in which both prefixes and suffixes are attached to a stem. Consequently, he reclassified the words according to their stems, i.e. a word such as *mpho* (gift) is not found under the letter /m/ but rather under /f/, which starts the stem *-fa* (give). This means that to find the word *mpho*, one has to look under the stem *-fa*. Paroz also added some new words and changed the title of the dictionary to *Southern-Sotho–English Dictionary*. The revised and reclassified edition is what Ambrose calls "the true seventh edition" (2006: 5), which was published in 1950 using the Lesotho orthography and in 1961 using the Republic of South Africa's orthography (8th edition). According to Paroz (1950), the main difference between Mabile and Dieterlen's *Sesuto–English Dictionary* and *Southern-Sotho–English Dictionary* lies in the classification of lexical entries. Whereas the first six editions of the dictionary (up to 1937) used the word approach whereby entries could be looked up under the first letter of the word (mostly in the singular form), the seventh and later editions (from 1950 onwards) followed the stem approach, whereby words carrying prefixes were classified according to the first letter of the stem as illustrated above with the entry *mpho*. The current study looks at the sixth edition (1937) but the reprint (2000) of that edition is utilised. The dictionary includes words from different subject fields such as initiation, poetry, dance, food, history and plants to mention a few.

2.2 *Sethantšo sa Sesotho*

The *Sethantšo sa Sesotho* was compiled by Batho Hlalele (a former Catholic priest) in 2005. According to Ambrose (2006), the author spent over 40 years collecting and recording in writing the meanings of words in Sesotho. *Sethantšo sa Sesotho* is a general dictionary which consists of words from various subject fields such as initiation, poetry, dance, food, history, proverbs and idioms. The dictionary uses phonemic sorting for the arrangement of its words while following the normal alphabetical order of English that is, from the letter "a" to "u". It offers detailed information regarding word categories, for example, in the case of nouns, it presents the class number to which the noun belongs, its plural prefix and information on the origin of the noun in question where applicable. In the case of verbs, various verbal suffixes are provided to indicate different extensions that apply to them. Past tense forms are also provided. In addition, decimal points are used to show places of junction between morphemes that make up each word. Examples of word usage are also offered to indicate how certain words can be used in context. Furthermore, lexicographic labels are provided to give the user an idea regarding which field or subject a particular word belongs to. These features make it unique because it differs from others which are restricted in nature such as Matšela's *Sehlahosi: Sesotho Cultural Dictionary* of 1994 and Pitso's 1997 thesaurus called *Khetsi ea Sesotho*.

According to the information provided in the back matter of Hlalele's dictionary (2005), *Sethantšo sa Sesotho* is the first Sesotho monolingual dictionary of its kind produced by a mother-tongue speaker. It encourages users to speak and write "Sesotho sa 'Mankhonthe'" which can be translated as 'the real Sesotho' or 'the original Sesotho'. The dictionary is used as a reference book by students at secondary and high school level. It is also useful to student teachers, teachers, lecturers and university students who are doing African languages and literature and libraries all over the Southern African Development Community (SADC) countries. *Sethantšo sa Sesotho* has not yet been revised since its publication (according to the researcher's knowledge).

3. Comparing dictionaries

Comparative studies between dictionaries are made for various reasons including the evaluation of dictionary use while reading and writing, reasons for dictionary consultation, knowledge of words and assessment of the users' needs which determine the dictionary plan. For instance, scholars such as Prinsloo (2005), Laufer (2000), Nesi (2000) and Lomicka (1998) deal with the effectiveness of paper dictionaries versus electronic dictionaries during a reading comprehension experiment. Shiqi (2003) analyses dictionaries that were produced over a period of time while Rundell (2008), Hatherall (1986) and El-Badry (1986) survey dictionaries which derive from the same source to identify the changes that occurred over time. This study follows those carried

out by scholars such as Shiqi (2003), Rundell (2008), Hatherall (1986) and El-Badry (1986).

Shiqi (2003) analyses ancient and modern Chinese monolingual dictionaries from the ninth century BC to 2002. The study looks at the development of these dictionaries in terms of their classification, arrangement of words, number of entries, how words are explained, and types of words included, such as names of implements, geographical features, names of plants and animals as well as kinship terms. The study reveals that ancient dictionaries were used as a basis upon which the modern dictionaries are compiled. The ancient dictionaries are smaller and were created by individuals while the modern ones are larger and produced by groups of scholars. Words and characters are mostly arranged according to the radical order (that is, characters of the same radical are grouped together) both in the ancient and modern dictionaries. Explanation of words is brief in the ancient dictionaries compared to the modern ones and the number of entries increased. Both the ancient and modern dictionaries contain common and specialised terms. This means that the modern ones are improved and add on to what was already presented. For instance, the modern dictionaries cover scientific and technical terms from more than 120 disciplines.

Rundell (2008) studied the recent developments in English monolingual dictionaries. The study deals with the extent to which the advanced *English Monolingual Learner's Dictionary* (MLD) has moved on from Hornby's *Idiomatic and Syntactic English Dictionary* (ISED) of 1942. The study establishes that the dictionary now has broadened to encompass such areas as pragmatics, cultural allusion, encyclopaedic information and guidance on every aspect of grammar and usage. Again, monolingual learner's dictionaries moved away from the model of the native-speaker's 'dictionary of record' towards a more 'utilitarian' lexicography, in which the needs of the user take precedence over all other factors.

Hatherall (1986) compares the *Duden Rechtschreibung*'s 1985 edition by Leipzig and the 1986 edition by Mannheim and reveals that the editions differed significantly from each other mainly because they stemmed from different publishers and different editorial boards.

El-Badry (1986) surveys seven Arabic–English and eight English–Arabic dictionaries in order to trace the development of the bilingual lexicography of these two languages in terms of the explicit or implicit plans of their respective authors and the sources they draw on. The study found that Arabic–English dictionaries used source material from several contemporary bilingual dictionaries and an Arabic monolingual dictionary. The English–Arabic dictionary utilised bilingual dictionaries of Arabic and French plus other linguistic and literary works of classical writers.

Studies similar to those undertaken by Shiqi, Rundell, Hatherall and El-Badry have not yet been done in Sesotho dictionaries. This study therefore attempts to bridge that gap by establishing the relationship between the *Sesuto–*

English Dictionary and the *Sethantšo sa Sesotho* to determine what needs to be done to improve dictionaries in Sesotho.

4. Comparison of the *Sesuto–English Dictionary* and the *Sethantšo sa Sesotho*

To establish whether or not there is a relationship between the *Sesuto–English Dictionary* and the *Sethantšo sa Sesotho*, the designs and entries of the two dictionaries were compared. Dictionary design involves proper planning of the structure of the dictionary in question. According to Gouws and Prinsloo (2005: 13-16), the plan is concerned with the direct lexicographic issues and focuses on aspects such as the lexicographic functions, dictionary typology, target user, structure of the dictionary, and lexicographic presentation.

Lexical entry 'refers to the entry in a dictionary of information about a word' (<http://www.thefreedictionary.com>). The study therefore compared all lexical entries, also known as headwords, contained in the *Sethantšo sa Sesotho* with those that are presented in *Sesuto–English Dictionary*, that is, all the items shared by the two dictionaries were identified as well as those that are peculiar to *Sethantšo sa Sesotho* only. This means that the study focuses only on the lexical items that are included in the *Sethantšo sa Sesotho* while those that appear only in the *Sesuto–English Dictionary* are left out. This is done to establish how close or far the two dictionaries are and to find out if the new dictionary is better than its predecessor.

4.1 Differences

The two dictionaries are different in the sense that, the *Sesuto–English Dictionary* was compiled by missionaries in the nineteenth century to assist them to learn and understand Sesotho so that they could evangelise the Basotho. On the other hand, the *Sethantšo sa Sesotho* was written by a Mosotho in the twenty-first century to help the Basotho to use the language appropriately.

The dictionaries are also different in that the former is bilingual while the latter is monolingual. One might argue that this fact does not place the two dictionaries on an equal footing for the purpose of a comparative analysis. However, only the Sesotho section of the bilingual dictionary is compared to entries in the monolingual dictionary. In addition, these dictionaries are the only Sesotho dictionaries of note available currently and much can be learnt from a comparison between the two, even though the former has the non-Sesotho speaker in mind as its target user, while the latter is aimed at both the non-Sesotho and mother-tongue speakers. The dictionaries are of different sizes. Unlike Shiqi's (2003) ancient dictionaries which were smaller than the modern ones, the *Sethantšo sa Sesotho* (recent dictionary) is smaller than the *Sesuto–English Dictionary* with 9,561 lexical items compared to the 20,039 items

included in the Sesotho section alone of the bilingual dictionary. The discussion below will focus on the arrangement of words, word-division, derivation, noun classes and plural morphemes.

4.1.1 Arrangement of words

The two dictionaries use the same orthography and the same orthographical alphabet except that the *Sesuto-English Dictionary* includes the letters *d*, *g* and *v* which are not in the *Sethantšho sa Sesotho*. All the entries in both dictionaries start with the lemma, which appears in bold type. In both the *Sesuto-English Dictionary* and the *Sethantšho sa Sesotho*, words are arranged in an ordinary alphabetical order, however, in the *Sesuto-English Dictionary* the article stretches are represented by monographs, whereas the *Sethantšho sa Sesotho* has utilised phonemic sorting. For instance, in *Sethantšho sa Sesotho*, words are arranged as follows: A, B, Ch, E, F, H, Hl, I, J, K, Kh, K'h, L, M, N, Ng, Ny, O, P, Ph, Pj, Psh, Q, Qh, R, S, Sh, T, Th, Tj, Tl, Tlh, Ts, Tš, U. The digraphs and trigraphs hl [ɬ]; kh [kxh]; k'h [kʰ]; ng [ŋ]; ny [ɲ]; pj [pʃ]; psh [pʃʰ]; qh [tʰ]; sh [ʃ]; th [tʰ]; tj [tʃ]; tl [tɬ]; tlh [tʰɬ]; ts [ts]; tš [tʃʰ] are thus treated as separate article stretches. This implies that words such as *hopola* (remember) and *hula* (pull) appear before the word *hlaba* (prick or sting) in this dictionary. The order itself might cause a problem especially during the first consultation of the dictionary because guidance is not provided to help users know how to search for words. This means that only users who are experts in Sesotho might find the order of words in the *Sethantšho sa Sesotho* easier to understand than those who are learning the language, particularly if they are not sure of the spelling of a word. The latter group might thus not find the phonemic sorting of words user-friendly, despite the fact that the dictionary is intended for both mother-tongue and second-language learners as stipulated in the back-matter, namely that it can be used by students and lecturers of African languages and literature in all the SADC countries. The dictionary does not conform to the user-perspective approach which expects dictionaries to serve the specific needs and research skills of specific target user groups, that is, dictionaries need to provide in the real needs of real users and take into consideration the users' reference skills (Gouws and Prinsloo 2005: 3). Again, those who know the spelling are also likely to be confused because they might think that the word(s) they are looking up are not in the dictionary yet they are there but placed where users are not expecting to find them. According to Prinsloo (2013: 247), dictionaries that use phonemic sorting instead of an alphabetical order, irritate users. He further states that even though the phonemic sorting is based on sound grammatical considerations, users regard it as user-unfriendly.

4.1.2 Word-division

In the *Sesuto-English Dictionary*, word-division is not indicated while in the

Sethantšo sa Sesotho verb-roots are separated from the verbal ending/suffixes by a dot [.] to show users where different suffixes can be inserted, because in most cases the verb root does not change. For instance, the lemma **kheloh.a** (err/turn from) consists of:

Verb-root	+	verbal-ending
kheloh	+	a

This indicates that the word is made up of two parts which are /kheloh-/ and /-a/. The first part of the word (i.e. the root) cannot change whereas the second one can change. According to Guma (1971) the verbal root is the central morpheme, which cannot change even after all affixes, whether prefixal, infixal or suffixal, have been removed. This information enables users to know where to insert or not to insert any morpheme. Some of the morphemes that can be put into that slot include past-tense morphemes.

4.1.3 Derivation

In the *Sesuto–English Dictionary*, derivative forms are presented in the dictionary article of the lemma and are followed by explanations of their meanings. For example:

talima, v.t., to look at, to contemplate, to consider, to watch; to concern one; *talimana*, to look at one another, to be parallel; *taba ena e talimane le 'na*, that matter concerns me; *italima*, v.r., to look at oneself; *talimisa*, v.t., to cause to look at, to help to consider a question; to direct toward ... (Mabille and Dieterlen 2000: 436).

In the *Sethantšo sa Sesotho*, on the other hand, the derived words appear as separate headwords as in the following extracts:

talim.a(.ile & .me) /*kutu-ketso*/ ho sheba ho hong kapa e mong; ho boha ho hong ... (<talima) (Hlalele 2005: 260)

talim.an.a(.e) /*kutu-ketso*/ ho shebana; ho bohana; ho halimana. (<talima) (Hlalele 2005: 260)

talim.el.a(.etse) /*kutu-ketso-ketsetso*/ ho sheba ho hong ka morero o itseng ... (<talima) (Hlalele 2005: 261).

talim.is.a(.itse) /*kutu-ketso-ketsiso*/ ho etsa hore ho talingoe ... (<talima) (Hlalele 2005: 261)

The inclusion of (<talima) at the end of each of these articles shows that the headwords are derived from *talima* which means to look at or to watch. Some users may believe that the above headwords are not related and that also contributed to the number of lexical items treated in *Sethantšo sa Sesotho*. One would

assume that if related words are treated as separate headwords, one could have expected an increase in the number of lexical entries in Hlalele (2005) — yet this is not the case.

4.1.4 Noun classes and plural morphemes

Information regarding the noun classes and the plural morphemes is not offered in *Sesuto–English Dictionary* whereas it is provided in *Sethantšo sa Sesotho*, as in the following examples:

tinkana, n., ox with horns bent forward (Mabille and Dieterlen 2000: 457)

tinkana (li.) /lereho 9/ poho kapa pholo e linaka li koropeletseng ka mahlong (Hlalele 2005: 269).

The (li.) is a plural morpheme of *tinkana* and the number (9) indicates the noun class of the headword. Provision of the plural morphemes and the noun classes is essential for students and other people who may want to learn the language, hence, making the dictionary user-friendly. These are the major differences seen in the two dictionaries. The following section deals with the similarities.

4.2 Similarities

The gap between the last edition of the *Sesuto–English Dictionary* (i.e. 1937, before it was revised following the stem approach) and the publication of the *Sethantšo sa Sesotho* (2005) is roughly 68 years. It is therefore, surprising to see that the two dictionaries share the following features: use of foreign sounds and sound patterning, illustrative phrases and use of old/obsolete words. One would have expected a greater measure of modernisation in *Sethantšo sa Sesotho*.

4.2.1 Use of non-standard symbols and atypical sound patterning

The two dictionaries make use of some symbols which are not part of the standard practical orthography of Sesotho as is evident in their use of *d* and *g*. The symbol *d* (phonemically /l/) is utilised in words such as *daemane* (diamond) instead of *taemane* and *diabolosi* (devil) instead of *liabolosi* (Mabille and Dieterlen 2000: 54). The same dictionary also uses the symbol *g* (phonemically /x/) in words such as *gansi* (goose) for *khantši* and *galasi* (glass) for *khalasi* (Mabille and Dieterlen 2000: 69) etc.

Likewise, the *Sethantšo sa Sesotho* also uses the symbol *d* which is not represented in the Sesotho orthography. This is evident in its inclusion of words such as *adora* (to adore), *adoreha* (adorable) (Hlalele 2005: 1) and *sanadere* (particular type of gun) (Hlalele 2005: 233).

It is true that the sound [d] is part of the spoken language, but the symbol *d* is not included in the inventory of Sesotho orthography. The sound [d] is a variant of the phoneme /l/ and is perceived when the phoneme /l/ is followed by the vowels [i] or [u], i.e. when there are syllables with (l + i) = *li*; and (l + u) = *lu*. The syllables (*li*) and (*lu*) in Sesotho are pronounced as [di] and [du]. Hence, the first syllable of the Sesotho greeting *Lumela* does not sound like [lu] in Luke but rather like [du]. Hlalele (2005: v) mentions that 'd' is realised when 'l' is used with the vowels 'i' and 'u', but when 'd' is followed by the vowels 'a', 'e' and 'o' it changes to 't'. However, he failed to apply that rule to the words *adora*, *adoreha* and *sanadere*. Hlalele contradicts himself, since he says:

... *puo efe kapa efe e na le nteteroane ea eona e sa itšetlehang ho tsa puo tse ling. Haeba taba li tsamaea ka nepo, le mainahano a tsepameng, puo ka 'ngoe e latela tsela ea eona ea mongolo e sa pepang mongolong oa puo tse ling* (2005: iv).

(... each language has its own sound system which does not lean on other languages. If things go the right way based on the right thinking, each language should use its own orthography without leaning on other languages — own translation).

According to this statement, each language should use its own orthography as it is a language in its own right. However, based on Hlalele's use of symbols which are not part of the standard Sesotho orthography, one gets confused because it looks like there are exceptional cases which allow users to use *d* and not *t* even though Hlalele himself mentioned that the letter *d* should be changed to *t* when followed by the vowels *a*, *e* and *o*. Dictionaries are expected to provide users with information (for example, spelling) that is valid and acceptable, however, the inclusion of this type of information may mislead learners in particular. They might believe that the mentioned symbols can be used yet they are not among the standard Sesotho symbols.

Furthermore, both dictionaries utilised atypical sound patterning. This is seen in the inclusion of words such as *tramontene* or *tramtene* (turpentine) (p. 473) in the *Sesuto–English Dictionary* and *trakema* (drachma) and *trakone* (dragon) (p. 273) in the *Sethantšo sa Sesotho*. The dictionaries (especially *Sethantšo sa Sesotho*) did not attempt to adapt atypical sound combinations to comply with the open syllable system of Sesotho, whereby unacceptable consonant clusters should be separated by vowels. Even though /t/ and /r/ are among the phonemes of Sesotho, they are not among the consonants that can form consonant clusters. A common Sesotho syllable structure consists of a consonant and a vowel (Guma 1971: 25). On the other hand, vowels are correctly added at the end of these words to comply with the Sesotho syllable structure in a word. All Sesotho words end with vowels except for words ending with *ng* /ŋ/.

It seems that where the missionaries were unable to represent particular Sesotho sounds in the standard orthography, they utilised symbols from European languages to stand in for sounds which they could not represent otherwise. This is reasonable and understood for foreign language speakers and

particularly the missionaries for they were the first to put Sesotho into writing. However, the continued use of non-standard symbols and atypical sound patterning in *Sethantšo sa Sesotho* does not reflect that the new dictionary was produced by a mother-tongue speaker nor that it has moved away from the *Sesuto–English Dictionary*.

4.2.2 Illustrative phrases and sentences

The use of similar illustrative phrases is another factor that links the *Sesuto–English Dictionary* and the *Sethantšo sa Sesotho*. It seems that the majority of illustrative sentences which are used in the *Sethantšo sa Sesotho* are duplications of the ones used in the *Sesuto–English Dictionary*. For example:

khala, n., crab; *likhala tsa molapo o le mong*, (crabs of the same brook, people of the same kind) (Mabille and Dieterlen 2000: 127)

khala² (li.) /lereho 9/ phoofotsoana e nyenyane e phelang metsing e tsamaeang ka lekeke. **ml.** *khala tsa molapo o le mong*: batho ba morero o le mong, ba mekhoha e tšoanang, ba sepheo se tšoanang, ba utloanang (Hlalele 2005: 80)

khanyapa, n., a fabulous water serpent; *selemo sa Khanyapa*, 1840 (Mabille and Dieterlen 2000: 129)

khanyapa (li.) /lereho 9/ pula e ngata hoo meholi e phuphuthang fatše 'me lifate li kotohang ka metso; noha eo ho hopoloang hore ke ea metsi 'me ha e falla nakong ea lipula tsa melupe ea litloebelele e heletsa matlo 'me e fothola lifate. *Selemo sa khanyapa*: selemo se hlahlamang komello e kholo ea lerole le leholo le lefubelu sa 1840 sa pula e bongata bo tšabehang (Hlalele 2005: 82).

phōnyōnyō, n., something one cannot seize or hold; *ho tšoara phonyonyo*, to try and to fail (Mabille and Dieterlen 2000: 352).

phonyonyo (#bongata) /lereho 9/ eng le eng e se nang botšoareho. **ml.** *ho tšoara phonyonyo*: ho tšoara 'mamphela ka sekotlo; ho ba bothateng; ho itšoarella ka mohatl'a pela (Hlalele 2005: 182)

The italicised phrases or sentences occur in both dictionaries as seen in the above extracts. There are several instances of this, and that proves that the two dictionaries are somehow related. The use of similar illustrative phrases in the dictionary that was published many years after the *Sesuto–English Dictionary* suggests that the *Sethantšo sa Sesotho* has not moved away from the former. It is assumed that if a different source had been used instead of *Sesuto–English Dictionary*, the illustrative sentences could have been different.

Ilson (1986) posits that there is nothing wrong with using information from existing dictionaries, because lexicographers have opportunities to add value to the existing data in order to maximise the usefulness of a new dictionary for users. Bothma and Tarp (2012) concur that lexicographers do not only make use of existing lexicographical tools but they reuse and recreate existing data from the database, internet and elsewhere. Again, this is in line with the theory of adaptation, which stipulates that 'art is derived from other arts' (Hutcheon 2013: 2), which simply means that a new text is created with material from elsewhere, i.e. the product is an 'extended reworking of other texts [and] adaptations are often compared to translations' (Hutcheon 2013: 16). This indicates that in adaptation, changes can occur in terms of the order of items/events, reduction or expansion of some material that can lead to major differences between the source, and the adapted text.

4.2.3 Use of old/obsolete words

Words which were used during the compilation of *Sesuto-English Dictionary* (old/obsolete words) are also presented in *Sethantšo sa Sesotho* as if they are common. These words are mostly used by old people and are not common to the contemporary users as they are not found in the majority of literary texts or newspapers. The following words show evidence of such instances:

Table 1: Obsolete words

Unfamiliar	Familiar	Translation
lekhono	Lefutso	Heredity, resemblance
lesafo	Lelapa	Family
lesela	Lesholu	Thief
mefuthaketso (here, an indication is given in <i>Sesuto-English Dictionary</i> that the word is the old name for trousers)	Borikhoe	Trousers
'moana	Matekoane	Dagga
lekhonya	Lekhooa	White person
lengeto	Leeto	Journey
letsiboho & tsiboho	Leliboho	Ford
seate	Leoatle	Ocean, sea
senyabela	Leoto	Foot

The words presented in Table 1 above are rarely used but they are presented as if they are common in *Sethantšo sa Sesotho*. This dictionary provided these words without indicating through the use of lexicographic labels that they are archaic. For instance, the *Sesuto–English Dictionary* revealed that a word such as *mefuthaketso* (trouser) refers to the 'old' name for trouser but Hlalele presented it as if it is a normal word. Zgusta (1971) posits that all obsolete and regional words should be labelled as such by a sign or label because if this were not done, the word would be regarded as normal or current. The fact that *mefuthaketso* was already considered 'old' when the *Sesuto–English Dictionary* was compiled, shows that there is a possibility that users might not encounter it in their daily conversations.

In some instances, the *Sethantšo sa Sesotho* uses unfamiliar words as the headwords and the common words are only found in the explanation of the words in question. That is, the commonly used words do not occur as headwords in the dictionary. When going through the explanation, one notices that the word refers to a known item, which is not in the dictionary. The following extracts bear testimony to such occurrences:

lekhono (#bongata) /*lereho* 5/ lefutso; tšoano e tsoeleletseng
(Hlalele 2005: 115).

letsiboho (ma.) /*lereho* 5/ moo ho tšeloang nokeng; leliboho
(Hlalele 2005: 119).

In the above extracts, the words *lefutso* and *leliboho* are common but they are not treated as headwords in this dictionary. The fact that Hlalele used the common words while explaining the meanings of the words considered unfamiliar, shows that he was aware of their existence, but he did not include them as main lemmata for some reasons known to him. This type of presentation does not benefit the users who only know the currently used words because it is difficult to anticipate that the known words would appear under the explanation of the meanings of the less familiar words. As a result, one may conclude that *Sethantšo sa Sesotho* seems to have been neglecting the current generation since its focus is similar to that of Mabile and Dieterlen. If Hlalele wanted users to have knowledge of both versions of the words (i.e. former and current usage), he should have included the unknown as well as the known lexical items as headwords in the dictionary. Based on these findings, it is evident that the new dictionary has not distanced itself from the old one and that the changes that have occurred in Sesotho have been neglected.

The shift from dictionaries compiled by the missionaries to modern dictionaries is expected to be seen through the inclusion of current terminology. Mtuze (1992) emphasises that the latest developments are reflected in a dictionary by including neologisms introduced into the lexicon via current politics, technology, diseases, etc. The high frequency words are expected to be given appropriate treatment and consideration in monolingual dictionaries more

than in other dictionaries because they are widely used in textbooks (Holi 2012). If we concur with Mtuze's (1992) idea, that the dictionaries produced by the missionaries contain many words that have fallen into disuse, and have limited vocabulary, then Sesotho lexicography has not yet moved away from the past. Many words which are currently used do not occur in this new Sesotho dictionary. Therefore, the study concludes that *Sethantšo sa Sesotho* has moved away only slightly from the *Sesuto-English Dictionary*. It is as if it was intended for the same target users (i.e. Mabile and Dieterlen's target group). It is also assumed that much of what Hlalele has produced may soon be of little value to the current generation because many of the changes in the Sesotho language were neglected in his dictionary. As a result, the *Sethantšo sa Sesotho* is not considered to be better than *Sesuto-English Dictionary*.

5. Conclusion and recommendations

In conclusion, it was found that the two dictionaries chosen for comparative analysis in this study, revealed pertinent differences and similarities. They are different in their typology since the *Sesuto-English Dictionary* is bilingual and was produced by missionaries while the *Sethantšo sa Sesotho* is monolingual and was compiled by a Sesotho mother-tongue speaker. The former was published in the 19th century while the latter was published in the 21st century. The old dictionary is large since it consists of 20,039 headwords in the Sesotho section whereas the number of headwords in the new dictionary is less (only 9,561), contrary to the trend found in studies by Rundell (2008), Hatherall (1986), and El-Badry (1986) which revealed that new dictionaries (particularly, those derived from the former ones) were larger than the old ones and showed a spectacular increase of words over the years and throughout the editions. The *Sesuto-English Dictionary* is alphabetically ordered while the *Sethantšo sa Sesotho* followed phonemic sorting. Again, the old dictionary does not show word division but the new one does. Derived forms are treated under the same dictionary article in the old dictionary while in the new one they are presented as separate items and the word from which the word is derived is indicated at the end of the dictionary article. In addition, the old dictionary does not provide plural morphemes and classes of nouns while in the new one they are offered. The manner in which the *Sethantšo sa Sesotho* has presented information is considered beneficial to the contemporary user with regard to the indication of noun classes, plural morphemes and word-division. Regarding the similarities, the two dictionaries make use of some symbols which are not part of the standard Sesotho orthography and atypical sound patterning. This suggests that the new dictionary does not fully meet the needs of the current generation. The study concludes that the new dictionary has not distanced itself much from the old one and that the information contained in the new dictionary may soon lose its usefulness.

The study led to the realisation that there is need to produce a new mono-

lingual dictionary to improve the existing one. The dictionary to be produced should contain most current words which have entered Sesotho due to science and technology, borrowing, diseases, abuse, politics etc. that have never been written down in dictionaries and other words which are frequently used as well as words from the existing dictionaries. It is necessary to provide detailed information regarding the pronunciation of some Sesotho phonemes which could potentially be pronounced differently by people who are not familiar with the language. It is also recommended that Sesotho dictionaries should be compiled by groups of scholars and not by individuals.

6. References

- Ambrose, D.** 2006. *Lesotho Annotated Bibliography: Dictionaries and Phrase Books*. Maseru: House 9 Publications, National University of Lesotho.
- Awak, M.K.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa: Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–25 March 1989*: 8-18. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Bothma, T.J.D. and S. Tarp.** 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.
- Busane, M.** 1990. Lexicography in Central Africa: The User-perspective, with Special Reference to Zaïre. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989*: 19-35. Exeter: Exeter University Press.
- El-Badry, N.** 1986. The Development of the Bilingual English–Arabic Dictionary from the Middle of the Nineteenth Century to the Present. Hartmann, R.R.K. (Ed.). 1986. *The History of Lexicography*: 57-64. Studies in the History of the Language Sciences 40. Amsterdam: John Benjamins.
- Gouws, R.H.** 2005. Lexicography in Africa. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics*: 95-101. Second edition. Oxford: Elsevier.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Guma, S.M.** 1971. *An Outline Structure of Southern Sotho*. Pietermaritzburg: Shuter and Shooter.
- Hatherall, G.** 1986. The *Duden Rechtschreibung* 1880–1986: Development and Function of a Popular Dictionary. Hartmann, R.R.K. (Ed.). 1986. *The History of Lexicography*: 85-98. Studies in the History of the Language Sciences 40. Amsterdam: John Benjamins.
- Hlalele, B.** 2005. *Sethantšo sa Sesotho*. Maseru: Longman.
- Holi, I.H.A.** 2012. Monolingual Dictionary Use in an EFL Context. Available at: <http://dx.doi.org/10.5539/elt.v5n7p2>. [Accessed on 13 March 2014.]
- Hutcheon, L.** 2013. *A Theory of Adaptation*. New York: Routledge.
- Illson, R.F.** 1986. Lexicographic Archaeology: Comparing Dictionaries of the Same Family. Hartmann, R.R.K. (Ed.). 1986. *The History of Lexicography*: 127-136. Studies in the History of the Language Sciences 40. Amsterdam: John Benjamins.
- Laufer, B.** 2000. Electronic Dictionaries and Incidental Vocabulary Acquisition: Does Technology Make a Difference? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 849-854. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

- Lomicka, L.** 1998. "To Gloss or Not to Gloss": An Investigation of Reading Comprehension Online. *Language Learning and Technology* 1(2): 41-50. Retrieved from <http://llt.msu.edu/vol1num2/pdf/article2.pdf>. [Accessed on 19 March 2017.]
- Mabille, A. and H. Dieterlen.** 2000. *Sesuto-English Dictionary*. Reprint. Morija: Morija Sesuto Book Depot.
- Makoni, S. and P. Mashiri.** 2007. Critical Historiography: Does Language Planning in Africa Need a Construct of Language as Part of its Theoretical Apparatus? Makoni, S. and A. Pennycook (Eds.). 2007. *Disinventing and Reconstituting Languages*: 62-89. Clevedon, UK: Multilingual Matters.
- Mangoaela, Z.D.** 1921. *Lithoko tsa Marena a Basotho*. Morija: Morija Sesuto Book Depot.
- Matšela, F.Z.A.** 1994. *Sehlahosi: Sesotho Cultural Dictionary*. Maseru: Macmillan Boleswa.
- Mofolo, T.** 1907. *Moeti oa Bochabela*. Morija: Morija Sesuto Book Depot.
- Mofolo, T.** 1925. *Chaka*. Morija: Morija Sesuto Book Depot.
- Mtuzze, P.T.** 1992. A Critical Survey of Xhosa Lexicography 1772-1989. *Lexikos* 2: 165-177.
- Nesi, H.** 2000. On Screen or in Print? Students' Use of a Learner's Dictionary on CD-ROM and in Book Form. P. Howarth and R. Hetherington (Eds.). 2000. *EAP Learning Technologies. BALEAP Conference Proceedings*: 106-114. Leeds: Leeds University Press.
- Nkomo, D.** 2008. *Towards a Theoretical Model for LSP Lexicography in Ndebele with Special Reference to a Dictionary of Linguistic and Literary Terms*. Unpublished M.Phil. Thesis. Stellenbosch: Stellenbosch University.
- Otlogetswe, T.J.** 2013. Introducing *Tlhalosi ya Medi ya Setswana*: The Design and Compilation of a Monolingual Setswana Dictionary. *Lexikos* 23: 532-547.
- Paroz, R.A.** 1950. *Southern Sotho-English Dictionary*. Morija: Morija Sesuto Book Depot.
- Pitso, T.T.E.** 1997. *Khetsi ea Sesotho*. Cape Town: CTP Book Printers.
- Prinsloo, D.J.** 2005. Electronic Dictionaries Viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.
- Prinsloo, D.J.** 2013. Issues in Compiling Dictionaries for African Languages. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 232-255. London/New York: Bloomsbury.
- Rundell, M.** 2008. Recent Trends in English Pedagogical Lexicography. Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*: 221-243. New York: Oxford University Press.
- Shiqi, X.** 2003. Chinese Lexicography Past and Present. Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts*: 158-173. London/New York: Routledge.
- www.kwintessential.co.uk/lang/. [Accessed on 19 July 2014.]
- www.thefreedictionary.com/lexical+entry by Farlex. [Accessed on 24 July 2014.]
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

Performing Non-sexism via Degendering Phoric Forms in English: The Gap between Rules and Practice as Observed in the 9th Edition of *Oxford Advanced Learner's Dictionary of Current English*

Nshindi-Germain Mulamba, *Department of English Letters and
Civilization, University of Lubumbashi, DR Congo*
(mulambagermain56@gmail.com)

and

Francis Ngoyi Crequi Tshimanga, *Department of English and African
Culture, Teacher Training College of Mbujimayi, DR Congo*
(ftshimangangoyi@gmail.com)

Abstract: Non-sexism is one of the outstanding and obvious proofs of how social and cultural changes are taken into account in the English language. However, it is still a usage problem for natives and learners alike. This paper uses the degendering of phoric elements as an illustration of efforts in the 9th edition of *Oxford Advanced Learner's Dictionary of Current English* (OALD9) to capture and express social and cultural changes in the English language. It also attempts to point out some mismatches between rules and actual practice in this enterprise. Finally, it intends to show how the dictionary may act as an ideological tool, imposing, sustaining, highlighting, and perpetuating some points of view to the detriment of others. In practice, the use of more than one structure to achieve non-sexism makes fluency difficult and writing cumbersome.

Keywords: GENDER, PHORIC ELEMENTS, SEXISM, NON-SEXISM, DICTIONARY, IDEOLOGY, USAGE PROBLEM

Résumé: La pratique du non-sexisme à travers la neutralisation du genre des anaphores en anglais: L'écart entre les règles et la pratique tel que perçu dans la neuvième édition d'*Oxford Advanced Learner's Dictionary of Current English*.

L'emploi du langage non sexiste en anglais est une preuve patente de la prise en compte par la langue des mutations sociales et culturelles. Cependant celui-ci demeure un problème d'usage tant pour les locuteurs natifs que pour les apprenants. Cet article puise dans la neutralisation du genre des anaphores comme un exemple des efforts déployés par la neuvième édition d'*Oxford Advanced Learner's Dictionary of Current English* (OALD9) afin de cerner et exprimer lesdits changements dans

la langue anglaise. Cet article essaie également d'identifier les points de convergence et/ou de divergence entre les règles établies et la pratique réelle dans la neutralisation du genre structurelle de la langue anglaise. En définitive, il envisage démontrer comment le dictionnaire peut servir d'outil idéologique, imposant, soutenant, mettant en exergue, et perpétuant certains points de vue au détriment d'autres. En pratique, l'emploi multiple des structures pour réaliser le non-sexisme rend difficile la fluidité du langage et l'écriture encombrante.

Mots clés: GENRE, ANAPHORES, SEXISME, NON-SEXISME, DICTIONNAIRE, IDEOLOGIE, PROBLEME D'USAGE

1. Introduction

For decades, gender-referring usage in English has tended to become an issue not only in English teaching but even in English native speakers' everyday communication (Mulamba and Tshimanga 2006). Language users have been confronted with the rise of new grammatical, lexical and discourse rules which compel them to become more and more aware of how they use language. That is, far from remaining purely descriptive, some grammars and studies have tended towards 'prescriptivism' to impose the new rules. The latter serve, among other things, to account for the changes in the language usage as a result of cultural and social changes, viz. gender.

To state it differently, these new rules challenge the traditional ones (e.g. the generic masculine) to such an extent that language use has become less and less intuitive even for native speakers. At times, speakers have to stop and think about the appropriate form to use in order not to commit a social blunder. The case in question herein is that of sexist language and how to avoid it. The present paper thus aims at addressing sexism through the treatment of degendering phoric forms in English in the *Oxford Advanced Learner's Dictionary of Current English*, 9th edition (OALD9 henceforth). By so doing, we try to capture how this dictionary tackles this issue and how it suggests solutions.

Different studies on language and society have shown that language is political (Bourdieu 1992). That is, it ideologically privileges some points of view while downplaying others. From a systemic functional perspective, language being made up of various systems which in turn comprise various subsystems (Halliday 2004), the choice of a given system to the detriment of another or others is to be considered as always motivated. It is in this framework that the use of masculine phoric elements as epicenes has been decried for being gender-biased (Mills 1995). As an offshoot to this situation, several strategies have been suggested as a solution to this problem. To this end, the paper draws on OALD9 to address these various suggested strategies (i.e. expectations or rules) and match them with actual achievements (i.e. practice).

After a short discussion of the methodological issues, the paper will first explore gender-indexing phoric elements in English, pointing out when they may be gendered or ungendered. Second, it will address their treatment in

OALD9 with a view to finding out whether or not there is a gap between the rules and actual practice in this respect. Finally, the paper will raise theoretical implications for the findings.

2. Methodological considerations

As regards data collection, this paper is basically a desk research (Sunderland 2006: xxiii) drawing on the skim-reading of OALD9 with a view to pointing out and describing gender-indexing phoric elements. That is, the lemmata selected are those belonging to the subclass of central pronouns which reflect the male-female natural gender contrast as well as their epicene counterparts (see section 3.1). In addition, some idiomatic expressions which include these phoric elements in either their definiendums or definiens are also selected to illustrate the dictionary's taking stance on structural gender-indexing in actual use.

As to the analysis of the data, the paper draws on lexical semantics (Van Roey 1990). Consideration will be taken of both the form (i.e. definiendum) and the meaning (i.e. definiens) of the lemma (i.e. phoric element) under study to point out whether it is gendered or ungendered. Furthermore, a selective choice will be made of Chomsky's Binding Theory (Cook 1988) in order to capture the reference relationships between antecedents and their phoric elements. Finally, the paper also draws on Critical Discourse Analysis (CDA) (Lazar 2005, Fairclough 2003) in order to show how the choice and treatment of a given phoric element may be ideologically loaded.

3. On gender-indexing phoric elements

The term *gender-indexing phoric elements* denotes gender-referring pronouns and adjectives which may function cataphorically or anaphorically (Halliday and Hasan 1976). Stated differently, they are those phoric elements likely to convey structural sexism or non-sexism, i.e. sexism or non-sexism which operates through syntagmatic relation within phrases, clause, sentences and/or text (Musau 1995, Mulamba and Tshimanga 2006, Tshimanga 2012, 2016, 2017a, 2017b and 2018). Such sexism is made explicit through the phrase structure or syntax, and particularly by considering coreferentiality between antecedents and their phoric elements (Cook 1988). Therefore, the linguistic context in which the gender-referring item is used becomes of paramount importance as it helps in deciding whether the item considered is gender-neutral or gender-biased. However, in the present study, the consideration of both the definiendum and the definiens of the phoric element will be of equal help.

Structural sexism is more inclusive than the sole consideration of phoric elements. For example, in her discussion of sexism at phrase and clause levels, Mills (1995) targets: (1) pre-constructed phrases, and (2) sentences, which, in turn includes proverbs, metaphors as well as jokes and humour. As already

clarified earlier, in the framework of this situation, structural sexism will be restricted to the relationship of phoric elements and their antecedents. This being the case, the focus will be put on how gender is constructed through the central pronouns (Halliday and Hasan 1976) in both nominal and determiner functions (Quirk et al. 1985).

3.1 Masculine pronouns as generic forms

The term *pronoun* needs clarifying as it may be differently considered in terms of terminology and scope. From a terminological point of view, the term *pronoun* may be regarded as misnomer since it suggests that a pronoun deputizes for a noun in the structure. Instead, pronouns are surrogates to noun phrases whose case, number and gender-marking they reflect (Quirk et al. 1985). It is the latter characteristic, that is, gender-marking, which is of relevance to structural gender-indexing in English. In other words, not all pronouns are concerned with structural gender-indexing, but only those which actually express natural gender differentiation. In this regard, Quirk et al. (1985: 314) show that only '[s]ome 3rd person pronouns and *wh*-pronouns do ... express natural gender distinction.' The following instances illustrate this view.

- (1) *The boy_i who_i is outside the theatre has just wounded himself_i.*
- (2) *The girl_i who_i is outside the theatre has just wounded herself_i.*

The reflexive pronouns *himself* in (1) and *herself* in (2) respectively denote the masculine and the feminine, which corresponds to the male-female natural gender distinction. Besides, the relative pronoun *who* denotes the personal gender as contrasted with the nonpersonal one.

It is worth warning that *wh*-pronouns, although expressing the contrast personal vs. non-personal, are not to be considered for structural gender-indexing on the ground that they do not show masculine vs. feminine distinction. Therefore, only the wider class of central pronouns is concerned by structural gender-indexing issues. In Quirk et al.'s (1985) understanding, these pronouns are called central pronouns because they share the features which are particularly characteristic of pronouns, *viz.* contrast of person, gender, and subjective/objective case. Furthermore, '[a]lthough these central pronouns fulfil different syntactic functions, they have obvious morphological resemblances.' The latter reason justifies the inclusion in the table below of the possessive *her*, for example — which is determinative in function and consequently cannot function alone as a surrogate to a noun (Quirk et al. 1985: 345-346). Halliday and Hasan (1976) use the term *personal pronouns* to refer to the traditional personal pronouns, possessive adjectives and pronouns, as well as reflexive pronouns; and justify this view with the centrality of these categories and their relevance to the descriptions of personal roles within discourse. Similarly, in their discussion of gender distinction in central pronouns in English, Quirk et

al. (1985: 341) hold that '[g]ender distinctions are largely restricted to 3rd person singular pronouns of the categories of personal, possessive, and reflexive pronouns'. For the sake of inclusivity, the term gender-indexing *phoric element* will be used. The following table is an illustration.

Table 1: Gender-indexing phoric elements in English

	Category	1	2	3		4
Personal	Masculine	he	him	his		himself
	Feminine	she	her	her	hers	herself

(Adapted from Quirk et al. 1985: 341)

In the above table, the following categories of central phoric elements are clarified, (1) subjective, (2) objective, (3) possessive, and (4) reflexive. Besides, the above table shows that the subjective personal pronoun *he*, together with its objective, possessive, and reflexive counterparts, is explicitly masculine.

However, although grammarians, for example, Quirk et al. (1972 and 1985), Thomson and Martinet (1980), to mention but a few, define the pronoun *he* as masculine (i.e. as referring to men and male animals), they equally consider it as generic. That is, *he* is regarded as gender-indefinite or unmarked pronoun used to refer to both males and females simultaneously, as O'Grady et al. (1993: 434) demonstrate it: 'We do not employ *he*, for instance, to refer only to males. In Standard English, it is used as a generic third person singular pronoun when the sex of the referent is unknown or irrelevant'.

Such a use of phoric elements as generic is said to be gender-biased. Mills (1995: 87) defines generic forms as 'those elements in language which perpetuate a view of male as a norm or universal and the female as deviant or individual'. Generally speaking, such kind of sexism known as *structural* or *grammatical sexism* (Musau 1995; Mulamba and Tshimanga 2006) is made overt through a syntactic and/or semantic relation known as *reference*. This cohesive device includes, among other relations, exophora, also known as *deixis* (Neveu 2011: 111) or situational reference, and endophora or textual reference (Halliday and Hasan 1976) depending on whether it operates outside or within text. Endophora is further divided into anaphora (backward reference) and cataphora (forward reference) (Halliday and Hasan 1976).

The term *anaphora* denotes a within-discourse referential relation between two linguistic expressions, one of which known as *anaphor* or *recall form*, receives its interpretation of the referential sense of the other, called anaphor source or antecedent (Neveu 2011: 37). By contrast, the term *cataphora* denotes a within-discourse referential relation between two linguistic expressions, one of which referentially dependent (i.e. *cataphor* or forward-pointing phoric element) receives in retroactive appearance its interpretation from another expression which is postposed to it (Neveu 2011: 72-3). The terms *anaphora* and *cataphora*

are also respectively called 'unmarked-order anaphora' and 'marked-order anaphora' (Huddleston 1984) or 'syntactically-controlled anaphora' and 'pragmatically-controlled anaphora' (Hankamer and Sag 1976, cited in Huddleston 1984: 298).

In addition, Huang (2010: 9) uses the terms *anaphor* and *anaphoric* as synonymous with *anaphora* and considers the latter to have three distinct senses in contemporary linguistics. In its first sense, this term 'can be used to refer to a relation between two or more linguistic elements, in which the interpretation of one element (called an anaphor or anaphoric expression) is in some way determined by the interpretation of another element (called an antecedent). And, the 'linguistic elements that can be employed to encode an anaphoric relation in this general sense range from phonetically unrealized gaps/zero anaphora/empty categories through pronouns and reflexives to various reference-tracking systems like gender/class, switch-function, and switch-reference'.

The second sense of the term is the one used by Chomsky's 'generative syntax for reference to a NP which has the features [+anaphor, -pronominal] versus pronominal as a NP with the features [-anaphor, +pronominal]'. In keeping with this view, Chomsky states three basic principles: A: An anaphor is bound in its domain; B: A pronominal is not bound in its domain, and C: An R-expression is free (see Cook 1988).

Finally, in the third sense, the term can be used to refer to an anaphoric expression whose antecedent comes earlier as opposed to 'cataphora/cataphor/cataphoric', whereby the antecedent comes later. In fact, this latter sense is related to the first which is its superordinate. To avoid this confusion, use will be made of the term *anaphora* in Huang's (2010) sense (3) (i.e. anaphora vs. cataphora), and the term *phoric element* for the superordinate or as synonymous to *anaphora* in Huang's (2010) sense (1). Besides, Huang (2010: 9) shows that 'anaphora can be (intra-)sentential, in which case the anaphoric expression and its antecedent occur within a single simplex or complex sentence. It can also be discursal, in which case the anaphoric expression and its antecedent cross sentence boundaries. In terms of syntactic category, Huang (2010: 9) distinguishes between (1) NP-anaphora and (2) VP-anaphora. Halliday and Hasan (1976) use the general term *reference*, which is wider in scope and subsumes endophoric vs. exophoric reference, but also personal, demonstrative and comparative reference.

Sexism is displayed in the first type of reference (i.e. personal reference) which is realized through the use of personal pronouns (i.e. the traditional personal pronouns, possessives and reflexives). Here, a masculine item denoting the 'other roles' i.e. third person singular, is used as coreferential with nouns of the personal dual class, the pronouns *everyone*, *everybody*, *someone*, *somebody*, *one*, together with their alternatives and possible combinations, to mention but these (Mulamba and Tshimanga 2006). Such a use of the pronoun *he* or its related forms as phoric (i.e. point back or forward) (Halliday and Hasan 1976)

to dual nouns and pronouns which are not necessarily masculine is and has been considered as a case of prevailing sexism in the English language. For Mills (1995: 87), '[t]he generic "he" is perhaps the most well-known example of gender specific or sexist language, and is frequently referred to as "he-man" language'.

As Mills (1995: 88) observes, different researchers have shown that generic pronouns are often understood as referring to male. For example, students who were asked to complete fragmentary stories which contained the generic pronoun 'he', in most cases, completed the stories using 'he' as a sex-specific pronoun or a male named character. In addition, she paraphrases Robertson's research on visualization and generic pronouns which supports that people tend to visualize male participants when the supposed generic pronoun 'he' is used. Sometimes, sexism in the use of the generic *he* is implied by other non-verbal elements, such as images, which accompany the message. That is, although this phoric element may be used as gender-inclusive, its masculine-specific character will be revealed by the image accompanying it (Mills 1995: 88).

Furthermore, some people consider the use of the generic pronoun as confusing. This stems from the fact that, in a given context, such a pronoun can be interpreted as used generically or gender-specifically. Finally, also considered as sexist is the fact that gender-specific pronouns are used to refer to practitioners of some stereotypically male or female professions. Besides, when the sex of the person is not known, it is commonly assumed that the person is male (Mills 1995: 88). As can be noticed, such kind of gender-based bias is subtle as it is not explicitly made but implicated and is to be arrived at by drawing inferences from what is said.

3.2 Suggested phoric element-degendering strategies

To level sexism in the use of explicitly masculine phoric elements as epicene, various strategies have been suggested. According to Mills (1995), these include: (1) the use of plural pronouns, (2) the use of conjoined forms 's/he', '(s)he', 'he or she', 'she or he'; (3) passivization, (4) the use of the female pronoun as generic with a proviso, (5) the use of the male pronoun as generic with a proviso, (6) the use of alternate pronouns, and (7) the use of singular 'they' (see O'Grady et al. 1993; Quirk et al. 1985; Swan 1984). However, without providing any statistics related to the frequency of each of these alternate forms, Mills (1995: 97) notes that some people use 'she' for the first occurrence of the pronoun, and 'he' for the next, and so on; which has the disadvantage of being potentially confusing.

The use of plural pronouns implies pluralization of the whole clause while the use of the female pronoun as generic is based on a formal argument assuming that 'she' can be seen to contain 'he'. The use of alternate pronouns consists in using 'she' and 'he' in turn. That is, for example, 'she' is used for the

first occurrence of the pronoun and 'he' for the next, and so on. The following instances illustrate:

- (3) *Each student_i submitted his_i assignment after he_i had completed it.*
- (4) *All students submitted their_i assignments after they_i had completed it.*
- (5) *Each student's assignment was submitted after it had been completed.*
- (6) *Each student_i submitted his/her_i assignment after s/he_i had completed it.*
- (7) *Each student_i submitted his_i assignment after she_i had completed it.*
- (8) *Each student_i submitted her_i assignment after she_i had completed it.*
- (9) *Each student_i submitted their_i assignment after they_i had completed it.*

Instance (3) is the case of the use of the masculine as generic. Given that this is considered as gender-biased, its use has to be accompanied by a proviso. The latter is a note or a kind of disclaimer warning the reader that the use of *he* is intended as generic. As can be noticed, this should be done at the beginning of a book, for example, and especially in the introduction. Example (4) illustrates the use of plural phoric elements yielding the overall pluralization of the whole clause to neutralize gender-bias.

However, were the gender-inclusive term *student* replaced by a generic masculine one such as *man*, pluralization would be of little help. Instance (5) clarifies the use of passivization to level gender-bias. Actually, the aim here is to remove any phoric element which would reflect gender discrimination. Therefore, any rephrasing in this way may equally be welcome provided that it is not confusing. Example (6) is a case in point of the joint use of masculine and feminine phoric elements while (7) illustrates the alternate use of the masculine and feminine as coreferential with the same antecedent. This strategy seems confusing insofar as the phoric elements may be interpreted as coreferential to different antecedents. Instance (8) is all about the use of the feminine phoric element as gender-inclusive while (9) illustrates the use of plural *they*.

Each of the above strategies has its specificity of use. For example, the use of the masculine phoric element as generic with a proviso is more likely to occur in the written rather than in the spoken medium. Similarly, while the singular *they* might occur in both media, it is generally regarded as occurring in the informal style. Accordingly, commenting on the latter case, Swan (1984: 236) writes that '[a]nybody, anyone, somebody, someone, nobody, no-one, everybody and everyone are used with singular verbs ... we often use *they*, *them* and *their* to refer to these words, especially in an informal style'. Admittedly, usage has evolved since as shown in other posterior studies below. For example, drawing on *The Comprehensive Grammar of the English Language*, Sunderland (2006: 12) sustains that the singular *they* is 'now increasingly accepted even in formal usage'. To put it clearly, the singular *they* is used as coreferential to either an indefinite pronoun or a noun phrase whose head is a dual noun. However, wherever there is a gender-specific item, use is made of either *he* or *she* exclusively (O'Grady et al. 1993: 434).

Similar to the above use is the case in which the indefinite *one* occurs as antecedent. Gender-bias can be avoided, here, by repeating *one* as phoric to *one* in every configuration, for example '*One should never criticize if one is not sure of one's facts*'. This is actually the use of the indefinite *one* which denotes 'people in general' or 'I' when the speaker is referring to himself or herself. Such a 'use of *one* is very formal and now sounds old-fashioned'. Thus, the pronouns *you* and *I* are currently much more usual for the first and the second meaning respectively (Hornby 2015: 1076). However, the joint use of the masculine and the feminine or the use of the singular *they* would be equally acceptable.

Furthermore, the zest to wipe out sexism based on generic *he* has led some researchers to coin "unisex pronouns" SE (he/she), SEM (him/her), SES (his/her/hers), and SEMSELF (himself/herself) (Abdel-Nabi and Hilfi 1989, Musau 1995). Full capitalization of these forms may be intended for highlight. Actually, such an attempt is one of the response strategies to the problem generated by the absence of a true sex-neutral 3rd person singular pronoun in English (Herbert and Nykiel-Herbert 1986: 52). Accordingly, they argue 'Specific proposals for new pronouns abound, e.g. *co, tey, hesh, thon, xe, E, po*.' Such a coinage might have been carried out on the basis of some diachronic or synchronic linguistic justification. For example, Miller and Swift (1977) (cited by Herbert and Nykiel-Herbert 1986: 52) argue that 'such coinages are not new: *thon*, for example derives from *that one* and was first recorded in 1859 and has been listed in many standard reference works'.

Table 2 below recapitulates on different proposals (i.e. expectations) regarding the use of gender-inclusive pronouns or epicenes, as well as their masculine and feminine counterparts.

Table 2: Gender-indexing phoric elements with suggested revisions

		Personal		Possessive		Reflexive	
		Subjective	Objective	Determinative	Independent		
Personal	Masculine	1	<i>he</i>	<i>him</i>	<i>his</i>		<i>himself</i>
	Feminine	2	<i>she</i>	<i>her</i>	<i>her</i>	<i>hers</i>	<i>herself</i>
	Epicene	3	(s)he, s/he	him/her	his/her	his/hers	him-/herself
		4	<i>they</i>	<i>them</i>	<i>their</i>	<i>theirs</i>	<i>themselves</i>
		5	SE	SEM	SES		SEMSELF
		6	<i>co, tey, hesh, thon, xe, E, po</i>				
Nonpersonal		<i>it</i>		<i>its</i>		<i>itself</i>	

The above table needs commenting for clarification. Pronouns in (1) are explicitly masculine and illustrate the proposal according to which the masculine may be used as generic with a proviso. Although explicitly feminine, (2) illustrates the

proposal supporting the use of the feminine as generic while (3) illustrates the joint use of the masculine and feminine as epicene. Option (4) illustrates the use of the singular *they* while options (5) and (6) are all about completely invented forms, without current relevance to everyday language use. While forms in (1) to (4) are actually used in everyday discourse, those in (5) and (6) are still but mere suggestions. Forms in (1) to (5) have a complete distribution across different types of central pronouns: personal, possessive, and reflexive. Although the table also outlines nonpersonal phoric elements for a complete picture of the situation, it leaves out two of strategies described above consisting, in turn, (1) in repeating indefinite *one* and in (2) alternatively using the masculine and feminine in the same text. This is justified by two facts: (1) the impossibility to tabulate the latter strategy and (2) the irrelevance of the indefinite *one* to the issue of indicating the speech roles (see Halliday and Hasan 1976).

By way of a conclusion, it is worth noting that the discussion of degendering strategies in this section is intended to serve illustrative purpose. This stems from the fact that the scope of sexism is so wide that it cannot be restricted to word or sentence level. That is, it can be conveyed in rather subtle ways at discourse level and involving, for example, talk in interaction where it may be deployed through a wide range of other strategies. On the other hand, designing strategies for the degendering of the English language is not the whole story. Rather, it is just a step on the scale of gender-neutral language implementation; and the latter is not without its problems, as will be clarified below.

3.3 Treatment of gender-indexing phoric elements in OALD9

Since structural or grammatical gender-indexing is made possible through the cohesive relation of reference, it is important to discuss in turn different phoric elements which enter this relationship. Such phoric elements are those belonging to the category of central pronouns, and which explicitly refer to gender-referring antecedents. These gender-indexing phoric elements are discussed in the following examples.

	<i>Definiendum</i>	<i>Definiens</i>
(10)	<i>He</i>	<i>Pron. (used as the subject of a verb) (1) a male person or animal that has already been mentioned or is easily identified.... (2) (becoming old-fashioned) a person, male or female, whose sex is not stated or known, especially when referring to sb mentioned earlier or to a group in general: Every child needs to know that he is loved. (saying) He who (=anyone who) hesitates is lost → NOTE AT GENDER (OALD9: 723)</i>
(11)	<i>She</i>	<i>(used as the subject of a verb) a female person or animal that has already been mentioned or is easily identified: 'What does your sister do?' 'She's a dentist.' → NOTE AT GENDER (OALD9: 1425)</i>

- | | | |
|------|-----------------|---|
| (12) | <i>S/he</i> | <i>pron. Used in writing by some people when the subject of the verb could be either female (she) or male (he): If a student does not attend all the classes, s/he will not be allowed to take the exam. (OALD9: 1425)</i> |
| (13) | <i>Themself</i> | <i>(the reflexive form of they) used instead of himself or herself to refer to a person whose sex is not mentioned or not known: Does anyone here consider themself a good cook? HELP: Although themself is fairly common, especially in spoken English, many people think it is not correct. (OALD9: 1622)</i> |
| (14) | <i>They</i> | <i>Used instead of he or she to refer to a person whose sex is not mentioned or not known: If anyone arrives late they'll have to wait outside. → NOTE AT GENDER (OALD9: 1625)</i> |

Example (10) reveals that, although becoming old-fashioned, the subjective masculine pronoun *he* is considered as generic, that is, as also referring to both masculine and feminine antecedents. This would lead to the interpretation that the dictionary still, somehow, perpetuates some fossils of biased usage as far as gender representation in phoric elements is concerned. However, the presence of the diachronic label *becoming old-fashioned* might be considered as a strategy aimed at discouraging the user in the choice of this phoric element.

By contrast, its feminine counterpart, i.e. the subjective pronoun *she* in (11) is considered as specific. This shows that its generic meaning commonly found in feminists' writers has been simply omitted. This omission might also be regarded as contribution to the obfuscation of femininity. The form *s/he* in (12) is used in writing to refer to gender-inclusive antecedents. In fact, we are of the view that the feminine generic meaning could at least be signalled, although with some usage comments.

Besides, in (13) the reflexive pronoun *themself* is a recent coinage regarded as a gender-neutral phoric form coreferential with a gender-inclusive indefinite pronoun or a noun phrase whose head is a dual noun. On the other hand, the use of the phoric element *themself* as epicene has not yet reached general consensus. That is the reason why it is followed by a usage note which warns the user that it is still a case of divided usage.

Finally, (14) is the case of singular *they* used as coreferential with a dual antecedent. Once again, this entry is accompanied with a cross-reference leading the user to discover more information on the item considered.

Gender-indexing through phoric elements in OALD9 can be presented in tabular form as below.

Table 3: Gender-indexing phoric elements in OALD9

		Personal		Possessive			
		Subjective	Objective	Determinative	Independent	Reflexive	
Personal	Masculine	1	<i>he</i>	<i>him</i>		<i>his</i>	<i>himself</i>
	Feminine	2	<i>she</i>	<i>her</i>	<i>her</i>	<i>hers</i>	<i>herself</i>
	Epicene	3	<i>s/he, they</i>	<i>them</i>	<i>their</i>	<i>theirs</i>	<i>themselves</i>
Nonpersonal		4		<i>it</i>	<i>its</i>		<i>itself</i>

This table shows both personal gender-indexing in (1) to (3), and nonpersonal gender-indexing in (4), with only the former being considered relevant to the issues of sexist and non-sexist gender-encoding. The inclusion of both personal and nonpersonal gender-indexing is intended to provide a thorough picture of gender differentiation in the central pronouns. The phoric element *he* — together with its different related objective, possessive or reflexive forms — in (1) stands for both specific masculine reference, which is its canonical use, and gender-inclusive reference, that is, gender-indexing subsuming both male and female. In addition, the phoric element *she* — including its related forms — in (2) illustrates feminine reference.

It is worth pointing out, by the same token, that the gender-inclusive meaning of *she* is spoken about in the more-about-gender usage note, but is not lexicalized in OALD9. Finally, *s/he* and *they* in (3) — as well as their related forms — illustrate coreferentiality with gender-inclusive antecedents. Note also that other proposals for conjoining the masculine and the feminine — such as *(s)he*, *he/she*, *he* or *she*, *his/her*, *his* or *her*, etc. — are not lexicalized in OALD9, although they are discussed in the more-about-gender usage note.

3.4 Gender construction through gender-indexing phoric elements

3.4.1 Lexicalization of gender-indexing phoric element

As to gender construction through phoric elements, the analyses have revealed that OALD9 keeps perpetuating the use of generic *he*. Conversely, it makes use of the joint form *s/he*, the singular *they*, and the gender-inclusive reflexive phoric form *themselves*, all of which are explicitly lexicalized in the dictionary. Apart from these, the generic feminine *she*, and the joint forms *(s)he*, *his/her*, *he* or *her* are made mention of in the more-about-gender usage note but are not lexicalized in the word list. Table 4 summarizes lexicalization and lexical gapping of gender-referring phoric elements in OALD9.

Table 4: Lexicalization of gender-indexing phoric elements

N ⁰	Phoric form	Description	OALD9
1	<i>He</i>	Masculine	+
2	<i>He</i>	Generic	+
3	<i>She</i>	Feminine	+
4	<i>She</i>	Generic	(+)
5	<i>(S)he</i>	Generic	(+)
6	<i>S/he</i>	Generic	-
7	<i>They</i>	Singular	-
8	<i>Themselves</i>	Singular	+

This table presents different phoric forms as well as their occurrences in OALD9. The plus sign (+) means that the item is lexicalized in the dictionary while the minus sign (-) means that the item does not occur. By contrast, the plus sign enclosed in parentheses ((+)) means that the item is only mentioned but not lexicalized in the word list. Similarly, it is important to note that, apart from *themselves* which is a reflexive pronoun; the others are subjective personal pronouns and are intended as canonical forms. As such, they stand for other related forms of the central pronoun subclass. Finally, the subjective personal pronouns *(s)he* in (5) and *s/he* in (6) are cases of blending.

These blends offer an insight from the theoretical matter stand and from that of gender-based ideology. From the theoretical point of view, they reveal two new tendencies of spelling blends which we call 'parenthetical blend' as in *(s)he* and 'slashed or slanted blend' as in *s/he*. These add up to the commonest typologies of 'solid blend' as in *brunch* from *breakfast + lunch*, and 'hyphenated blend' as in *wi-fi* (from *wireless fidelity*). However, cases of 'open blends' on the pattern of 'open compound' are still to find. In *s/he* or *(s)he*, the slash and the parentheses are inserted for ideological purposes. That is, they filter the overlap between the use of feminine *she* as generic and these new coinages, intended to be sex-inclusive. Were the slash and parentheses deleted, the debates of using a given specific form as generic would result and reduce its chance of consensual acceptability.

Note that this terminology is coined after that of compounds subsuming solid compounds as in *classroom* where the two components are tied up together, open compounds such as *food box* where the two component parts are separated from each other by a space, and hyphenated compounds as in *short-sighted* whose elements are conjoined by a hyphen. From the perspective of gender ideology, the use of a slant or slash (/) in *s/he* and that of parentheses in

(s)he is ideological as it is intended to preclude the spelling of the blend as *she*, which would be a case of reversing the marking.

Interestingly, this tendency does occur in some writing, most of which by feminist writers. Furthermore, from a theoretical perspective, it can be argued that these blends have introduced a new role of the slant or slash as well as that of parentheses, which is of playing the role of conjoining parts of a blend. This role is coterminous with that of the hyphen. Next, would one dig up the usual meaning associated with these writing mechanics, that is separating alternatives constituent for the slash or stroke and enclosing optional information for the parentheses (Quirk et al. 1985), this would send us back to square one.

To conclude, Table 4 above has important implications for the understanding of the central pronouns functioning as gender-inclusive phoric elements. Therefore, the current state of these phoric elements may be presented as in Table 5 below.

Table 5: Degendering phoric elements: Current state of the art

	Achievements	Personal		Possessive		Reflexive
		Subjective	Objective	Determinative	Independent	
OALD9	1	* <i>he</i>	<i>him</i>	<i>his</i>		<i>himself</i>
	2	* <i>she</i>	<i>her</i>	<i>her</i>	<i>hers</i>	<i>herself</i>
	3	(s) <i>he, s/he</i>	<i>him/her</i>	<i>his/her</i>	<i>his/hers</i>	<i>him-/herself, themself</i>
	4	<i>they</i>	<i>them</i>	<i>their</i>	<i>theirs</i>	<i>themselves</i>
	5	<i>SE</i>	<i>SEM</i>	<i>SES</i>		<i>SEMSELF</i>
	6	<i>co, tey, hesh, thon, xe, E, po</i>				

This table presents the current configuration of the central pronouns actually used and/or simply proposed in the degendering of the English language. It comprises three different strata. The first, which coincides with the view still perpetuated in OALD9, is represented by option (1) in the table. That is, it is all about the use of the masculine phoric element as generic. Two options may be noted. First, if native speakers have actually stopped using the generic *he*, and OALD9 still somehow maintain it, it may be said to be norm-prescriptive in maintaining or perpetuating the status quo. The other side of the coin is that this dictionary might be said norm-descriptive in reflecting actual usage. This may transpire through a careful consideration of the sentence *Every child needs to know that he is loved* (OALD9: 723), which has been taken from naturally occurring evidence. The second stratum comprises options (2) to (4) which are specific to OALD9. Option (2) is all about the reversal of the marking or the use of the feminine as generic. This option is likely to raise contention associated with option (1); that is why both of them are starred.

Since OALD9 simply describes option (2), it is said to be norm-descriptive in this regard. Options (3) and (4) are about the joint use of masculine and feminine phoric elements and the use of the singular *they*. In option (3) the blend *s/he* is lexicalized while *(s)he* is simply mentioned. It is also important to note the emergence of the gender-inclusive reflexive phoric element *themselves* in the same option. Finally, the third stratum constitutes lexical gaps or forms not yet attested. It is made up of option (5) which has the subjective as well as its objective, possessive, and reflexive counterparts; and option (6) whose elements are cross-functional. These gaps would be justified by their absence from the corpus since they are artificially coined and not actual specimens of language in use.

3.4.2 Dictionary use and exclusion of gender-indexing phoric elements

The earlier discussion was devoted to lexicalization or gapping of gender-indexing phoric elements as well as their treatment in OALD9. This subsection attempts to address actual examples of use of gender-indexing phoric elements in order to point out which ones of such forms are actually used in this dictionary. This would reveal which stance the dictionary takes in this regard. To achieve this task, some idiomatic expressions were randomly selected as illustrations given the presence of a gender-referring phoric element in their structure. Consideration will be made of both their definiendums and the definiens.

	<i>Definiendum</i>	<i>Definiens</i>
(15)	<i>Every man for himself</i>	<i>(saying) People must take care of themselves and not give or expect any help (OALD9: 945)</i>
(16)	<i>He who laughs last laughs longest</i>	<i>(saying) used to tell sb not to be too proud of their present success; in the end another person may be more successful (OALD9: 880)</i>
(17)	<i>Leave sb to their own devices</i>	<i>To leave sb alone to do as they wish, and not tell them what to do (OALD9: 420)</i>
(18)	<i>Put sb/sth through their/its paces</i>	<i>to give sb/sth a number of tasks to perform in order to see what they are capable of doing (OALD9: 1105)</i>
(19)	<i>Walk sb off their feet</i>	<i>(informal) to make sb walk so far or so fast that they are very tired (OALD9: 1751)</i>

Regarding actual use of gender-indexing phoric elements in the definiendum, it can be argued that only the generic masculine, examples (15) and (16), and the singular *they*, examples (17) to (19), are attested. By contrast, in the definiens, gen-

eralized pluralization, examples (15), and the singular *they*, examples (16) to (19) are actually used. However, it is worth noting that the gender-indexing phoric forms in the definiendums were probably taken as such from a corpus as naturally occurring specimens rather than lexicographer-intentionally selected. Conversely, those in the definiens may be considered the other way around as they are part of the lexicographers' choice of the defining vocabulary. The forthcoming section summarizes and concludes the overall discussion.

4. Conclusion

This paper was based on the premise that non-sexism is one of obvious proofs of the interaction between social and cultural changes in the English speech community. Therefore, for illustrative purposes, it drew on OALD9 in order to address the degendering of phoric elements and the related efforts to capture and express social and cultural changes in the English language. It has been found out that OALD9 draws on the potential provided by the English language in order to put forward a variety of degendering strategies, to the exclusion of those artificially proposed by various scholars.

On this ground, OALD9 is to be regarded as an invaluable tool for non-native speakers insofar as it draws their attention to and raises their awareness of the issues which are often left unaddressed in common (elementary) grammar books. However, it is important to add that this dictionary does not propose a monolithic solution to the degendering of phoric elements, nor does it take into account all the proposed degendering strategies. For example, the use of feminine as generic is not even mentioned. Besides, despite the coinage of epicene forms, the generic masculine phoric elements still appear. Although this state of affairs may show that the dictionary aims at being neutral in just describing language in use, it also reveals that it takes a stance in privileging some strategies to the detriment of others. For instance, in the defining vocabulary, the use of generalized pluralization and the singular *they* are intentionally favoured by the lexicographers.

This situation has three main insightful theoretical consequences. First, there is still a real gap between rules and actual practice. Second, non-sexism is still a usage problem insofar as all the language users do not agree about the rules. Even native speakers have to learn these new conventions — a violation of the psycholinguistic saying according to which language is what a native speaker says it is. Finally, the use of more than one structure to achieve non-sexism is seen as making fluency difficult and writing cumbersome.

References

- Abdel-Naby, M. and H. Hilfi. 1989. Introducing the Unisex Pronoun. *English Teaching Forum* 27(3): 42-43.
- Bourdieu, P. 1992. *Language and Symbolic Power*. Cambridge: Polity Press.
- Cook, V.J. 1988. *Chomsky's Universal Grammar: An Introduction*. Oxford: Basil Blackwell.
- Fairclough, N. 2003. *Analysing Discourse: Textual Analysis for Social Research*. New York: Routledge.
- Halliday, M. 2004. *An Introduction to Functional Grammar*. London: Routledge.
- Halliday, M.A.K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hankamer, J. and I. Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 7(3) 391-426.
- Herbert, R.K. and B. Nykiel-Herbert. 1986. *Explorations in Linguistic Sexism: A Contrastive Sketch*. Available at: <http://wa.amu.edu.pl/psicl/files/21/05Herbert&Nykiel-Herbert.pdf>.
- Hornby, A.S. 2015. *Oxford Advanced Learner's Dictionary of Current English*. Ninth Edition. Oxford: Oxford University Press.
- Huang, Y. 2010. Pragmatics of Anaphora. Cummings, L. (Ed.). 2010. *The Routledge Pragmatics Encyclopedia*: 9-13. London/New York: Routledge.
- Huddleston, R. 1984. *Introduction to the Grammar of English*. Cambridge/New York: Cambridge University Press.
- Lazar, M.M. (Ed.). 2005. *Feminist Critical Discourse Analysis: Gender Power and Ideology in Discourse*. New York: Palgrave MacMillan.
- Miller, C. and K. Swift. 1977. *Words and Women: New Language in New Times*. Garden City, NY: Anchor Press/Doubleday.
- Mills, S. 1995. *Feminist Stylistics*. London: Routledge.
- Mulamba, N.G. and N. Tshimanga. 2006. Non-sexism as a Usage Problem in English. *Annales de l'I.S.P.-Mbuji mayi* 14: 133-152. Mbuji mayi: CANDIP-I.S.P.
- Musau, G.M. 1995. *Sexism and Africanisms*. English for Africa Series (Book Three and Book Four). Licence Dissertation. Mbuji mayi: I.S.P., English Department.
- Neveu, F. 2011. *Dictionnaire des sciences du langage*. Second revised and expanded edition. Paris: Armand-Colin.
- O'Grady, W., M. Dobrovolsky and M. Aronoff. 1993. *Contemporary Linguistics: An Introduction*. Second Edition. New York: St. Martin's Press.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Sunderland, J. (Ed.) 2006. *Language and Gender: An Advanced Resource Book*. London: Routledge.
- Thomson, A.J. and A.V. Martinet. 1980. *A Practical English Grammar*. Third Edition. Oxford: O.U.P.
- Tshimanga, F.N.C. 2012. Sexisme dans les Anglicismes. *LITERA* 25(2): 1-31.
- Tshimanga, F.N.C. 2016. Gender Bias in Bantu Languages: The Case of Cilubà (L31). Abbou, J. and F.H. Baidier (Eds.). 2016. *Gender, Language and the Periphery: Grammatical and Social Gender from the Margins*: 129-164. Amsterdam: John Benjamins.
- Tshimanga, F.N.C. 2017a. Degendering the English Language: Achievements and Gaps, The Case of *Oxford Advanced Learner's Dictionary*. Third Edition. Unpublished M.A. Dissertation. Kinshasa: University of Kinshasa, Faculty of Arts — Department of English Letters and Civilization.

- Tshimanga, F.N.C.** 2017b. Lexical Gender-Indexing, Sexism, and Non-Sexism in The *Oxford Advanced Learner's Dictionary*. Third Edition. *Annales de la Faculté des Lettres et Sciences Humaines*, Université de Kinshasa. Vol. 18: 67-91.
- Tshimanga, F.N.C.** 2018. Gender-indexing Phoric Elements in Idioms: Peeping into *Oxford Advanced Learner's Dictionary*. Third Edition. *Journal of Advanced Linguistic Studies* 7(1-2): 121-144..
- Van Roey, J.** 1990. *French-English Contrastive Lexicology: An Introduction*. Louvain-La-Neuve: Peeters.

Teaching Lexicography as a University Course: Theoretical, Practical and Critical Considerations

Tvrtko Prčić, *Department of English, Faculty of Philosophy, University of Novi Sad, Serbia (tvrtko.prcic@gmail.com)*

Abstract: This paper offers an account of a university course in Lexicography, designed particularly for advanced EFL students. The discussion and accompanying material derive from around fifteen years of the author's experience in developing and teaching this course. The aims of the paper are twofold: firstly, to draw attention to the fact that lexicography today is no longer just 'the art and craft of' dictionary making of yesteryear but a fully-fledged applied linguistic discipline, with its own theory, methodology and practice, some of which is teachable as a dedicated academic subject; and secondly, to encourage the teaching of practical and theoretical lexicography to university students by putting forward concrete proposals. The exposition is divided into five parts, as follows: Section 1 brings some opening remarks, including that on the scientific status of lexicography; Section 2 reviews theoretical aspects regarding the design of this course; Section 3 describes practical aspects regarding the implementation of the course; Section 4 examines critical aspects regarding the evaluation of the course; and Section 5 recapitulates the main points of the paper and projects the modifications to the course in the future.

Keywords: LEXICOGRAPHY AS A UNIVERSITY COURSE, TEACHING LEXICOGRAPHY, EFL UNIVERSITY STUDENTS, THEORETICAL CONSIDERATIONS, DESIGN OF THE COURSE, PRACTICAL CONSIDERATIONS, IMPLEMENTATION OF THE COURSE, CRITICAL CONSIDERATIONS, EVALUATION OF THE COURSE

Opsomming: Die onderrig van leksikografie as universiteitskursus: Teoretiese, praktiese en kritiese oorwegings. In hierdie artikel word verslag gelewer oor 'n universiteitskursus in die leksikografie wat spesiaal ontwerp is vir gevorderde EVT-studente. Die bespreking en gepaardgaande materiaal spruit uit die outeur se ervaring van ongeveer vyftien jaar in die ontwikkeling en onderrig van hierdie kursus. Die doel met die artikel is tweevoudig: eerstens, om die aandag daarop te vestig dat die leksikografie vandag nie meer net die "art and craft" van woordeboekmaak van die verlede is nie, maar 'n volledig toegepaste linguïstiese dissipline met 'n eie teorie, metodologie en praktyk, waarvan gedeeltes onderrig kan word as 'n toegewyde akademiese vak; en tweedens, om die onderrig van die praktiese en teoretiese leksikografie by universiteitstudente met behulp van konkrete voorstelle te bevorder. Die uiteensetting wat in in vyf dele verdeel is, is as volg: Afdeling 1 bevat 'n paar inleidende opmerkings, insluitende dié oor die wetenskaplike status van die leksikografie; Afdeling 2 beskou teoretiese aspekte rondom die ontwerp van die kursus; Afdeling 3 beskryf die praktiese aspekte rondom die implementering van die

kursus; Afdeling 4 ondersoek die kritiese aspekte rakende die evaluering van die kursus; en Afdeling 5 som die belangrikste argumente van die artikel op en voorspel toekomstige veranderings aan die kursus.

Slutelwoorde: DIE LEKSIKOGRAFIE AS UNIVERSITEITSKURSUS, DIE ONDERRIG VAN LEKSIKOGRAFIE, EVT-UNIVERSITEITSTUDENTE, TEORETIESE OORWEGINGS, ONTWERP VAN DIE KURSUS, PRAKTIESE OORWEGINGS, IMPLEMENTERING VAN DIE KURSUS, KRITIESE OORWEGINGS, EVALUERING VAN DIE KURSUS

1. Opening remarks: lexicography as a university course

Not many people thought at the time that the short and unassuming title "Lexicography as an Academic Subject" of a conference presentation and later a paper by J. Sinclair (1984) was to become a pioneering and seminal work that, as if institutionally, triggered the process of a radical change in perspective on lexicography — from what was then largely perceived as "the art and craft of" dictionary making (cf. Landau 2001) towards a rediscovery and redefinition of its true identity and status. The ongoing scholarly debate, which takes place in journals, conference proceedings, books and handbooks, revolves around three sets of interrelated key issues: firstly, the scope of lexicography; secondly, the scientific status of lexicography; and thirdly, the teaching of lexicography as an academic subject (for a representative cross section, see Adamska-Salaciak 2019; Atkins 2008; Bergenholtz 2012; Bergenholtz and Gouws 2012; Bogaards 2010; Gouws 2012; Hartmann 2001; Magay 2000; Margalitadze 2018; Martynova et al. 2015; Nkomo 2014; Tarp 2008, 2012, 2018; Wiegand 1984).

In view of the current debate, this paper aims at making a contribution to the third set of key issues, by proposing a synopsis of a ready-to-teach university course in Lexicography, with comments on the theoretical, practical and critical aspects of its creation, development and realization. The paper describes the course in Lexicography, which has been taught since the academic year 2006/07, at the Department of English, Faculty of Philosophy, University of Novi Sad, initially as part of the BA programme in English Language and Literature, and now of the MA programme in English Language and Linguistics. The course takes a 12-week semester and is taught 4 classes per week to between 15 and 30 students. Combining practical and theoretical components, it is jointly taught by Tvrško Prčić, the creator and developer of the course and author of all of its teaching materials, the bulk of which is included in this paper, and Gordana Lalić Krstin, the associate teacher and student adviser.

Concerning the scientific status of lexicography and, consequently, its teachability at the university, this course builds around a range of criteria, which fully inform its contents. In a nutshell, lexicography can be, and is to be, regarded as a science — specifically, an applied linguistic science — by virtue

of exhibiting customary properties of science and scientific/scholarly research, the most important being the following (cf. Prčić 2018: 28-29):

- precisely delineated scope, topics and aims of research,
- theoretical and methodological apparatus, and technical terminology,
- empirical data and research methods with easily made generalizations,
- research results able to be made known at the practical level in the shape of ready dictionaries and at the theoretical, methodological and practical levels in the shape of scholarly papers, books, handbooks, conference presentations, etc.,
- applicability of research results in everyday practice, including popularization of scholarly achievements,
- applicability of theoretical and practical research results in university courses, at MA and PhD studies, including the writing of MA theses and PhD dissertations,
- contribution to the advancement of lexical, lexicological and lexicographic research, as well as of knowledge of language use and general knowledge, and
- existence of professional associations, scholarly journals, books, handbooks, conferences, books of proceedings and bibliographies of dictionaries and scholarly works.

The remainder of the exposition will be organized into four parts, as follows: matters of theoretical nature, relative to the design of this course, will be presented Section 2; matters of practical nature, relative to the implementation of the course, will be dealt with in Section 3; matters of critical nature, relative to the evaluation of the course, will be looked at in Section 4; and finally, by way of a conclusion, a summary of the main points accompanied by an overview of the future modifications to the course will be offered in Section 5.

2. Theoretical considerations: design of the course

This section explains the theoretical principles which underpin the design of this course in its various dimensions. The discussion will focus on three major aspects, in the following way: the preliminaries, with the course's objectives, prerequisites and requirements (2.1), the syllabus, with the sets of topics to be covered in the course (2.2), and the micro-dictionary project, with a sketch of an exercise in practical and theoretical lexicography (2.3).

2.1 The preliminaries

Starting with the objectives of this course, in general terms, its aim is to highlight and to approach to the students dictionaries as objects of study in their own right, as opposed to the utilitarian view of dictionaries as tools and/or

aids used in reading/hearing, writing/speaking and translating activities. Put differently, with the explanatory power of automotive imagery, this course aims not to teach people how to drive a car but how to know thoroughly its parts and their workings, and, ultimately, to disassemble and reassemble a car successfully. Intent on familiarizing the students with dictionaries from within, rather than from without, this course has the following specific objectives to accomplish:

- to get the students acquainted with basic terms and concepts pertaining to practical and theoretical lexicography,
- within the domain of practical lexicography, to get the students acquainted with principles of modern monolingual and bilingual lexicography, and practices of dictionary making, and
- within the domain of theoretical lexicography, to get the students acquainted with aspects of scientific examination of dictionaries and people professionally involved with them, and methods of dictionary research.

The prerequisites to this course, which make it possible for the students to choose, attend and profit from it, are threefold:

- necessary for practical and theoretical lexicography is prior knowledge of lexicology, and especially:
 - organization of the lexicon (building on, but not restricted to, Aitchison 2012; Carter 2012; Crystal 2019; Lehrer 1974; McCarthy 1990),
 - meanings and uses of words and their paradigmatic and syntagmatic relations (cf. Cruse 2010; Kövecses 2010; Lipka 2002; Nida 1975; Prčić 2016; Taylor 2003), and
 - internal structure and formation of (new) words (cf. Bauer 1983, 2003, 2004; Bauer and Huddleston 2002; Dixon 2014; Matthews 1991; Plag 2018; Quirk et al. 1985: Appendix I),
- also necessary for practical and theoretical lexicography is prior knowledge of types of dictionaries and of major dictionaries of English, and especially:
 - dictionaries recommended for the use by EFL students (cf. Prčić 1996), and
 - efficient use of print and online dictionaries in passive and active functions, and in translation (cf. Leaney 2007; Mugglestone 2011), and
- necessary for bilingual and bilingualized lexicography is prior knowledge of basic principles of:
 - contrastive and contact linguistics (cf. Đorđević 2004; Filipović 1986; Furiassi, Pulcini and Rodríguez González 2012; Krzeszowski 1990; Prčić 2019), and
 - translation studies (cf. Baker 1992; Hatim 2001; Hatim and Mason 1990; Newmark 1988; Prčić 2019).

Students wishing to sign up for this course in Lexicography have had the opportunity of taking and passing the undergraduate courses in Lexical Semantics and Pragmatics, Lexical Morphology and General Principles of Translation, which, between them, provide good grounding in the requisite skills.

And lastly, the requirements of this course are also threefold, the joint results of which contribute to the formation of the final mark:

- mastery of the practical and theoretical knowledge acquired during the course, as laid down in the syllabus, and verified and evaluated in a written 4-page 2-hour test (50% of the final mark),
- compilation of a micro-dictionary, as the students' individual lexicographic project (40% of the final mark), and
- attendance at, and active participation in, the classes (10% of the final mark).

In what follows, the syllabus and the micro-dictionary project will each be explained and exemplified in detail.

2.2 The syllabus

In line with the objectives stated above, the syllabus of this course has been designed so as to present a complete and balanced picture of essentials of lexicography, for students at an advanced level of proficiency in English language and linguistics, and has been developed so as to provide a carefully selected variety of lexicographically relevant content considered teachable, because important and interesting enough to the students and, above all, readily discoverable by their own research. Consequently, the syllabus covers topic areas ranging from dictionary-related terminologies and typologies to aspects of dictionary structure and compilation to aspects of dictionary scholarly applications. The focus of this course, to reiterate, is not on dictionaries themselves but on all that is relevant and teachable *about* dictionaries. That is why the syllabus does not cover topic areas with which the students are presupposed to be familiar, because these areas are, as already mentioned, the prerequisites to the course.

With the theoretical boundaries of the course clearly marked out, its contents will now be presented. The syllabus is divided into three thematic parts, which, in turn, are all subdivided into several thematic units. Here are the thematic parts and their associated units in full.

Part 1, headed *Basic Concepts*, consists of three units and comprises topics related to the identity of lexicography and its place within the science of language, and the identity, history and types of dictionaries alongside their place among other reference works:

- UNIT 1 — Scope of lexicography: dictionary making (practical branch) and dictionary research (theoretical branch). Overlaps with other linguistic disciplines: lexicology, grammar, phonology, stylistics, sociolinguistics, psycholinguistics, applied linguistics, corpus linguistics, contrastive linguistics, contact linguistics. Main types of reference works: dictionary, thesaurus, encyclopedia, atlas, almanac, style guide.
- UNIT 2 — Landmarks in English-language lexicography. Major contemporary British, American and other dictionaries of English.
- UNIT 3 — Criteria for classifying dictionaries. Types of dictionaries: general-purpose, encyclopedic, specialized, terminological; print, electronic; monolingual, bilingual, bilingualized, multilingual; native-speaker, learner's (pedagogical); passive/receptive, active/productive; semasiological, onomasiological; hybrid dictionaries.

Part 2, headed *Dictionary Making*, consists of five units and comprises topics related to the internal organization of monolingual and bilingual dictionaries, and the process of compilation of dictionaries:

- UNIT 4 — Stages of the lexicographic process: preparation, data gathering, data processing, editing, publishing; planning and implementation phases.
- UNIT 5 — Components of dictionary structure: macrostructure (word list); outside matter (front, middle, back); megastructure; microstructure (entry): headword, form-related comment (spelling, pronunciation, grammar), content-related comment (definition, usage, examples), etymology; subheadword; mediostructure (cross-referencing); access structure. Reduced dictionary structures.
- UNIT 6 — Dealing with words, phraseological units and bound lexical elements in monolingual dictionaries. Approaches to definition; controlled defining vocabulary in learner's dictionaries.
- UNIT 7 — Monolingual and bilingual lexicography and dictionaries: similarities and differences. Theoretical and microstructural innovations in modern bilingual dictionaries. Semantic and pragmatic approaches to translation equivalence. Dealing with false friends, culture-specific words and anglicisms in bilingual dictionaries. The bilingualized learner's dictionary.
- UNIT 8 — Corpus-based analysis of language and corpus-based lexicography: definition of corpus; types of corpora; major English-language corpora. Uses of corpora in language study; lexicographically relevant research questions; ways of querying a corpus; concordancing; KWIC concordances.

Part 3, headed *Dictionary Research (Metalexicography)*, consists of two units and comprises topics related to people professionally engaged in matters lexicographic and dictionaries as objects of scholarly attention and research:

- UNIT 9 — The four protagonists: the compiler, the user, the teacher, the researcher. Dictionary history. Dictionary typology. Dictionary structure. Dictionary use. Dictionary criticism. Dictionary IT.
- UNIT 10 — The dictionary user; reference needs; reference skills. Dictionary awareness; user training; efficient and creative use of dictionaries in language learning, language teaching, translation, lexical and other research.

2.3 The micro-dictionary project

Complementary to its dimension of imparting and acquiring specialized practical and theoretical knowledge about lexicography, this course has an applied, and creative, dimension as well, in that each student does an individual project aimed at producing a micro-dictionary. The idea behind this project is that, during the process of making the micro-dictionary, the students are given a unique opportunity, within the set framework of an exercise in real-life application of practical and theoretical knowledge, to demonstrate their level of mastery of the essentials of lexicography provided in the course. The project, carried out in agreement and in consultation with the teacher(s), involves two parts: compiling the dictionary itself and writing the text of the front matter to the dictionary.

Front matter consists of three separate components: a typological identification of the dictionary, a description of the five stages of the lexicographic process and a user's guide, all three offering, in a concise overview, technical details about a specific dictionary. On the other hand, the dictionary consists of a word list, in the form of headwords, and obligatory information pertaining to them, in the form of a lexical entry, which may have full or reduced structure, depending on the type of the dictionary; optionally, visual illustrations and back matter may be included, if deemed necessary and/or useful enough. The total number of headwords normally ranges between 15, with full structure, and 30, with reduced structure, and they can be either English only, or English–Serbian, or Serbian–English, or bilingualized.

Regarding the topics of the dictionaries, the students are given *carte blanche* to choose and propose what their dictionaries will be about, in accordance with their personal interests and preferences, which are indeed limited only by their very lively imagination. To spark that imagination, the students are provided with the following checklist of suggested topic areas:

- the headword consists of or deals with: lexical fields, sense relations, metaphors and metonymies, collocations, idioms, neologisms, technical terms, regional words, British-American words, slang words, false friends, culture-specific words, anglicisms, exclamations, prefixes, suffixes, initial combining forms, final combining forms, names (personal, geographical, institutional and other), fictional names (characters in books, films, cartoons, etc.), titles (of books, newspapers, films, songs, etc.).

The micro-dictionary is submittable either in computer-printed comb-bound paper form or, since recently, (to save ever-shrinking shelf room in the teachers' office) in computer-typeset electronic form as a PDF file. At this stage, however, micro-dictionaries are not yet realized in electronic, online or offline, form, because the students lack specialist knowledge in software programming and therefore their focus is placed on classical, pre-electronic, methods of dictionary making. Prior to embarking on the research project, the students receive *Guidelines for Compiling the Micro-dictionary*, which is reproduced in Appendix 1.

3. Practical considerations: implementation of the course

This section presents the practical details relative to the successful implementation of this course. The discussion will focus on two major aspects, in the following way: providing appropriate teaching materials, with an overview of the sources put at the students' disposal (3.1), and addressing the research topics, with an overview of the practices employed by the students and the teachers for working on the topics (3.2).

3.1 Providing appropriate teaching materials

In order that any course, including this one, achieves its objectives and produces desired outcomes, provision of appropriate teaching materials to the students is of utmost importance. By 'appropriate teaching materials' are meant materials carefully prepared and/or selected by the teacher(s) and purposed to furnish reliable, efficient, accessible and exploitable sources of information about various aspects of lexicography. In this particular course the teaching materials serve two complementary functions: firstly, taking the students step-by-step through the syllabus, there are worksheets with assignments; and secondly, offering titles of, and, where necessary, links to, print and electronic sources for the students to consult, there are three reference lists, made up of a reading list, a dictionary list and a resource list. All these materials will now be described in some detail.

Worksheets with assignments contain questions and instructions matching the thematic units of the syllabus. The assignments, 60 of them altogether, have been specially devised and developed for this course and phrased in a way to prompt the students to some creative action, typically involving advance reading of texts and/or exploring of dictionaries or websites followed by making summaries of the findings and reporting the results either of some online research, or of comparing and contrasting certain dictionary properties, or of voicing informed opinions, or of expressing critical judgements, or of simply providing answers to questions. It may be as well to point out that each of the 60 assignments, often consisting of several interrelated questions and/or

instructions, can be viewed as a micro-topic affording a focused insight into a facet of lexicography. Progressing through the assignments from number 1 to number 60 gradually builds up a composite picture of practical and theoretical lexicography, as set forth in the syllabus. And when the final assignment has been completed that picture becomes complete. The integral text of the assignments, entitled *Research Topics on Lexicography*, is available in Appendix 2; the headings inserted signal the connection of sets of assignments to individual thematic units of the syllabus.

In the province of the reference lists, the reading list, as first, contains titles of textbooks and articles recommended for the course, and is divided into three categories:

- Obligatory reading brings first *Teaching and Researching Lexicography* (Hartmann 2001), selected as a coursebook, because it is admirably suited to the scope of this course, whose creation it has motivated to a great extent and inspired part of the assignments in *Research Topics on Lexicography*. Complementing the coursebook are selected chapters from *Lexicographic Description of English* (Benson et al. 1986), *Corpus Linguistics. Investigating Language Structure and Use* (Biber et al. 1998) and *Engleski u srpskom [English within Serbian]* (Prčić 2019), and a specialized paper on hybrid dictionaries (Hartmann 2005).
- Further reading brings titles recommended for expanding the students' knowledge and including notably the following: Béjoint 2000, 2010; Cowie 2002, 2009; Hartmann and James 1998; Landau 2001; McEnery and Wilson 2001; Svensén 2009; Teubert and Čermáková 2007; Yong and Peng 2007; Zgusta 1971.
- Advanced reading brings titles recommended for gaining highly specialist knowledge and including notably the following comprehensive handbooks: Atkins and Rundell 2008; Biber and Reppen 2015; Durkin 2015; Fontenelle 2008; Fuertes-Olivera 2018; Jackson 2013; Van Sterkenburg 2003.

Secondly, the dictionary list contains titles of dictionaries recommended for exploitation in this course. It lays emphasis on the 'fabulous five' advanced learner's dictionaries — *OALD*, *LDOCE*, *COBUILD*, *CALD* and *MEDAL*, in print and electronic forms, and on the online editions of renowned and respected British and American medium-size native-speaker, or collegiate, general-purpose dictionaries. The full list of these titles is available in the *References, A. Dictionaries* section.

And lastly, the resource list contains an inventory of websites recommended for exploration in this course. Grouped under 15 headings, the websites to be visited serve as a supplementary, supporting source of information to the students, especially in the areas of discovering about the history of major English-language dictionaries and of looking into electronic corpora. The full list of these websites, entitled *Internet Resources for Research Topics*, is available in Appendix 3.

Viewed as a whole, the three reference lists and the assignments constitute a meticulously thought-out, self-contained and stimulating pack of teaching materials provided for this course in Lexicography and is usable equally in teaching and in student dictionary making.

3.2 Addressing the research topics

Methodologically speaking, this course has been conceived with a heuristic approach to teaching and learning in mind (cf. Kumaravadivelu 2003; Sale 2015; Seliger 1975; Takimoto 2008), so as to be both student-centred and student-friendly in fostering the students' active acquisition of information by guiding them step-by-step through the topics of the syllabus. This inherently self-study method puts the students in control of the amount of knowledge they gain, of the ways in which they gain it and of the pace at which they gain it. What is especially important, the students are given an opportunity to compare the knowledge they have acquired with the knowledge acquired by their fellow-students, to each other's mutual benefit.

Activities of gaining knowledge are normally done at home, whereas those of comparing are performed in class. Self-study at-home activities comprise advance reading of texts and advance observing of dictionaries, and subsequent writing of short pieces of the kind demanded by assignments on the worksheets. In-class activities comprise discussions about the findings among fellow-students and/or between the students and the teachers, concise oral reports by the students and questions posed either by the students or the teachers.

As it may have been noticed in the *Research Topics on Lexicography* (reproduced in Appendix 2), the nature of specific assignments is determined by their formulations. More precisely, assignments framed as instructions (like Assess the various ranges ..., Topic 9) and as questions (like What is the main purpose ...?, Topic 20), typically indicate explicit at-home activities and, at the same time, implicit in-class activities, since the latter deal with collective assessments of the individual results of the former. Exceptionally, however, certain kinds of at-home activities can be conducted in class, but this practice is to be employed sparingly and with a good reason, because otherwise the concept of this course being based on the students' advance preparation would be spoilt. In addition, assignments containing reference to print or electronic sources represent exercises in assisted (guided) research, whereas those with no reference to sources are exercises in non-unassisted (free) research, which calls for an even greater involvement on the part of the students. And the micro-dictionary compilation project is an exercise in mixed research, because it starts from the general references provided and continues with the sources ferreted out by the students.

Finally, a glimpse at the roles of the two protagonists of this course. It will have definitely become apparent that the students here are cast to play the lead role, in that, firstly, it is them who actively earn their knowledge in practical

and theoretical lexicography, exchange bits of it with their fellow-students and receive appropriate feedback and proper guidance from the teachers; and, secondly, it is the students who demonstrate their proficiency in essentials of lexicography by devising and creating their own first-ever micro-dictionary. Complementary to this, the teachers fill the guiding role, in that they guide the students, more or less visibly, through the intricacies of lexicography, firstly, by carefully organizing the course; secondly, by carefully providing appropriate teaching materials; thirdly, by carefully overseeing the students' micro-dictionary projects; and fourthly, by carefully moderating discussions in class, making effective and timely interventions relating to explanations of difficult points, adjudications on disputed points, pertinent questions to provoke further discussions and exchanges, and, most significantly, constructive feedback to keep the students interested and motivated both in acquiring lexicography and in implementing the knowledge thus acquired. According to this conception, the teacher is seen as an organizer, coordinator, stimulator and explainer, and not an erstwhile boring soliloquist deliverer of facts to a bored, passivized and unenthusiastic audience.

4. Critical considerations: evaluation of the course

This section shares critical insights into the reception and appraisal of this course on the part of both the students and the teachers. The discussion will focus on two major aspects, in the following way: student satisfaction, with empirical results of an evaluation questionnaire filled in at the end of the course (4.1), and teacher satisfaction, with critical and self-critical observations and assessments (4.2).

4.1 Student satisfaction

Judging by their actions and reactions, the students enjoy taking this course. This can be seen, firstly, in the regular attendance of the classes by their great majority; secondly, in their predominantly keen interest in the matters practical and theoretical being covered; thirdly, in the assiduity with which they do their advance preparation for the classes; and, fourthly, in the extraordinary creativity, dedication and pride with which they produce their first micro-dictionaries.

Substantiated empirically, the results of the End-of-course Feedback Questionnaire, filled out anonymously in January 2019, by the latest generation of 21 students, regular attendees of the course, have shown that, on a scale of four, most of them rate this course quite favourably — as 'very useful' (8 respondents), 'useful' (11), 'not useful' (1), 'waste of time' (0), with one circling none of the options.

In response to the question What, if anything, do you like about this course?, the students have, almost without exception, selected the micro-dic-

tionary project as their personal high point, followed by praise of the way the course is organized and the material handled by the teachers. On the other hand, the question *What, if anything, should be improved in this course?*, has brought to the fore that many students still want more practice and less theory in the teaching process as well as their still greater involvement in the course. Asked about the topic areas they find not so useful and stimulating, and thus dispensable, around two thirds of the respondents have indicated, some of them quite strongly, historical aspects of dictionaries and lexicography. In contrast, as especially useful and stimulating topic areas, a half of the students have mentioned the stages of the lexicographic process and, even more, comparison of the realizations in various dictionaries of the elements of the micro-structure.

It may be interesting to reveal that the provocative question *What do you think of the explore-resources-on-your-own-ahead-of-the-class method of work in this course?*, with three options to choose from, viz. 'refreshing and highly efficient' — 'don't know, rarely used it myself' — 'prefer old-style teacher-talking method' has elicited 7 vs 10 vs 4 votes, respectively. And lastly, turning to student self-evaluation, when enquired about *How satisfied are you with your own performance in the dictionary-making project?*, the option 'quite satisfied' has been underlined by 12 students, 'so-so' by 8, 'could/should have been better' by zero, with one underlining none of the options.

As it can be concluded from the teachers' observations and the students' feedback answers, they are, on the whole, satisfied with this course in many respects. However, their critical comments and constructive suggestions drive home an honest and important message to the teachers and provide major input into the planned refinements and adjustments to the course, especially in terms of the students' increased active involvement in its at-home and in-class implementation.

4.2 Teacher satisfaction

Judging by their own critical and self-critical eyes, the teachers find this course a pleasure to teach. It is surely not only because of the positive feedback from the students. It is because, in addition to the positive student feedback, all three objectives of the course have been accomplished, in some cases with flying colours, as demonstrated by the following data about the fulfilment of the requirements of this course, set out earlier:

- the written 4-page 2-hour test, assessing the students' practical and theoretical knowledge of lexicography, usually receives marks between 8 and 10 (on a 6-point scale, from 5, fail, to 10, outstanding), with 9 being the most frequent mark overall,
- the micro-dictionary project normally receives 10 or 9 (and rarely 8), and
- attendance at, and active participation in, the classes varies between 7 and 10,

which, taken together, average out to the final mark of 9 in Lexicography, not only with the latest generation of students but similarly so with previous ones as well.

The fact that not every student is eager to do their homework, especially by the self-study method still off-putting to some, nor to attend the classes regularly, nor to take part in the discussions is, and remains, the weakest spot of this course. And it is the teachers' sole responsibility to find convenient ways to motivate, interest and involve such students more actively, probably by first approaching them individually and handing them out tailor-made assignments to better understand, maybe attempt and eventually even take to the self-study method.

In contrast, the micro-dictionary, both while it is being produced and when it has been finished, is by far the peak of this course — equally for the students and for the teachers. At the very beginning of this course, some 15 years ago, the teachers felt apprehensive exactly about the very compilation of the micro-dictionary. And they were quickly disabused by the very first generation of the students. Now, with close to 100 micro-masterpieces of student lexicography, neatly stacked on the shelves in our office, as if waiting to go on some public exhibition, all these projects witness to the creative, imaginative and resourceful traits of their compilers — in terms of the topics selected, the way in which headwords have been processed and the way in which the dictionaries have been realized. To be true, not all of these micro-dictionaries are flawless and irreproachable pieces of lexicography, methodologically, linguistically, technically or typographically. However, all of them, without exception, offer ample and solid proof that their student compilers have got the hang of the ABC of dictionary making. And this is precisely what matters here: the students are given a unique opportunity to come up with their own ideas and to put them into practice. And they seize the opportunity and take maximum advantage of it.

The theoretical, methodological and practical challenges that have led up to this bold piece of lexicography and are all involved in the running of this course, are as follows: conceiving the course, designing its syllabus, preparing and selecting appropriate teaching materials, organizing at-home and in-class student and teacher activities, creating the written test — and, lastly, devising and seeing through the micro-dictionary projects, during which the students' original but vague ideas gradually transform into a concrete dictionary of their own making. It is exactly these components, in their totality, and especially their most impressive net results, that make this course a genuine pleasure to teach.

5. Summing up and looking ahead

In this paper, a synopsis of a university course in Lexicography has been presented from three interrelated angles: firstly, within theoretical considerations,

the focus has been on the design of the course and included discussion of the preliminaries (objectives, prerequisites and requirements), of the syllabus and of the micro-dictionary project. Secondly, within practical considerations, the focus has been on the implementation of the course and included discussion of the teaching materials provided and used, and of the methods used for addressing the research topics. And thirdly, within critical considerations, the focus has been on the evaluation of the course and included discussion of student satisfaction, based on an end-of-course evaluation questionnaire, and of teacher satisfaction, based on critical and self-critical insights. This paper also contains three full lists of references (reading, dictionary, resources) used in the course, as well as three appendixes, with texts of the following: (1) Guidelines for Compiling the Micro-dictionary, (2) Research Topics on Lexicography, and (3) Internet Resources for Research Topics. In this way, the interested readers are faced with three, possibly four, options: to start their own course in Lexicography by adopting the solutions provided here, or by adapting the present solutions to their own teaching needs and wants, or by working out their own solutions from scratch — or they can even opt not to teach Lexicography at all.

Regarding modifications to the course, contemplated for the near future and inspired by the critical evaluation input, both that described above and that accumulated from earlier-taught courses in Lexicography, these modifications are likely to go along four lines:

- increasing and further stimulating active participation of the students in class,
- expanding Unit 7 of the syllabus with topics about e-lexicography and producing online and offline e-dictionaries, by exploiting the tools at the e-dictionary platforms *Lexonomy* and *Webonary* (both for online dictionaries), *Lexique Pro* (for offline dictionaries) and the most powerful *tlTerm, Terminology Software* (for online and offline dictionaries),¹
- expanding Unit 9 of the syllabus with the topic Dictionary culture, which could be broadly defined as an acquired ability to use dictionaries efficiently and a regular habit of resolving all moot points of usage lexical, grammatical, phonological and orthographic by consulting reputable dictionaries and other reference works (cf. Prčić 2018); in doing so, at the same time a proposal is made for an extension of the scope of dictionary research (metalexigraphy), as outlined in Hartmann 2001, and
- lastly, but not quite in the near future, adapting the course to online implementation, by exploiting the tools at *Moodle*,² or a similar open-source learning platform.

To end with, arguably the most necessary and useful at the moment would be the preparation and publication, preferably in electronic form, of an interactive and ever-evolving companion to this course, entitled (provisionally) *Dictionaries in Close-up. An Advanced EFL Student's Resource Book*, that would function as a step-by-step guide through practical and theoretical lexicography, and would

be based on the currently espoused self-study principle copiously supplemented by exercises in gaining hands-on experience throughout.

Endnotes

1. The addresses of the four e-dictionary platforms:
Lexonomy: <https://www.lexonomy.eu/>
Webonary: <https://www.sil.org/dictionaries-lexicography/webonaryorg>
Lexique Pro: <http://www.lexiquepro.com/>
t!Term, Terminology Software: <https://tshwanedje.com/terminology/>
2. The address of the learning platform:
Moodle: <https://moodle.org/>

Acknowledgements

This is a substantially expanded, reorganized and rewritten text of the talk entitled "Teaching Lexicography to EFL University Students: Theoretical and Practical Aspects", co-authored by Gordana Lalić Krstin and Tvrtko Prčić, and co-presented at the *3rd International Conference on English Studies "English Language and Anglophone Literatures Today" (ELALT 3)*, Novi Sad, March 2015. This paper is part of the research on Project No. 178002, entitled *Languages and Cultures in Time and Space*, which is financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia.

For constructive comments and suggestions the author wishes to thank two anonymous reviewers.

References

A. Dictionaries

Printed

- Crystal, D.** 2008. *A Dictionary of Linguistics and Phonetics*. Sixth edition. Oxford: Blackwell.
- Lea, D. and J. Bradbery (Eds.)**. 2020. *Oxford Advanced Learner's Dictionary*. Tenth edition. Oxford: Oxford University Press.
- Fox, Ch. and R. Combley (Eds.)**. 2014. *Longman Dictionary of Contemporary English*. Sixth edition. Harlow: Pearson Education.
- Matthews, P.H.** 2007. *The Concise Oxford Dictionary of Linguistics*. Second edition. Oxford: Oxford University Press.
- McArthur, T. (Ed.)**. 2005. *Concise Oxford Companion to the English Language*. Second edition. Oxford: Oxford University Press.
- McIntosh, C. (Ed.)**. 2013. *Cambridge Advanced Learner's Dictionary*. Fourth edition. Cambridge: Cambridge University Press.

- Quinion, M.** 2002. *Ologies and Isms. A Dictionary of Word Beginnings and Endings*. Oxford: Oxford University Press.
- Rundell, M. (Ed.)**. 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan Publishers.
- Sinclair, J. (Ed.)**. 1991. *Collins COBUILD English Guides 2: Word Formation*. London: HarperCollins Publishers.
- Sinclair, J. (Ed.)**. 2018. *Collins COBUILD Advanced Learner's Dictionary*. Ninth edition. Glasgow: HarperCollins Publishers.

Online

- Affixes: The Building Blocks of English*: <http://www.affixes.org/>.
- American Heritage Dictionary of the English Language, the (AHD)*: <https://ahdictionary.com/>.
- Cambridge Advanced Learner's Dictionary and Thesaurus (CALD)*: <https://dictionary.cambridge.org/>.
- Chambers 21st Century Dictionary (C21CD)*: <https://chambers.co.uk/>.
- Collins Online English Dictionary (COED)*: <https://www.collinsdictionary.com/dictionary/english>.
- Longman Dictionary of Contemporary English (LDOCE)*: <https://www.ldoceonline.com/>.
- Macmillan English Dictionary for Advanced Learners (MEDAL)*: <http://www.macmillandictionaries.com/>.
- Merriam Webster's Collegiate Dictionary (MWCD)*: <https://www.merriam-webster.com/>.
- Merriam Webster's Learner's Dictionary (MWLD)*: <http://learnersdictionary.com/>.
- Oxford Advanced Learner's Dictionary (OALD)*: <https://www.oxfordlearnersdictionaries.com/>.
- Oxford Dictionary of English (ODE)*: <https://www.lexico.com/>.
- Random House Learner's Dictionary of American English (RHLD)*: <http://www.wordreference.com/definition/>.
- Random House Unabridged Dictionary (RHUD)*: <https://www.dictionary.com/>.
- Webster's New World College Dictionary (WNWCD)*: <https://www.yourdictionary.com/>.
- Wiktionary, the Free Dictionary*: <https://en.wiktionary.org/>.
- WordReference Random House Unabridged Dictionary of American English (RHUAE)*: <https://www.wordreference.com/>.

Downloadable

- TheSage: English Dictionary and Thesaurus*: <https://www.sequencepublishing.com/1/thesage/thesage.html>.
- WordWeb*: <https://wordweb.info/free/>.

B. Other literature

- Adamska-Salaciak, A.** 2019. Lexicography and Theory: Clearing the Ground. *International Journal of Lexicography* 32(1): 1-19.
- Aitchison, J.** 2012. *Words in the Mind. An Introduction to the Mental Lexicon*. Fourth edition. Oxford: Wiley-Blackwell.
- Atkins, B.T.S.** 2008. Theoretical Lexicography and Its Relation to Dictionary-Making. Fontenelle, Th. (Ed.). 2008. *Practical Lexicography. A Reader*: 31-50. Oxford: Oxford University Press.

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Baker, M. 1992. *In Other Words. A Coursebook on Translation*. London/New York: Routledge.
- Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, L. 2003. *Introducing Linguistic Morphology*. Second edition. Edinburgh: Edinburgh University Press.
- Bauer, L. 2004. *A Glossary of Morphology*. Edinburgh: Edinburgh University Press.
- Bauer, L. and R. Huddleston. 2002. Lexical Word Formation. Huddleston, R. and G.K. Pullum (Eds.). 2002. *The Cambridge Grammar of the English Language: 1621-1721*. Cambridge: Cambridge University Press.
- Béjoint, H. 2000. *Modern Lexicography. An Introduction*. Oxford/New York: Oxford University Press.
- Béjoint, H. 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Benson, M., E. Benson and R.F. Ilson. 1986. *Lexicographic Description of English*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. 2012. What is a Dictionary? *Lexikos* 22: 20-30.
- Bergenholtz, H. and R.H. Gouws. 2012. What is Lexicography? *Lexikos* 22: 31-42.
- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D. and R. Reppen (Eds.). 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Bogaards, P. 2010. Lexicography: Science Without Theory? De Schryver G.-M. (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*: 313-322. Kampala: Menha Publishers.
- Carter, R. 2012. *Vocabulary. Applied Linguistic Perspectives*. London/New York: Routledge.
- Cowie, A.P. 2002. *English Dictionaries for Foreign Learners. A History*. Oxford: Oxford University Press.
- Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography. Volume I: General-purpose Dictionaries. Volume II: Specialized Dictionaries*. Oxford: Oxford University Press.
- Cruse, A. 2010. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Third edition. Oxford: Oxford University Press.
- Crystal, D. 2019. *The Cambridge Encyclopedia of the English Language*. Third edition. Cambridge: Cambridge University Press.
- Dixon, R.M.W. 2014. *Making New Words. Morphological Derivation in English*. Oxford: Oxford University Press.
- Durkin, Ph. (Ed.). 2015. *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press.
- Dorđević, R. 2004. *Uvod u kontrastiranje jezika*. 6. izdanje. [An Introduction to Contrasting Languages. Sixth edition]. Beograd: Filološki fakultet.
- Filipović, R. 1986. *Teorija jezika u kontaktu. Uvod u lingvistiku jezičnih dodira*. [Theory of Languages in Contact. An Introduction to Contact Linguistics]. Zagreb: Jugoslavenska akademija znanosti i umjetnosti, Školska knjiga.
- Fontenelle, T. (Ed.). 2008. *Practical Lexicography. A Reader*. Oxford/New York: Oxford University Press.
- Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*. London/New York: Routledge.
- Furiassi, C., V. Pulcini and F. Rodríguez González (Eds.). 2012. *The Anglicization of European Lexis*. Amsterdam/Philadelphia: John Benjamins.

- Gouws, R.H.** 2012. Theoretical Lexicography and *the International Journal of Lexicography*. *International Journal of Lexicography* 25(4): 450-463.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Hartmann, R.R.K.** 2005. Pure or Hybrid? The Development of Mixed Dictionary Genres. *Facta Universitatis. Series: Linguistics and Literature* 3(2): 193-208. Also accessed at: <http://facta.junis.ni.ac.rs/lal/lal2005/lal2005-06.html> (20 March 2020).
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. New York/London: Routledge.
- Hatim, B.** 2001. *Teaching and Researching Translation*. Harlow: Pearson Education.
- Hatim, B. and I. Mason.** 1990. *Discourse and the Translator*. London: Longman.
- Jackson, H. (Ed.).** 2013. *The Bloomsbury Companion to Lexicography*. London/New York: Bloomsbury Academic.
- Kövecses, Z.** 2010. *Metaphor. A Practical Introduction*. Second edition. Oxford: Oxford University Press.
- Krzyszowski, T.** 1990. *Contrasting Languages. The Scope of Contrastive Linguistics*. Berlin/New York: Mouton de Gruyter.
- Kumaravadivelu, B.** 2003. *Beyond Methods. Macrostrategies for Language Teaching*. New Haven: Yale University Press.
- Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Leaney, C.** 2007. *Dictionary Activities*. Cambridge: Cambridge University Press.
- Lehrer, A.** 1974. *Semantic Fields and Lexical Structure*. Amsterdam/London: North-Holland / New York: American Elsevier.
- Lipka, L.** 2002. *English Lexicology. Lexical Structure, Word Semantics, and Word-Formation*. Third edition. Tübingen: Gunter Narr.
- Magay, T.** 2000. Teaching Lexicography. Heid, U., S. Evert, E. Lehmann and Ch. Rohrer (Eds.). 2000. *Proceedings of the 9th EURALEX International Congress EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*: 443-451. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Margalitadze, T.** 2018. Once Again Why Lexicography Is Science. *Lexikos* 28: 245-261.
- Martynova, I.N. et al.** 2015. Modern Approaches to Teaching English Lexicography at the University of Freiburg, Germany. *Journal of Sustainable Development* 8(3): 220-226.
- Matthews, P.H.** 1991. *Morphology*. Second edition. Cambridge: Cambridge University Press.
- McCarthy, M.** 1990. *Vocabulary*. Oxford: Oxford University Press.
- McEney, T. and A. Wilson.** 2001. *Corpus Linguistics. An Introduction*. Second edition. Edinburgh: Edinburgh University Press.
- Mugglestone, L.** 2011. *Dictionaries. A Very Short Introduction*. Oxford: Oxford University Press.
- Newmark, P.** 1988. *A Textbook of Translation*. New York/London: Prentice Hall International.
- Nida, E.A.** 1975. *Componential Analysis of Meaning*. The Hague/Paris: Mouton.
- Nkomo, D.** 2014. Teaching Lexicography at a South African University. *Per Linguam* 30(1): 55-70.
- Plag, I.** 2018. *Word-Formation in English*. Second edition. Cambridge: Cambridge University Press.
- Prčić, T.** 1996. Jednojezični rečnici engleskog jezika opšte namene: selektivna bibliografija s komentarima. [Monolingual General-purpose Dictionaries of English: A Selected Annotated Bibliography]. *Glossa* 2: 51-57.

- Prčić, T.** 2016. *Semantika i pragmatika reči*. 3., elektronsko, izdanje. [Semantics and Pragmatics of the Word. Third, electronic, edition]. Novi Sad: Filozofski fakultet. Accessed at: <http://digitalna.ff.uns.ac.rs/sadrzaj/2016/978-86-6065-356-9> (20 March 2020).
- Prčić, T.** 2018. *Ka savremenim srpskim rečnicima*. 1., elektronsko, izdanje. [Towards Modern Serbian Dictionaries. First, electronic, edition]. Novi Sad: Filozofski fakultet. Accessed at: <http://digitalna.ff.uns.ac.rs/sadrzaj/2018/978-86-6065-454-2> (20 March 2020).
- Prčić, T.** 2019. *Engleski u srpskom*. 3., elektronsko, izdanje. [English within Serbian. Third, electronic, edition]. Novi Sad: Filozofski fakultet. Accessed at: <http://digitalna.ff.uns.ac.rs/sadrzaj/2019/978-86-6065-512-9> (20 March 2020).
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik.** 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Sale, D.** 2015. *Creative Teaching. An Evidence-Based Approach*. Singapore: Springer.
- Seliger, H.W.** 1975. Inductive Method and Deductive Method in Language Teaching: A Re-examination. *International Review of Applied Linguistics (IRAL)* 13(1): 1-18.
- Sinclair, J.** 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Ed.). 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 3-12. Tübingen: Max Niemeyer.
- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Takimoto, M.** 2008. The Effects of Deductive and Inductive Instruction on the Development of Language Learners' Pragmatic Competence. *The Modern Language Journal* 92: 369-386.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S.** 2012. Do We Need a (New) Theory of Lexicography? *Lexikos* 22: 321-332.
- Tarp, S.** 2018. Lexicography as an Independent Science. Fuertes-Olivera, P. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 19-33. London/New York: Routledge.
- Taylor, J.R.** 2003. *Linguistic Categorization. Prototypes in Linguistic Theory*. Third edition. Oxford: Oxford University Press.
- Teubert, W. and A. Čermáková.** 2007. *Corpus Linguistics. A Short Introduction*. London/New York: Continuum.
- Van Sterkenburg, P. (Ed.).** 2003. *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Tübingen: Max Niemeyer.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague: Mouton.

Appendix 1: Guidelines for compiling the micro-dictionary

1. Structure of the dictionary:
 - Obligatory part 1: title page; consult the template for the layout and typography of the required elements.
 - Obligatory part 2: front matter, which should contain the following:
 - (1) a typological identification of the dictionary, headed *1. Typological Identification*,
 - (2) a description of the five stages of the lexicographic process as implemented in the given project, headed *2. Stages of the Lexicographic Process*; it should end with a list of all print and electronic sources used, and their full bibliographic data, headed *Print Sources* and *Electronic Sources*, respectively, and
 - (3) a user's guide, headed *3. User's Guide*, which should contain information on (a) the target audience, (b) the number of the headwords included, (c) the organization of an entry, with the presentation system of the micro-structural elements clearly shown.
 - Obligatory part 3: the dictionary, headed *The Dictionary*.
 - Optional part 1: visual illustrations, captioned as appropriate.
 - Optional part 2: back matter, containing an appendix, headed as appropriate.
 - The basic language of the project is English.
2. Technical matters:
 - The project should be submitted either as a computer-printed hard copy (in comb binding) or as a PDF file.
 - The copy submitted is non-returnable, so make another one for your own reference.
 - IPA fonts can be downloaded from: <http://www.phon.ucl.ac.uk/home/wells/fonts.htm> (the file *Doulos ipa-samd roman* is enough; to install the font: Control Panel / Fonts / File / Install New Font; to insert characters in Word: Insert / Symbol / Font, choose the name of the font and then the appropriate character).
3. Grading of the project will take into account the following:
 - the overall organization of the front matter and the dictionary, especially the types of information provided,
 - the layout and typography of the entries, especially the systematicity, efficiency and user-friendliness of the presentation of the information in the dictionary,
 - linguistic and factual accuracy of the information in the dictionary, especially the spelling, the definitions and the translations,
 - how well researched and realized the entire project is.
 - IMPORTANT: During the evaluation discussion, you may be asked questions of practical and/or theoretical nature in connection with the project.
 - VERY IMPORTANT: All cases of plagiarism, i.e. using material and/or ideas from explicitly unacknowledged sources, will be heavily penalized!

Appendix 2: Research topics on lexicography

SPANNING UNIT 1 OF THE SYLLABUS:

1. What are the two branches of lexicography (Hartmann 2001: Ch 1)? Discuss the overlaps of lexicography with other linguistic disciplines.
2. Define and exemplify the six main types of reference works: (a) dictionary, (b) thesaurus, (c) encyclopedia, (d) atlas, (e) almanac, (f) style guide. Could you identify the similarities and differences between them?

SPANNING UNIT 2 OF THE SYLLABUS:

3. The following ten dictionaries can be said to represent landmarks in English-language lexicography: (a) *A Dictionary of the English Language* (1755), (b) *An American Dictionary of the English Language* (1828), (c) *Roget's Thesaurus of English Words and Phrases* (1852), (d) *The Oxford English Dictionary* (1928), (e) *The Oxford Advanced Learner's Dictionary of Current English* (1948), (f) *Webster's Third New International Dictionary, Unabridged* (1961), (g) *Collins COBUILD English Language Dictionary* (1987), (h) *The New Oxford Dictionary of English* (1998), (i) *Longman Language Activator* (1998), (j) *WordNet*. Find out the significance of each of these dictionaries.
4. Read carefully Samuel Johnson's "Preface to the Dictionary" (Resource 1) and discuss his approach to (a) spelling, (b) etymology, (c) explanation, and (d) examples. What was his motive for writing this dictionary?
5. Analyse Johnson's definitions of the words *citizen*, *civilize*, *colony*, *commerce*, *curiosity*, *national*, *native_{adj}*, *pirate*, *trade*, *translation* (Resource 2) and compare them with Webster's definitions (Resource 3). Now establish whether, and to what extent, the various senses and their definitions differ from today's senses and their definitions, as recorded in the online editions of *The American Heritage Dictionary of the English Language* and of *The Merriam-Webster's Collegiate Dictionary*.
6. Read carefully the two extracts from Noah Webster's "Dissertations on the English Language" (Resource 4) and discuss the two ideas he proposes and advocates there. What are the spelling reforms that Webster implemented in his dictionaries (Resource 5)? How do Johnson's and Webster's views on spelling differ?
7. Explore the website of *The Oxford English Dictionary* (Resource 6), focusing on the following: (a) Dictionary milestones, (b) OED editions and facts, (c) How to use the OED, (d) Rewriting the OED, (e) Updates to the OED. What makes this dictionary, and especially its latest (third), online, edition, a unique piece of lexicography, not only in the English-speaking world?
8. Compile a list with exact titles of major contemporary British, American, Canadian and Australian dictionaries of English. What was the title of the first American dictionary?

SPANNING UNIT 3 OF THE SYLLABUS:

9. Assess the various ranges of criteria put forward for classifying dictionaries (Hartmann 2001: Ch 5.3-5.4).
10. Discuss the differences between the following types of dictionaries: (a) general-purpose, encyclopedic, specialized and terminological, (b) print and electronic, (c) mono-

lingual, bilingual, bilingualized and multilingual, (d) native-speaker and learner's (pedagogical), (e) passive/receptive and active/productive, (f) semasiological and onomasiological.

11. Comment on the advantages and disadvantages of electronic dictionaries (available over the internet, in mobile phones and tablets, on CD-ROMs and DVD-ROMs or as self-contained hardware) over print ones.
12. Define the concept of hybrid dictionary. What are its currently available types (Hartmann 2005)?

SPANNING UNIT 4 OF THE SYLLABUS:

13. Discuss the five stages of the lexicographic process: (a) preparation, (b) data gathering, (c) data processing, (d) editing, (e) publishing. What does each of these stages involve (Hartmann 2001: Ch 2.2-2.3)? At which stage must decisions be made about the target audience and their assumed reference needs, the type of the dictionary, the scope of the dictionary and the information categories included? What relationship is there between the five stages and the planning and implementation phases?

SPANNING UNIT 5 OF THE SYLLABUS:

14. Explain the difference between the following components of dictionary structure: (a) macrostructure, (b) megastructure, (c) microstructure, (d) mediostructure. What is meant by access structure (Hartmann 2001: Ch 5.1-5.2)?
15. Discuss the factors which influence qualitative and quantitative aspects of a dictionary's word list, i.e. the type/nature and the number of the words included in it.
16. Comment on the process of deciding whether or not a neologism finds its way into *The Oxford English Dictionary*. What important factors must be taken into account in making the decision (Resource 7)?
17. Identify and analyse front, middle and back matter of your learner's dictionary. Which of the three components of outside matter must never be neglected both by the dictionary compiler and by the dictionary user — and why?
18. Define the three main components of microstructure: (a) headword, (b) form-related comment, which comprises spelling, pronunciation and grammar, and (c) content-related comment, which comprises definition, usage and examples, all of which appear within an entry, the basic reference unit of a dictionary (Hartmann 2001: Ch 5.1-5.2).
19. Study carefully the pages from the *Longman Dictionary of Contemporary English* (2004), the "How to Use this Dictionary" pages from the *Oxford Dictionary of English* (2003) and the "Vodič kroz Rečnik" pages from *Du yu speak anglosrpski? Rečnik novijih anglicizama* (2001), and discuss the microstructural features of the three dictionaries. Find out how these structures differ from one another — and why.
20. Study carefully the mediostructure, i.e. cross-referencing, of your learner's dictionary and discover the system underlying it. What is the main purpose of cross-references in dictionaries? What information is typically cross-referred?
21. In many learner's and native-speaker dictionaries there is an increasing variety of additional information provided between, or even within, individual entries. Identify such information in your learner's dictionary and the way(s) it is presented

there. Could you say to which part of dictionary structure this information would belong?

22. Some types of dictionaries have reduced structure, which departs from the (full) structure outlined above. Name at least five types of such dictionaries, establish in what respect(s) their structure is reduced and say why this is the case.
23. Most British and most American dictionaries use different notational systems to indicate the pronunciation of headwords. What does the difference between them consist in?
24. For some types of headwords dictionaries often fail to provide the pronunciation, which is left to be inferred by the user. Find out what these types of words are. How user-friendly is this method of selective non-provision of pronunciation?
25. Study carefully the grammatical information in your learner's dictionary and discover the system underlying it. There are basically two ways of presenting grammatical information in dictionaries. Can you identify them? Which method is more user-friendly?
26. Lexicographic definitions can be of three main types: (a) synonym definitions, (b) referential definitions, and (c) formulaic definitions, which are suited for particular defining tasks. Explain and exemplify each type of definition and specify the defining tasks for which they are most typically used. What is meant by circularity within a definition (Benson et al. 1986: Ch 4)?
27. The wording of (referential) definitions differs according to whether the headword belongs to the class of lexical words, like *flower*, *enter*, *warm*, or function words, like *and*, *above*, *wow*. Explain and exemplify this difference. Which type of lexical meaning do definitions of lexical words (try to) capture? And what meaning do definitions of function words (try to) capture?
28. Comment on the interrelation between the definition of lexical words, on the one hand, and the prototype of a class of extralinguistic entities, on the other. Which is more basic?
29. Study carefully the wording of the six definitions of the headword *horse*, taken from the current online editions of three learner's and three native-speaker dictionaries:
 - A large animal with four legs, a mane (= long thick hair on its neck) and a tail. Horses are used for riding on, pulling carriages, etc. (*Oxford Advanced Learner's Dictionary*).
 - A large animal with four legs that people ride on or use for carrying things or pulling vehicles. (*Cambridge Advanced Learner's Dictionary*).
 - A large strong animal that people ride and use for pulling heavy things. (*Longman Dictionary of Contemporary English*).
 - A large plant-eating domesticated mammal with solid hoofs and a flowing mane and tail, used for riding, racing, and to carry and pull loads. *Equus caballus*, family Equidae (the horse family), descended from the wild Przewalski's horse. The horse family also includes the asses and zebras. (*Oxford Dictionary of English*).
 - A large hoofed mammal (*Equus caballus*) having a short coat, a long mane, and a long tail, domesticated since ancient times and used for riding and for

drawing or carrying loads. (*American Heritage Dictionary of the English Language*).

- A large, solid-hoofed, herbivorous quadruped, *Equus caballus*, domesticated since prehistoric times, bred in a number of varieties, and used for carrying or pulling loads, for riding, and for racing. (*WordReference Random House Unabridged Dictionary of American English*).

Now comment on the following: (a) the comprehensibility, (b) the usefulness, and (c) the user-friendliness of individual definitions. Which of the above definitions, if at all, satisfies all of these criteria? What information should a good lexicographic definition offer?

30. Explain the concept of controlled defining vocabulary, which is used for writing definitions in British learner's dictionaries. What does the defining vocabulary of your learner's dictionary consist of? What are the advantages of this method? In what dictionary was it first used?
31. Study carefully the presentation of individual senses of polysemous headwords in your learner's dictionary and discover the system underlying it. There are currently three ways of visually setting apart different senses in dictionaries. Can you identify them? Which method do you find the most user-friendly?
32. Study carefully the usage labels in your learner's dictionary and discover the system underlying them. Which type of lexical meaning do they (try to) capture? Why is the provision of labels necessary, especially in learner's dictionaries?
33. Study carefully the use of illustrative examples in your learner's dictionary and discover the system underlying it. There are two basic structural types of examples in dictionaries. Can you identify them? Why is the provision of examples necessary, especially in learner's dictionaries? In what cases are examples usually not provided?
34. Which English-language dictionary offers the most detailed and reliable etymological information? As a rule, etymology is omitted from learner's dictionaries. Why is this so, do you think?

SPANNING UNIT 6 OF THE SYLLABUS:

35. Phraseological units and suffixal derivatives in most dictionaries receive the status of subheadwords. Study carefully the structure of these two kinds of subheadwords in your learner's dictionary and establish how it differs from that of headwords. What is meant by undefined run-on derivatives? Are all suffixations treated as run-ons? What about prefixal derivatives?
36. Bound lexical elements (i.e. affixes and combining forms) are treated slightly differently from free-standing words. Study carefully the entries of the prefixes *dis-* and *un-*, the suffixes *-er* and *-ish*, the ICFs *anthropo-* and *bio-*, and the FCFs *-logy* and *-phobia* in your learner's dictionary. How adequate, both qualitatively and quantitatively, do you find the information provided there? Where, if at all, do you see room for possible improvement?
37. All things considered, what microstructural features and ways of their presentation, in your opinion, greatly contribute to a more user-friendly learner's and native-speaker dictionary — and why?

SPANNING UNIT 7 OF THE SYLLABUS:

38. In addition to many similarities, monolingual and bilingual lexicography, and, consequently, dictionaries, have important differences. Can you name them? Compare two typical entries from a monolingual and a bilingual dictionary, and point out the similarities and differences between the information categories in those entries.
39. Study carefully the microstructures of the entries *enjoy* and *enlarge* from the *Oxford–Hachette French Dictionary* (2001) and of the entries *speech* and *speed* from the *Oxford–Duden German Dictionary* (2005), and identify the microstructural innovations that the new Oxford range of two-way bilingual dictionaries (from *and* into French, German, Spanish, Russian, etc.) has introduced into bilingual lexicography. What is the theoretical background to these innovations? How user-friendly do you find the groundbreaking approach implemented in these dictionaries — and why?
40. Summarize the proposed macrostructure and microstructure of a modern English–Serbian/Serbian–English general-purpose dictionary (Prčić 2019: Ch 19). What features make this dictionary stand apart from the rest of currently available English–Serbian and Serbian–English dictionaries? Which types of dictionary would this project aim to unite?
41. Discuss the semantic and pragmatic approaches to the provision of translation equivalents in bilingual dictionaries. Why is the latter approach linguistically far more justified than the former (Prčić 2019: Ch 19)?
42. Explain the difference between formal correspondents and functional-communicative equivalents, and ways of achieving them in the process of translation of meanings. What is meant by Zgusta's insertable and explanatory equivalents (Prčić 2019: Chs 19, 13)?
43. Discuss the ways proposed for dealing with three categories of headwords which are of great importance for bilingual dictionary making: (a) false friends, (b) culture-specific words, and (c) anglicisms (Prčić 2019: Chs 19, 14). Why are these three categories so important?
44. A typical bilingualized (semi-bilingual, translated) learner's dictionary contains the full English-language text of a dictionary into which translations of all individual senses of headwords and subheadwords, and the accompanying example sentences and phrases are inserted. Using the headword *break_{verb}* from your learner's dictionary, provide the bilingualized version of that entry. As a potential user, how would you rate this type of dictionary — and why?

SPANNING UNIT 8 OF THE SYLLABUS:

45. What is meant by corpus-based analysis of language in general and by corpus-based lexicography in particular? Explain the advantages of the corpus-based approach to language study over other approaches (Biber et al. 1998: Ch 1).
46. How is a corpus usually defined? What conditions must a collection of texts meet in order to be considered a corpus (Biber et al. 1998: Ch 1; Resources 8, 9)?
47. Corpora can be divided into the following types: (a) monolingual vs bilingual vs multilingual corpora, (b) reference vs monitor corpora, (c) comparable vs parallel (translation) corpora, (d) annotated (tagged) vs unannotated (untagged) corpora, (e) tree-

- banks, (f) balanced vs special corpora, (g) speech corpora, (h) native-speaker vs learner corpora, (i) synchronic vs diachronic (historical) corpora. Find out what each of these types involves (Resources 8, 10).
48. Provide a list of major synchronic English-language corpora, together with the type(s) to which they belong and their approximate size (Resource 11).
 49. Explore the website of *iWeb: The Intelligent Web-based Corpus*, the largest corpus in the world, focusing on the following: (a) the size, composition and structure of the corpus, (b) facts about the English language revealed by the corpus, (c) ways of searching the corpus, (d) ways of exploiting the resources of the corpus (Resource 12).
 50. Discuss the uses of corpora in the following areas of language study: (a) speech research, (b) lexical studies, (c) grammar, (d) semantics, (e) pragmatics and discourse analysis, (f) sociolinguistics, (g) stylistics, (h) language and linguistics teaching, (i) historical linguistics, (j) dialectology, (k) psycholinguistics (Resource 13).
 51. There are six major types of research questions to which corpus-based lexicography can provide answers: (a) meanings associated with particular words, (b) frequency of words relative to other related words, (c) dialectal, registral and historical association patterns of words, (d) collocations of words and their distribution across registers, (e) distribution of the senses and uses of words, (f) use and distribution of seemingly synonymous words. Discuss these six questions and illustrate some of the possible answers (Biber et al. 1998: Ch 2).
 52. Querying corpora can be performed in two ways: one is achieved by online access to the resources and the other by offline access. Explain the difference between the two ways of query.
 53. One of the commonest applications of corpus-based language research is concordancing, i.e. producing concordances. What is meant by concordance and by KWIC concordances (Resource 8)? Take a look at the concordances of the works of some famous English romantic poets and discuss the information available there (Resource 14).
 54. Using at least two online corpora of English, make concordance lists of the following words: (a) *lukewarm* vs *tepid* (literal vs metaphorical meanings), (b) *soar* (literal vs metaphorical meanings), (c) *hurl* vs *fling* (meaning distinctions), (d) *actually* (uses), (e) *highly* (right-hand adjective collocates), (f) *ruin* vs *destroy* (right-hand noun collocates), (g) *complaint* (left-hand verb collocates), (h) *heat* (left-hand adjective collocates), (i) *lorry* vs *truck* (frequency), (j) *roadmap* vs *road map* vs *road-map* (frequency) (Resource 15). Now analyse carefully the results and comment on the utility of the insights you have gained in these searches.

SPANNING UNIT 9 OF THE SYLLABUS:

55. Discuss the roles of the four protagonists in dictionary research: (a) the compiler, (b) the user, (c) the teacher, (d) the researcher (Hartmann 2001: Ch 3.1-3.2). Identify the points of similarity and the points of difference between them.
56. Find out the meaning of the prefix *meta-* in the word 'metalexicography', an alternative term for dictionary research.

57. Discuss the six areas of dictionary research: (a) dictionary history, (b) dictionary typology, (c) dictionary structure, (d) dictionary use, (e) dictionary criticism, (f) dictionary IT [= information technology] (Hartmann 2001: Chs 4-6; 2005).

SPANNING UNIT 10 OF THE SYLLABUS:

58. Define the concepts of (a) reference needs and (b) reference skills of the dictionary user (Hartmann 2001: Ch 6). Why is the dictionary user with both his/her reference needs and reference skills so important a factor in dictionary making?
59. What is meant by dictionary awareness (Hartmann 2001: Ch 6)? How can it be achieved and/or heightened both within the educational system and outside it? What should user training, ideally, consist of?
60. Discuss ways of efficient and creative use of dictionaries in (a) language learning, (b) language teaching, (c) translation, and (d) lexical and other research. Each of these four areas focuses on different aspects of dictionaries and their use. Can you identify those specific aspects? Are there any shared points of dictionary use between these four areas? What happens when the user receives no instruction in efficient and creative use of dictionaries?

Appendix 3: Internet resources for research topics

Click on a hyperlink to open a given page in your browser.

1. Johnson: Preface to the Dictionary
<https://johnsonsdictionaryonline.com/preface/>
2. Johnson: A select lexicon of words relating to travel and trade
https://www.norton.com/college/english/nael/18century/topic_4/johnson1.htm
3. Webster's 1828 Dictionary
<http://webstersdictionary1828.com/>
4. Webster: Dissertations on the English language
<http://www.whomes.uni-bielefeld.de/sgramley/Webster.pdf>
5. Spelling reform in English: Webster's attempt at systematization
<http://www.evertype.com/gram/webster-spelling.html>
6. The Oxford English Dictionary (OED)
<https://www.oed.com/>
Dictionary milestones
<https://public.oed.com/history/dictionary-milestones/>
OED editions and facts
<https://public.oed.com/history/oed-editions/>
How to use the OED (video guides)
<https://public.oed.com/how-to-use-the-oed/video-guides/>
Rewriting the OED
<https://public.oed.com/history/rewriting-the-oed/>
Updates to the OED
<https://public.oed.com/updates/>
7. How are words added to the OED?
<https://public.oed.com/how-words-enter-the-oed/>
8. Systematic Dictionary of Corpus Linguistics (Corpora)
<http://donelaitis.vdu.lt/publikacijos/SDoCL1.htm#Corpora>
9. What is a corpus, and what is in it?
<https://www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus2/2FRA1.HTM>
10. Preliminary recommendations on corpus typology
<http://www.ilc.cnr.it/EAGLES96/corpus/corpus.html>
11. Well-known and influential corpora: A survey
<https://www.lancaster.ac.uk/fss/projects/corpus/cbls/corpora.asp>
12. iWeb: The intelligent web-based corpus
<https://www.english-corpora.org/iweb/>
13. The use of corpora in language studies
<https://www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus4/4FRA1.HTM>
14. The web concordances
<http://www.concordancesoftware.co.uk/webconcordances/>
15. British National Corpus
<https://www.english-corpora.org/bnc/>
Corpus of Contemporary American English
<https://www.english-corpora.org/coca/>
WebCorp Live
<http://www.webcorp.org.uk/live/>
Just the Word
<http://www.just-the-word.com/>

Lexicographic Treatment of Negation in Sepedi Paper Dictionaries

D.J. Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (danie.prinsloo@up.ac.za)*

Abstract: Negation in Sepedi is a complex matter — a variety of negative morphemes are used across complicated verbal and copulative structures and these morphemes are not interchangeable. Dictionary users in a text production situation need to be guided to the correct compilation of negative constructions. In most cases negative morphemes are not lemmatised in Sepedi paper dictionaries or if they are, the lexicographic treatment is inadequate. This paper provides an inventory of negation from which an abstraction of the negative morphemes is given. This is followed by a critical evaluation of Sepedi dictionaries and suggested model entries for lexicographic treatment of negative morphemes.

Keywords: NEGATION, SEPEDI, DICTIONARIES, LEXICOGRAPHY, SEPEDI VERBAL MOODS, SEPEDI COPULATIVES

Opsomming: Leksikografiese bewerking van negatiewe vorming in Sepedi papierwoordeboeke. Negatiewe vorming in Sepedi is kompleks — 'n verskeidenheid negatiewe morfeme word in ingewikkelde verbale en kopulatiwe strukture gebruik, en hierdie morfeme is nie uitruilbaar nie. Woordeboekgebruikers in 'n teksproduksiesituasie moet gelei word om negatiewe konstruksies korrek te kan saamstel. In die meeste gevalle word negatiewe morfeme nie in Sepedi-papierwoordeboeke gelemmatiseer nie, of is die leksikografiese behandeling daarvan onvoldoende. Hierdie artikel bevat 'n inventaris van negatiewe vorming waaruit 'n abstraksie van die negatiewe morfeme gegee word. Dit word gevolg deur 'n kritiese evaluering van Sepedi woordeboeke en voorgestelde modelinskrifings vir leksikografiese behandeling van negatiewe morfeme.

Sleutelwoorde: NEGATIEFVORMING, SEPEDI, WOORDEBOEKE, LEKSIKOGRAFIE, SEPEDI WERKWOORD MODI, SEPEDI KOPULATIEWE

1. Introduction

For negation in isiZulu, Bosch and Faaß (2018: 1) state that "as an important instrument of language use, one would therefore expect aspects of negation to be dealt with in dictionaries". They also emphasize that corpus studies reveal an "impressive number" of complex morphological constructions expressing negatives. The situation for Sepedi is no different. Corpus studies reveal that negative morphemes in Sepedi are frequently used. They are, however, often

not lemmatised and treated in their own right in dictionaries. Prinsloo and Gouws (1996) attempted to reduce the negation strategies into a single so-called *ga/sa/se* convention. This convention has since been implemented in several dictionaries, e.g. *Oxford Bilingual School Dictionary* (ONSD) and *Popular Northern Sotho Dictionary* (POP). It is a powerful convention and reasonably user-friendly. The convention given in (1) caters for multiple negative constructions in the indicative, situative, relative, subjunctive, habitual, consecutive, infinitive and imperative moods and for the lemmatisation of verbs ending in *-e*. With a few exceptions verbs ending in *-e* are not lemmatised in Sepedi dictionaries.

(1)
rêke, rekê must buy; **..ga/sa/se..**~ not buy

What is required, however, for negative morphemes is that they should be separately lemmatised with full treatment. Traditional dictionaries reflect a strong word-bias, not focusing on items smaller than words such as morphemes, or multiword items, thus literally taking the alternative term for *lemma*, i.e. *headword* as the norm. This issue, however, falls beyond the scope of this article, cf. Gouws (1989) and Gouws and Prinsloo (2005) for a detailed discussion.

The aims of this article are to (a) give an overview of negation in Sepedi, presented as example driven rules, (b) extract the negation rules, (c) determine the frequency of use of negative morphemes in a Sepedi corpus, (d) critically evaluate the treatment of negation in Sepedi dictionaries and (e) compile model entries for Sepedi paper dictionaries. These model entries will firstly incorporate all the negative morphemes that have to be lemmatised and secondly, give a detailed scope of all the types of information and lexicographic conventions that have to be considered in the treatment of these lemmas.

2. Overview of negation in Sepedi

Negation is discussed in detail in Sepedi grammars such as Lombard, Van Wyk and Mokgokong (1985) and Poulos and Louwrens (1994) but is fragmented over different sections for the different verbal moods and copulatives. Inexperienced learners could find it difficult to get a full perspective on negation from grammar books. This article departs from such typical grammar-book descriptions but presents a summary of the full scope of negative constructions in only three tables. These tables will reflect an example-driven approach in a systematic way across different verbal moods and copulatives. The tables will also be used as a basis for the extraction of the negation rules, compilation of text boxes, running footers, back matter components, etc. in paper dictionaries. For the interested reader the full scope of verbal moods and copulatives is summarised in appendices A and B.

3. Negation of verbal moods

In this article the distinction of eight verbal moods for Sepedi by Lombard et al. (1985) is followed. The negation rules indicated in tables 1 to 3 are intended for a step-by-step build-up of the negative morphemes which will be regarded as lemma candidates. These negative morphemes will also be used in the evaluation of 12 Sepedi dictionaries. Tables 1 to 3, as well as appendices A and B, can also be used in the back matter of dictionaries as a reference address for the articles of negation morphemes in the central text. So, for example, references from the articles for *ga*, *sa* and *se* 'not' in the central text to the back matter contextualise *ga*, *sa* and *se* in the broader perspective of verbs and copulatives and could be valuable for the user's cognitive needs.

Table 1: Negation strategies for verbal moods

Mood	Negation strategy	Example	Neg.
3.1 Indicative			
3.1.1 Pres.	ga + subject concord + verb stem ending -e	<i>Ga ke reke</i> 'I do not buy'	ga
3.1.2 Fut.	subject concord + ka se + verb stem ending -e	<i>Nka se reke</i> 'I shall not buy'	ka se
3.1.3 Past	1: ga se + alternative concord + verb stem 2: ga se + subject concord + verb stem ending -e 3: ga + subject concord + a + verb stem 4: ga + alternative concord + verb stem	<i>Ga se ka reka</i> 'I did not buy' <i>Ga se ke reke</i> 'I did not buy' <i>Ga ke a reka</i> 'I did not buy' <i>Ga ka reka</i> 'I did not buy'	ga se ga se ga ga
3.2 Situative			
3.2.1 Pres.	subject concord + sa + verb stem ending -e	<i>Ge ke sa reke</i> 'If I do not buy'	sa
3.2.2 Fut.	subject concord + ka se + verb stem ending -e	<i>Nka se reke</i> 'I shall not buy'	ka se
3.2.3 Past	subject concord + sa + verb stem	<i>Ge ke sa reka</i> 'If I did not buy'	sa
3.3 Relative			
3.3.1 Pres.	subject concord + sa + verb stem ending -e + -go/-ng	<i>Yo ke sa rekego</i> 'I who do not buy'	sa
3.3.2 Fut.	subject concord + ka se + verb stem ending -e + -go/-ng	<i>Yo nka se rekego</i> 'I who shall not buy'	ka se
3.3.3 Past	subject concord + sa + verb stem + -go/-ng	<i>Yo ke sa rekago</i> 'I who did not buy'	sa
3.4 Subjunctive	subject concord + se + verb stem ending -e	<i>Gore ke se reke</i> 'So that I do not buy'	se
3.5 Habitual	subject concord + se + verb stem ending -e	<i>Ke se reke</i> 'I (usually) do not buy'	se
3.6 Consecutive	alternative concord + se + verb stem ending -e	<i>Ka se reke</i> 'I (then) did not buy'	se
3.7 Infinitive	go + se/sa + verb stem ending -e	<i>Go se reke</i> 'not to buy'	se/sa
3.8 Imperative	se + verb stem ending -e se + ke + alternative concord + verb stem	<i>Se reke!</i> 'Do not buy!' <i>O se ke wa reka!</i> 'Do not buy!'	se

4. Negation of copulatives

Louwrens (1994) gives a brief description of copulatives as follows:

A term used to refer to structures which incorporate a variety of prefixes (called particles in some grammatical descriptions of Northern Sotho) of which some resemble the subject concord (e.g. o bohlae, le bohlae, e bohlae, etc. 'he/it is clever'), whereas others have a unique form like ke (e.g. ke monna 'it is a man') and ga se (e.g. ga se monna 'it is not a man'), as well as a series of verb forms with the copulative verb stems -bê (e.g. ... gore a bê bohlae 'so that he may become clever'), -ba (e.g. ... a ba bohlae 'and then he became clever'), -le (e.g. ... gê a le bohlae 'if he is clever'), -se (e.g. ... gê a se bohlae 'if he is not clever') and -na (e.g. o na le lehufa 'she is jealous'). The copulative expresses the English verb to be, and is therefore translated with 'is', 'was', 'will be', 'has been', etc. (Louwrens 1994: 40)

The copulative in Sepedi is a very complex system — see appendix B for a detailed overview of the structure of static and dynamic copulatives. As for the verbal moods in table 1 the aim of the following section is to list the different negative copulative constructions and to extract the negative morphemes which will also be regarded as lemma candidates for Sepedi dictionaries.

4.1 Static copulatives

Louwrens (1991: 71) defines static as "referring to a state of rest which is characterised by the absence of motion or change". Three semantic relations are distinguished between a subject and a complement within the copulative, i.e. *identification/equality*, *descriptive* or *associative*.

Table 2: Negation strategies for static copulatives

4.1.1 Identifying			
Mood	Negation strategy	Example	Neg.
4.1.1.1 Indicative	<i>ga</i> + subject concord + complement <i>ga se</i> + complement	<i>Ga ke morutiši</i> 'I am not a teacher' <i>Ga se morutiši</i> 'It (he/she) is not a teacher'	ga ga se
4.1.1.2 Situative	subject concord + <i>se</i> + complement	<i>Ge ke se morutiši</i> 'If I am not a teacher' <i>Ge e se morutiši</i> 'If it (he/she) is not a teacher'	se
4.1.1.3 Relative	subject concord + <i>sego/seng</i> + complement	<i>Yo ke sego morutiši</i> 'I who am not a teacher' <i>Yo e sego morutiši</i> 'He/she who is not a teacher'	se(go/ng)
4.1.2 Descriptive			
4.1.2.1 Indicative	<i>ga</i> + subject concord + complement	<i>Ga ke bohlae</i> 'I am not clever' <i>Ga a bohlae</i> 'He/she/they is not / are not clever'	ga

4.1.2.2 Situative	subject concord + <i>se</i> + complement	<i>Ge ke se bohlale</i> 'If I am not clever' <i>Ge a se bohlale</i> 'If he/she/they is / are not clever'	se
4.1.2.3 Relative	subject concord + <i>sego/seng</i> + complement	<i>Yo ke sego bohlale</i> 'I who am not clever' <i>Yo a sego bohlale</i> 'He/she who is not clever'	se
4.1.3 Associative			
4.1.3.1 Indicative	<i>ga</i> + subject concord + <i>na</i> (le) + complement	<i>Monna ga a na (le) mpša</i> 'The man does not have (is not with) a dog'	ga
4.1.3.2 Situative	subject concord + <i>se</i> + <i>na</i> (le) + complement	<i>Ge monna a se na (le) mpša</i> 'If the man does not have (is not with) a dog'	se
4.1.3.3 Relative	subject concord + <i>se</i> + <i>nago/nang</i> (le) + complement	<i>Monna yo a se nago (le) mpša</i> 'The man who does not have (is not with) a dog'	se

4.2 Dynamic copulatives

Louwrens (1991: 71) defines dynamic as "referring to a state in which things are in motion and therefore changing". Dynamic copulative constructions are therefore often translated as '(to) become'. Similar to static copulatives, dynamic copulatives distinguish identifying, descriptive and associative relations between the subject and complement.

4.2.1 Identifying dynamic copulatives

As in the case of identifying static copulatives in table 2, first and second persons use their own subject concords, i.e. *ke*, *re*, *o* and *le* but third persons use a neutral subject concord *e*. In table 3 only examples with third person class 1 *monna* 'man' as subject will be given.

Table 3: Negation strategies for dynamic copulatives

4.2.1 Identifying			
Mood	Negation strategy	Example	Neg.
4.2.1.1 Indicative			
4.2.1.1.1 Pres.	<i>ga</i> + neutral subject concord + be + complement	<i>Monna ga e be morutiši</i> 'The man does not become a teacher'	ga
4.2.1.1.2 Fut.	Neutral subject concord + <i>ka se</i> + be + complement	<i>Monna e ka se be morutiši</i> 'The man will not become a teacher'	ka se
4.2.1.1.3 Past	1: <i>ga se</i> + neutral subject concord + ba + complement 2: <i>ga se</i> + neutral subject concord + be + complement 3: <i>ga</i> + subject concord + a + verb stem 4: <i>ga</i> + alternative concord + ba + complement	<i>Monna ga se ya ba morutiši</i> <i>Monna ga se e be morutiši</i> <i>Monna ga e a ba morutiši</i> <i>Monna ga ya ba morutiši</i> 'The man did not become a teacher'	ga se ga se ga ga

4.2.1.2 Situative			
4.2.1.2.1 Pres.	Neutral subject concord + sa + be + complement	<i>Ge monna e sa be morutiši</i> 'If the man does not become a teacher'	sa
4.2.1.2.2 Fut.	Neutral subject concord + ka se + be + complement	<i>Ge monna e ka se be morutiši</i> 'If the man will not become a teacher'	ka se
4.2.1.2.3 Past	Neutral subject concord + sa + ba + complement	<i>Ge monna e sa ba morutiši</i> 'If the man did not become a teacher'	sa
4.2.1.3 Relative			
4.2.1.3.1 Pres.	Neutral subject concord + sa + be-go/ng + complement	<i>Monna yo e sa bego morutiši</i> 'The man who does not become a teacher'	sa
4.2.1.3.2 Fut.	Neutral subject concord + ka se + be-go/ng + complement	<i>Monna yo e ka se bego morutiši</i> 'The man who does not become a teacher'	ka se
4.2.1.3.3 Past	Neutral subject concord + sa + ba-go/ng + complement	<i>Monna yo e sa bago morutiši</i> 'The man who did not become a teacher'	sa
4.2.1.4 Subjunctive	Neutral subject concord + se + be + complement	<i>Gore monna e se be morutiši</i> 'So that the man does not become a teacher'	se
4.2.1.5 Habitual	Neutral subject concord + se + be + complement	<i>Monna e se be morutiši</i> 'Usually the man does not become a teacher'	se
4.2.1.6 Consecutive	Neutral alternative concord + se + be + complement	<i>Monna ya se be morutiši</i> 'The man (then) did not become a teacher'	se
4.2.1.7 Infinitive	go + se/sa + be + complement	<i>Go se be morutiši</i> 'Not to become a teacher'	se sa
4.2.1.8 Imperative	se + be + complement se + ke + alternative concord + ba + complement	<i>Se be morutiši!</i> 'Do not become a teacher!' <i>O se ke wa ba morutiši!</i> 'Do not become a teacher!'	se

4.2.2 Descriptive dynamic copulatives

Descriptive dynamic copulatives use the same negation strategies as the descriptive identifying copulatives, cf. table 3 but all persons and classes use their own subject concords e.g. *Monna ga e be morutiši* 'the man does not become a teacher' (4.2.1.1, table 3) becomes *Monna ga a be bohlale* 'the man does not become clever' and will not be presented here as a separate full table for the different moods.

4.2.3 Associative dynamic copulatives

Associative dynamic copulatives use the copulative verb stem *ba* instead of *na* which is used in the associative static copulatives (cf. 4.1.3, table 2). Associative dynamic copulatives occur in all of the moods and follow the same negation

rules as ordinary verbs as in table 1. Associative dynamic copulatives also use the same negation strategies as associative static copulatives except for the situative and relative where the negation morpheme *-sa* is used in associative dynamic copulatives, e.g. *a sa be le* 'not be with' instead of *-se* (4.1.3.2, table 2) for static associative copulatives.

Due to space limitations the associative dynamic copulatives will also not be presented here as a separate full table for the different moods.

The reference numbers given in the first columns of tables 1 to 3 for the suggested lemma candidates *ga*, *sa*, *se*(*go/ng*), *ga se* and *ka se* in the final columns of these tables, are intended to guide the lexicographer in the microstructural treatment of these lemmas. The lexicographer should, namely, try to include examples representing the moods, morphological changes in subject concords, verbal endings, grammatical structures, etc. in following the reference numbers for each negation morpheme. See discussion below on the suggested scope for treatment and utilization of the reference numbers.

5. Frequency of use of negative morphemes in a Sepedi corpus

Negative morphemes are frequently used in Sepedi. Consider the following approximate counts for *ga*, *sa*, *se*, *ga se* and *ka se* in a corpus of 1.2 million tokens in table 4.¹

Table 4: Frequency counts for negative morphemes in Sepedi

Morpheme	Occurrences
<i>ga</i>	5,000
<i>sa</i>	3,000
<i>se</i>	5,000
<i>ga se</i>	300
<i>ka se</i>	1,800

6. A critical evaluation of the treatment of negation in Sepedi dictionaries

Twelve Sepedi dictionaries were studied for their lemmatisation and treatment of the negative morphemes *ga*, *sa*, *se*, *ga se* and *ka se* and will be briefly evaluated in this section. The dictionaries consulted were *Comprehensive Northern Sotho Dictionary* (henceforth referred to as GNSW), *Kort Drietalige Sakwoordeboek* (KDS), *Noord-Sotho-Afrikaans woordelys* (NAAN), *New English-Northern Sotho Dictionary* (NEN), *Oxford Bilingual School Dictionary* (ONSD), *Popular Northern Sotho Dictionary* (POP), *Pukuntšutlhaloši ya Sesotho sa Leboa* (PTLH), *Pukuntšu Dictionary* (PUKU), *Pukuntšu woordeboek* (PUKU'89), *Sediba* (SEDIBA), *Sesotho sa Leboa-*

English Pukuntšu Dictionary (SEPD) and Woordeboek Afrikaans–Noord-Sotho (WANS). Full references are given in the bibliography.

First, consider table 5 for ten dictionaries where Sepedi is the target language in English–Sepedi or, Afrikaans–Sepedi dictionaries.

Table 5: Dictionary articles for **not / nie**

PUKU'89	No entry
NEN	not , adv. se, ga, ga se
POP	not se , ga, ga se
ONSD	<p>not *** <i>adverb</i> ■ ga; sa; se; ga se ♦ Please do not walk on the wet floor. <i>Hle se sepele godimo ga lebato la go thapa.</i> ♦ You have not done your work. <i>Ga se wa dira mošomo wa gago.</i> ♦ I'm not very hungry. <i>Ga se ka swarwa ke tlala.</i></p> <p>◊ it is not so ■ ga go bjalo ♦ Your wife is pregnant; is it not so? <i>Mosadi wa gago o mmeleng, ga go bjalo?</i></p> <p>◊ not be/become ■ ga/sa/se (...) be; ga se (...) ba</p> <p>◊ not do ■ ga/sa/se (...) dire ♦ They will not do: anything. <i>Ba ka se dire selo.</i></p> <p>In English we often shorten not to n't. For example, You didn't close the door. In Northern Sotho, the notion of not (doing something) is expressed by a variety of negative morphemes, often combined with a change in the verbal ending. Examples of negative morphemes are shown as translation equivalents at the start of this dictionary entry.</p> <p>not be verb (isn't, not being, wasn't/weren't not been) See ISN'T, AREN'T</p> <p>The verb not be is very irregular. In the : Dynamic copulative constructions are therefore often translated as '(to) become)'present tense, for example, we have: I am not, you aren't/are not, she/he/it isn't/is not, we aren't/are not, both/all of you aren't/are not, they aren't/are not.</p>
SEPD	not <i>neg</i> ga; e sego
WANS	nie ga-, se-, ga se-
NAAN	nie , ga.
PUKU	nie ga, sa, se, ga se.
KDS	nie e sego not
SEDIBA	<p>not ga, sa, se <i>I do - run</i> ga ke kitime <i>he/she/it/they is/are - ga se I did - eat</i> ga se ka ja <i>he is - clever</i> ga a bohla <i>I do - know</i> ga ke tsebe [ne] (<i>while</i>) <i>he is - falling</i> a sa we <i>do - go home</i> se ye <i>gae do - come</i> (you, s:) <i>ankê o se tlê the water should - boil whatsoever</i> <i>mêitse a se kê a bela no longer</i> se sa not at all <i>ke/kê (a.v.) you should - start fearing</i> o se kê wa thôma <i>go fšêga they are not overcome at all</i> ga ba ke ba šitwa <i>you (pl.) must not damage it at all</i> le se keng la di senya</p> <p>not, b. - <i>se while we are not far away</i> re - <i>kgolê he/she was not</i> e bê e -</p>

PUKU'89 did not lemmatise *not*, NAAN only lists *ga* as a translation equivalent. NEN, SEPD and POP simply give the translation equivalents of *not* as *se(go)*, *ga* and *ga se*, thus missing out on *sa*. PUKU gives the translation equivalents *ga*, *sa*, *se* and *ga se*. None of the dictionaries with the exception of ONSD and SEDIBA gives information on grammar, pronunciation, examples of use, indicate circumflexes or tone. ONSD and SEDIBA do much better by giving a more exhaustive treatment including translated examples. ONSD also provides frequency and part of speech information as well as additional information in text boxes.

In tables 6 to 8, author inserted English translations for Afrikaans are given in double square brackets.

Table 6: Dictionary articles for *ga*

PUKU'89	<i>ga</i> ontk. -morf., ind. L: nie. [[negative morpheme indicative L (tone) not]]
NEN	<i>ga</i> , neg. part., not, not to; - <i>ke rate</i> , I do not want.
POP	<i>ga</i> not: <i>ga ke morutiši</i> I am not a teacher; at: <i>ga Madika</i> at Madika's place; if, when
ONSD	<i>ga</i> ¹ *** <i>negative morpheme</i> ■ (generally marks the negative form of indicative verbs) ♦ <i>ga ke nyake go robala ka pela. I don't want to go to sleep early.</i> ♦ <i>ga</i> (...) [+ VERB ENDING in -e] ■ do/does not ♦ Bašemane <i>ga ba rate go bala. Boys do not like to study.</i>
SEPD	<i>ga</i> ¹ <i>neg. the negative morpheme; not; place of; if; when</i>
NAAN	<i>ga</i> (negatiewe bywoord), nie. [[(negative adverb) not]]
PUKU	<i>ga</i> , van, na, nie, as, by [[of, to, not, if, at]]; - <i>e sa le</i> , dit is nie meer [[it is no more]]; - <i>se ka</i> , (daar is) baie [[there is plenty]].
SEDIBA	<i>ga</i> nie not - <i>ke kitime</i> ek hardloop nie <i>I don't run</i> - <i>se hy/sy/dit/hulle is nie he she/it/they is/are not</i> - <i>se ka ja</i> ek het nie geëet nie <i>I did not eat</i> - <i>a bohale</i> hy is nie slim nie <i>he is not clever</i>
PTLH	No entry
GNSW	GA (ga-) [ontkennende vormelemt] // [negative formative] <i>ga re bone</i> ons sien nie // we don't see

There is no entry for *ga* in PTLH but entries are given in the other nine dictionaries. *SEPD* has an entry for *ga* but does not lemmatise the possessive, *place of* and situative particle *if, when* as separate entries. This runs against the lemmatisation policy followed in this dictionary i.e. to separate such homonyms and to allocate homonym numbers. *POP* also treats the negative morpheme *ga* and the locative *ga* as well as *if, when* in the same entry. This is also the case in *PUKU* where *nie* [not] is given amidst four other unrelated translations. At least one of the examples given illustrates the use of *ga* in an identifying copulative containing the progressive *sa*. The second example *ga se ka*, '(daar is) baie' [there is plenty], however, makes no sense for any of the other related meanings, i.e. *of, to, if, by* given. *NEN* treats *ga* 'not' in a separate entry and gives an example. No examples are given in *SEPD*, *NAAN* and *PUKU*'89 and therefore are not of much help for text production.

The entries for *ga* in *SEDIBA* and especially *ONSD* are much more informative. *SEDIBA* does well in covering four different negative constructions in the example sentences given, i.e. negation in the indicative present (3.1.1, table 1), indicative identifying copulative (4.1.1.1, table 2), indicative past tense (3.1.3(1), table 1) and indicative descriptive copulative respectively. The examples also indicate two other important aspects, i.e. the use of subject concords of the first person as well as class 1 and that the past tense form of the verb changes back to the present tense. They also illustrate that unlike English, for example, Sepedi does not distinguish between *he is, she is, they are*, and *it is* — they all convert to *ke* 'it is' and *ga se* 'it is not' in the negative. *SEDIBA* also gives exam-

ples of *ga se* for both verbs and copulatives. PUKU'89 indicates tone by means of the letters L 'low' and H 'high'. GNSW also indicates tone e.g. high tones *é* and *ó* — low tones are unmarked. ONSD's entry for *ga* is of high lexicographic quality. It gives detailed grammatical guidance stating that *ga* is a morpheme, used in the negative of the indicative and that the verbal ending changes to *-e* illustrated by examples. It also gives frequency information by means of a one, two or three star convention — indicating in this case that *ga* is one of the top most frequently used tokens in a Sepedi corpus, cf. table 4.

Table 7: Dictionary articles for *sa*

PUKU'89	sa ⁴ , neg. morf. sit./rel. L; <i>gê o sa boe</i> , as jy nie terugkeer nie [[if you do not return]]; <i>gê o sa boa</i> , as jy nie teruggekeer het nie. [[if you did not return]]
NEN	sa , adv., still, yet, ke — ja , I am still eating;[:] — with the negative, no longer, not; — ntše , still.
POP	sa of, belonging to: selêpê sa ka my axe; not: gê ba sa boe if they do not return; still, yet: ba sa ithuta they are still studying; he/she/it then: sa fihla it then arrived; clear, dawn; disappear
ONSD	sa ⁴ *** <i>negative morpheme</i> ■ (marks the negative form of situative verbs) ♦ Ge ba sa ntefele nka se ba thuše gape. <i>If they don't pay me, I won't help them again.</i>
SEPD	No entry
NAAN	No entry
PUKU	sa , steeds, nog; - saam met die negatief nie meer nie; <i>ga a - rate</i> , hy wil nie meer nie; <i>ge a - rate</i> , as hy nie wil nie; - <i>le</i> , terwyl.
SEDIBA	sa nie <i>not a - we</i> (terwyl) hy nie val nie (<i>while</i>) he is not falling
PTLH	No entry
GNSW	SA (sa -) [ontkennende vormelement in infinitiewe en partisipiale wyses] // [negative formative in infinitive and participial moods]; <i>go sa bône</i> om nie te sien nie // not to see; <i>re sa bône</i> (terwyl) ons nie sien nie // we not seeing

In table 7 no entries were given for *sa* in SEPD, NAAN and PTLH. NEN and PUKU treat the negative morpheme *sa* within the article of *still*. If the user reads through all four entries given for *sa*, (s)he will eventually find the information given for the use of *sa* in negative constructions but the possibility of missing the information on negation is high. POP does even worse by treating the possessive, negative, subject concord and progressive in a single entry for *sa*. GNSW does well in stating that *sa* is used in negation of the infinitive and situative moods, illustrated by examples. SEDIBA gives the relevant basic information with an example from the situative. ONSD does well in providing a homonym number, frequency indication, part of speech as well as morphological information and a translated example in a relatively short entry. PUKU'89 is the only dictionary indicating negation with *sa* in the relative mood.

Table 8: Dictionary articles for *se*

PUKU'89	<i>se</i> ⁴ , neg. morf. L; [[negative morpheme L (low tone)]] <i>go se bônê</i> , om nie te sien nie. [[not to see]]
NEN	No entry
POP	<i>se</i> it/he/she; not
ONSD	<i>se</i> ⁴ ***copulative verb ■ is/are/am not ♦ O lebile kae ge e <i>se</i> go Tshikidi? <i>Where are you off to, if not to Tshikidi?</i> <i>se</i> ⁵ *** negative morpheme ■ (marks the negative form of subjunctive, imperative and consecutive verbs) ♦ Hlokomela gore o <i>se</i> forwe ke morekiši. <i>Watch out that you are not cheated by the trader.</i> ... <i>sego</i> ¹ *** copulative verb + rel. marker (go) e <i>se</i> ⁴ ■ who/which is/are/am not ♦ Ao ke mantso a gago e <i>sego</i> a ka. <i>Those are your words, which are not mine.</i>
SEPD	No entry
NAAN	No entry
PUKU	<i>se</i> , (neg form.), nie; <i>se</i> , onderwerp -, voorwerpskakeel v/d. <i>se</i> klas; <i>se sa</i> , nie meer nie. [[subject – object concord of the <i>se</i> class, no more]]
SEDIBA	<i>se</i> nie not - <i>ye gae</i> moenie huis toe gaan nie <i>don't go home ka</i> - mag/kan/sal - <i>may/can/shall/will</i> - <i>ankê o - tlê</i> moenie kom nie (jy) <i>don't come (you, s.) ga - ka ja</i> ek het nie geeet nie <i>I did not eat mêtse a - kê a bela</i> die water moet hoegenaamd nie kook nie <i>the water should not boil whatsoever - sa</i> nie meer nie <i>no longer</i> <i>se</i> nie w. (negatiewe kop. ww. b. not (negative cop. v.) <i>re-kgolê</i> terwyl ons nie ver is nie <i>while we are not far away e bê e- hy/sy</i> was nie <i>he/she was not</i>
PTLH	No entry
GNSW	SE (se-) [ontkenningsvormelement] // [neg. formative] nie // not; <i>gore re ~ rutê</i> sodat ons nie onderrig nie // so that we should not teach; <i>re ka ~ rutê</i> ons sal nie/kan nie onderrig nie // we shall not/cannot teach; <i>re ~ sa ruta</i> terwyl ons nie meer onderrig nie // while we no longer teach; <i>ga ~ ra ruta</i> ons het nie onderrig nie // we did not teach; <i>re ~ rute</i> ons onderrig nie gedurig nie // we do not teach habitually. SE (-se) (d.v.) nié // not; <i>re ~ ra ka ra ruta</i> terwyl ons nie onderrig het nie // while we did not [t]each

In table 8 no entries were given for *se* in SEPD, NEN, NAAN and PTLH. POP gives a translation equivalent following translation equivalents of *se* as a concord. PUKU treats the negative together with subject and object concords but the information on negation of the progressive is valuable. The entries given in ONSD, SEDIBA and GNSW are much more informative giving a variety of information types such as frequency, part of speech, multiple examples, negation in a number of different constructions, etc. They contain most of the elements required for a model entry for *se* in paper dictionaries, cf. discussion below. As far as *se ke* (3.8, table 1 and 4.2.1.8, table 3) is concerned, Lombard et al. (1985: 188) regard *ke* as an auxiliary verb stem. From a user perspective *se ke* should be lemmatised and treated or at least be treated in the articles of *se* and *ke*. SEPD has an entry for *se ke* with translation equivalents *do not*, *don't*. GNSW treats *se ke* in the article of *ke* with reference to its seventh lemma for *ka*. POP

has an entry for *se ke* with translation equivalent *must not* but under the lemma *seke*.

Table 9: Dictionary articles for *ga se*

PUKU'89	No entry
NEN	No entry
POP	No entry
ONSD	<i>ga se</i> ¹ *** copulative particle ■ is/are not ♦ Batho bao <i>ga se</i> ba lapa la ka. <i>Those people are not my family.</i> <i>ga se</i> ² *** negative morpheme ■ did not (marks the negative form of past tense indicative verbs) ♦ Ga se ke bone gore ke sefatanaga sa mang se se mo fološitšego mo kgorwaneng. <i>I did not see whose car dropped her off at the gate.</i>
SEPD	No entry
NAAN	No entry
PUKU	No entry
SEDIBA	No entry
PTLH	No entry
GNSW	No entry

SEDIBA gives an example of *ga se* (313(1)(2), 4.1.1, 4.2.1.1.3(1)(2)) in the negative of the past tense in the article of *se* in table 8 but only ONSD presents it as lemmas with full treatment.

Ka se (3.1.2, 3.2.2, 3.3.2, table 1) is not lemmatised in any of the dictionaries but ONSD gives *ka se* as part of an example for *sa* in table 7. A model entry for *ka se* will be suggested in the next paragraph.

7. The compilation of model entries for paper Sepedi dictionaries for the extracted negative morphemes

In the final columns of tables 1 to 3 above the negative morphemes *ga*, *sa*, *se(go/ng)*, *ga se* and *ka se* were extracted from the negation rules for verbal moods and copulatives. These morphemes are proposed as required lemmas in the macrostructure in Sepedi dictionaries. It will also be attempted to formulate the required essential scope of information types and lexicographic conventions to be considered by the lexicographer in their microstructural treatment in terms of grammatical and semantic information. The lexicographer must strive to cover the full scope of use of a specific negation morpheme in the lexicographic treatment. So, for example, should *ga se* not only be lemmatised but the compiler should also cover its use in verbal moods as past tense negative (3.1.3(1)(2), table 1) and as negative copulative (4.1.1.1, table 2). Such guidance is essential for especially text production with copulatives. No formal user studies of negation has as yet been done for Sepedi but a common error observed in tests of undergraduate students is incorrect negation in the indica-

tive of the static, identifying copulative (4.2.1.1.1, table 2). Learners would for example correctly construct the copulative *I am not a teacher* as *ga ke morutiši* but incorrectly apply negation with *ga* to *the man is not a teacher* as **monna ga o morutiši* or even **monna ga ke morutiši* instead of *monna ga se morutiši*. From the 10 dictionaries consulted for their treatment of *ga* in table 6, only ONSD provides the required guidance by lemmatising and treating *ga se*¹.

Lexicographers should firstly decide on the generic scope of information types to be presented, i.e. which data types should be considered in all cases. They should then decide on the conventions, structural markers and layout strategies to be used consistently. Consider the following recommended information types for negative morphemes in Sepedi:

- Homonym differentiation, preferably by homonym numbers presented as superscripts, cf. ONSD, SEPD, PUKU'89 in tables 6 to 9.
- Frequency information by means of a specific convention such as stars or filled/unfilled diamonds, cf. ONSD in tables 6 to 9.
- Part of speech.
- Translated examples.
- Pronunciation guidance.
- Grammatical and text production guidance, cf. ONSD in tables 6 to 9.
- Tonal indication low/high, cf. PUKU'89 in tables 6 to 8.
- Indication of circumflexes, cf. SEDIBA, GNSW and POP.
- Text boxes.
- Front and back matter.
- Cross references to internal and external sources.
- Running footers.

Consider also the following recommendations for dictionary conventions and structural markers:

- Lemma in boldface.
- Translation equivalents in Roman.
- Source language examples in boldface and italics.
- Translations of examples in italics.
- Use of indicators such as blocks and diamonds to separate data types, cf. ONSD.
- Hard carriage returns, i.e. information on a new line, cf. ONSD in table 6.

Note, however, that there are no binding rules for conventions and structural markers but the lexicographer will be well-advised to follow the practices that users became used to in other dictionaries such as presenting the lemma in lowercase and in boldface, translation equivalents in Roman, etc.

The suggested scopes and model entries for each of the negative morphemes *ga*, *sa*, *se*, *ga se*, and *ka se* are briefly presented in the following paragraphs.

For *ga* the scope includes the indicative present (3.1.1) and past (3.1.3(3)(4)) for verbs and all copulatives in the indicative of identifying (4.1.1.1), descriptive (4.1.2.1) and associative (4.1.3.1), static copulatives and the indicative of dynamic copulatives (4.2.1.1.1), (4.2.1.1.3(3)(4)). ONSD's entry in (2) gives good treatment for *ga* in respect of the negative of the indicative present (3.1.1).

(2)

ga¹ *** *negative morpheme* ■ (generally marks the negative form of indicative verbs)

◆ **ga** ke nyake go robala ka pela. *I don't want to go to sleep early.*

◇ **ga** (...) [+ VERB ENDING in -e] ■ do/does not

◆ Bašeman **ga** ba rate go bala. *Boys do not like to study.*

A reference to the lemmas *ga se*¹ and *ga se*² is advisable so that the user will not miss the treatment of the indicative past for verbs (3.1.3(3)(4), table 1) and the copulatives (4.1.1.1, table 2 and 4.2.1.1.3.(1)(2), table 3). This entry could be further enhanced by an example from class 1 to indicate that the subject concord changes from an *o* in the positive to *a* in the negative. Indication of circumflexes as in POP and PUKU is also recommended.

For *sa* the scope includes situative present (3.2.1), past (3.2.3), relative present (3.3.1), and past (3.3.3) in table 1 for verbs and all dynamic copulatives in the situative (4.2.1.2.1), (4.2.1.2.3) and relative (4.2.1.3.1), (4.2.1.3.3) in table 3.

PUKU'89 provides a good entry for *sa* given in (3).

(3)

sa⁴, neg. morf. sit./rel. L; *gê o sa boe*, as jy nie terugkeer nie [[if you do not return]]; *gê o sa boa*, as jy nie terugkeer het nie. [[if you did not return]]

The different *sa*-morphemes are separated and lemmatised with superscript homonym numbers. It was the only entry found indicating that *sa* is used as negation morpheme for both the situative and relative moods, illustrated by translated examples from present and past tense forms of the situative and indicating tone. The entry can be improved by adding frequency information, treatment and examples in respect of the past tense negatives (3.2.3 and 3.3.3, table 1) and dynamic copulatives (4.2.1.2.1, 4.2.1.2.3, 4.2.1.3.1 and 4.2.1.3.3) in table 3. Indication of circumflexes, cf. POP and PUKU, as well as tone on the vowels themselves is also recommended.

For *se* the scope includes consideration of the subjunctive (3.4), habitual (3.5), consecutive (3.6), infinitive (3.7) and imperative (3.8) for verbs (table 1); for static copulatives: identifying, situative (4.1.1.2), relative (4.1.1.3), descriptive, situative (4.1.2.2), relative (4.1.2.3), associative, situative (4.1.3.2), relative (4.1.3.3) (table 2) and for dynamic copulatives situative (4.2.1.2.1), (4.2.1.2.3) and relative (4.2.1.3.1), (4.2.1.3.3) in table 3.

SEDIBA provides good entries for *se* given in (4):

(4)

se nie *not* - **ye gae** moenie huis toe gaan nie *don't go home* **ka** - mag/kan/sal - *may/can/shall/will* - **ankê o - tlê** moenie kom nie (jy) *don't come (you, s.)* **ga - ka ja** ek het nie geet nie *I did not eat* **mêitse a - kê a bela** die water moet hoegenaamd nie kook nie *the water should not boil whatsoever* - **sa** nie meer nie *no longer*
se nie w. (negatiewe kop. ww. *b. not (negative cop. v.)*) **re- kgolê** terwyl ons nie ver is nie *while we are not far away* **e bê e-** hy/sy was nie *he/she was not*

SEDIBA's entries are strong on coverage, they give examples from the imperative, future tense, subjunctive mood, indicative past tense, use with the auxiliary verb stem *-ke*, negation with the progressive, copulative descriptive, and even some guidance on the use of a copula in a past tense construction. The run-on layout, although indented, is less user friendly, e.g. compared to ONSD's entry for *se* which also gives grammatical and frequency information. Homonym indication by means of superscript numbers is recommended.

For *ga se* the scope includes indicative past (3.1.3(1)(2)) for verbs in table 1 and dynamic copulatives (4.2.1.1.3(1)(2)) in table 3.

ONSD is the only dictionary treating *ga se* in an excellent way distinguishing with homonym numbers between the copulative particle and the negative morpheme.

(5)

ga se¹ *** *copulative particle* ■ is/are not

♦ Batho bao **ga se** ba lapa la ka. *Those people are not my family.*

ga se² *** *negative morpheme* ■ did not (*marks the negative form of past tense indicative verbs*) ♦ **Ga se** ke bone gore ke sefatanaga sa mang se se mo fološitšego mo kgorwaneng. *I did not see whose car dropped her off at the gate.*

Consider also the suggestion for model entries for *ga se* by Prinsloo (2002: 36) in example (6):

(6)

ga se [cop. part. neg.] it is not, **ga se phošo ya gago** it is not your fault; he/she/it is not, **Satsope ga se morutiši, ke mongwaledi** Satsope is not a teacher, she is a secretary; they are not, **dingaka ga se mahodu** doctors are not thieves ▶ **ke** ▶ BM1.1; ■■■■■■

...

se, ga ~ ▶ **ga se**

In (6) a cross-reference marker "▶" informs the user that more information is available in the back matter of the dictionary. For the user who looked up *se* but actually need information about *ga se* a reference entry is given at *se* with a cross-reference to *ga se* where a full treatment is given.

Provision of a text box at *ga se¹* in (5) for indicative, identifying, static copulatives (4.1.1.1, table 2) is recommended as learners or inexperienced users are often unfamiliar with the fact that no distinction is made in Sepedi for *he* is

not, *she* is not, *they* are not and *it* is not and often result in incorrect utterances for, e.g. he/she is a teacher or they are teachers such as **ga a morutiši* or **ga ba barutiši* respectively.

Unlike in English there are not different constructions in Sepedi for *he* is not, *she* is not, *they* are not and *it* is not — all convert to the same construction: **ga se 'it is not'**

Treatment of *ga se*² should include examples from the alternative negative strategies (3.1.3(1)(2), table 1) and could also be nicely complemented by a page footer containing the set of alternative concords used or a reference to the front, middle or back matter where all the alternative concords are presented together.

1st Person singular **ka**, 1st Person plural **ra**, 2nd Person singular **wa**, 2nd Person plural **la**, Class: 1 **a**, 2 **ba**, 3 **wa**, 4 **ya**, 5 **la**, 6 **a**, 7, **sa**, 8 **tša**, 9 **ya**, 10 **tša**, 14 **bja**, 15 – 18 **gwa**

For *ka se* the scope includes the future tense of the indicative (3.1.2), situative (3.2.2) and relative (3.3.2) in table 1, and the future tense of all dynamic copulatives in the indicative (4.2.1.1.2), situative (4.2.1.2.2) and relative (4.2.1.3.2) in table 3.

As mentioned above, none of the dictionaries lemmatised *ka se*. Consider an attempt towards a model entry for *ka se* in (7).

(7)

ka se^{1 ***} [kɑ se] (*negative of the potential ka*) ■ can not, may not. *Maripane a ka se kgônê. Maripane will not be able to nka se arabê potšišô yê I can not answer this question*

ka se^{2 ***} [kɑ se] (*negative of the future tense indicator tlo*) ■ shall/will not *re ka se boê we shall not come back gê ba ka se arabê if they will not answer yô a ka se ithutêgo ka mafolofolo he who will not study hard*

Provision of a text box at *ka se* is also recommended to inform the user about an important difference between the negative of the future in English and Sepedi.

In Sepedi there is no distinction between the negative forms for shall/will not (future tense) and can not (potential), both are expressed by **ka se**. You have to determine from the context which meaning applies.

The examples in *ka se*¹ indicate that the subject concord of class 1 is *a* and that *ke + ka* changes to *nka* and in *ka se*² illustrate and confirm the use of *ka se* in the indicative, situative and relative. In this way the use of examples are not merely intended to confirm the existence of a word in the language but adds to the meaning information.

The text boxes at *ga se* and *ka se* appropriately illustrate the nature and purpose of text boxes in terms of Gouws and Prinsloo (2010: 501):

Text boxes are salient dictionary entries and as such they are used to place more than the default focus on a specific data item ... guidance in terms of sense, contrasting related words, restrictions on the range of application, register, pronunciation, et cetera. ... Where dictionaries have a text production function data could be included in a text box to emphasise the use or non-use of certain combinations and collocations as well as proscriptive guidance.

As far as the required information for the treatment of negative morphemes is concerned, one should not attempt to compile a one-size-fits-all set of information types. For paper dictionaries the lexicographer should keep in mind that the size of the dictionary, i.e. available physical space for an article is always a crucial consideration. Therefore, one cannot expect lengthy detailed articles for negative morphemes in small bilingual dictionaries such as POP, NEN, PUKU, PUKU'89, SEDIBA, etc. The challenge is maximum utilization of limited space and from the dictionaries discussed in reference to tables 6 to 9, ONSD comes closest to covering the most important information types in relatively short articles and in a user-friendly layout.

As a final suggestion for model entries consider also the front matter entry in figure 1:

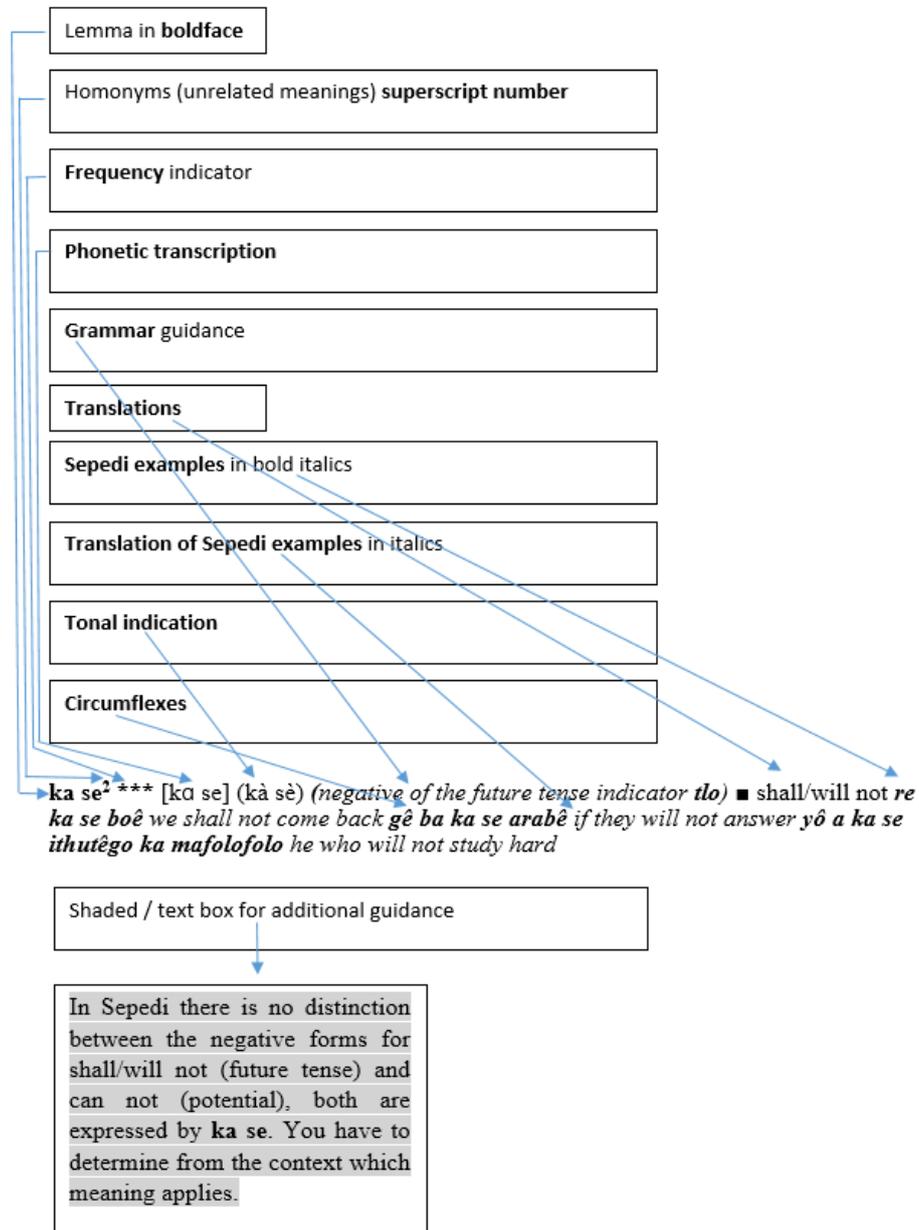


Figure 1: Suggestion for a model entry in the user's guide

Presenting figure 1 in the user's guide to the dictionary will guide the user to appreciation of the information types presented in the articles.

8. Conclusion

Within the space limitations of an article it has been attempted to give an overview of negation in verbs and copulatives and the frequency of occurrence of negative morphemes in Sepedi. The focus was on the different negation strategies and the compilation of a list of lemma candidates to be treated in dictionaries for this language. A critical evaluation of a number of Sepedi dictionaries and suggestions for the scope and conventions for the treatment of negative morphemes were given. Model entries for dictionary articles for negative morphemes and user guidance in the front matter as well as comprehensive back matter entries were proposed.

Acknowledgements

This research is supported by the South African Centre for Digital Language Resources (SADiLaR).

Endnote

1. The Sepedi corpus is marked-up for part of speech which helped to determine frequency counts for the different negative morphemes. However, in order to verify these counts, manual calculations were also done on sections of the corpus to compensate for tagging errors or tags not refined enough to rule out false positives. The counts given in table 4 are therefore only intended to confirm that negative morphemes are frequently used in Sepedi.

Bibliography

Dictionaries

- (GNSW) Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho-woordeboek / Comprehensive Northern Sotho Dictionary / Pukuntšu ya Sesotho sa Leboa*. Pretoria: J.L. van Schaik.
- (KDS) Grobler, G.M.M. 1991. *Die Kort Drietalige Sakwoordeboek*. Pretoria: Ad Donker.
- (NAAN) Kotzé, N.J. 1957. *Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho woordelys*. Johannesburg: Voorwaarts.
- (NEN) Kriel, T.J. 1976. *The New English-Northern Sotho Dictionary, English-Northern Sotho, Northern Sotho-English*. Johannesburg: Educum Publishers.
- (ONSD) De Schryver, G.-M. (Editor-in-chief). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*. Cape Town: OUP Southern Africa.
- (POP) Kriel, T.J., D.J. Prinsloo and B.P. Sathekge. 1997. *Popular Northern Sotho Dictionary, Northern Sotho-English, English-Northern Sotho*. Cape Town: Pharos.
- (PTLH) Mojela, M.V. (Ed.). 2007. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
- (PUKU) Kriel, T.J. 1983. *Pukuntšu Dictionary*. Pretoria: J.L. van Schaik.

- (PUKU'89) Kriel, T.J. and E.B. Van Wyk. 1989. *Pukuntšu woordeboek, Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho*. Pretoria: J.L. van Schaik.
- (SEDIBA) Lombard, D.P., R. Barnard and G.M.M. Grobler. 1992. *Sediba, Practical List of Words and Expressions in Northern Sotho, Northern Sotho-Afrikaans-English, English-Northern Sotho*. Pretoria: Via Afrika.
- (SEPD) Mojela, M.V., M.C. Mphahlele, M.P. Mogodi and M.R. Selokela. 2006. *Sesotho sa Leboa / English Pukuntšu Dictionary*. Cape Town: Phumelela.
- (WANS) Gerber, H.H. 2000. *Woordeboek Afrikaans-Noord-Sotho / Pukuntšu Seburu-Sesotho sa Leboa*. Eldoraingne: Arbeidsprestasie BK.

Other

- Bosch, Sonja and Gertrud Faaß.** 2018. Options for a Lexicographic Treatment of Negation in Zulu. Kernerman, I. and S. Krek (Eds.). 2018. *Proceedings of the LREC 2018 Workshop of GLOBALEX 2018, Lexicography & WordNets, 8 May 2018, Miyazaki, Japan*: 2-10. Miyazaki: Globalex.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. and D.J. Prinsloo.** 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*: 501-511. Leeuwarden: Fryske Akademy.
- Lombard, D.P., E.B. van Wyk and P.C. Mokgokong.** 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L. van Schaik.
- Louwrens, L.J.** 1991. *Aspects of Northern Sotho Grammar*. Pretoria: Via Afrika.
- Louwrens, L.J.** 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.
- Prinsloo, D.J.** 2002. The Lemmatization of Copulatives in Northern Sotho. *Lexikos* 12: 21-43.
- Prinsloo, D.J. and R.H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.

Appendix A: Verbal moods in Sepedi

	Positive	Negative
Indicative		
Pres	monna o bolela Sepedi 'the man speaks Sepedi'	monna ga a bolele Sepedi 'the man does not speak Sepedi'
Fut	monna o tlo bolela Sepedi 'the man will speak Sepedi'	monna a ka se bolele Sepedi 'the man will not speak Sepedi'
Past	monna o boletše Sepedi 'the man spoke Sepedi'	monna ga se a bolela Sepedi 'the man did not speak Sepedi'
Situative		
Pres	(ge) monna a bolela Sepedi '(if) the man speaks Sepedi'	(ge) monna a sa bolele Sepedi '(if) the man does not speak Sepedi'
Fut	(ge) monna a tlo bolela Sepedi '(if) the man will speak Sepedi'	(ge) monna a ka se bolele Sepedi '(if) the man will not speak Sepedi'
Past	(ge) monna a boletše Sepedi '(if) the man spoke Sepedi'	(ge) monna a sa bolela Sepedi '(if) the man did not speak Sepedi'
Relative		
Pres	monna yo a bolelago Sepedi 'the man who speaks Sepedi'	monna yo a sa bolelego Sepedi 'the man who does not speak Sepedi'
Fut	monna yo a tlogo bolela Sepedi 'the man who will speak Sepedi'	monna yo a ka se bolelego Sepedi 'the man who will not speak Sepedi'
Past	monna yo a boletšego Sepedi 'the man who spoke Sepedi'	monna yo a sa bolelago Sepedi 'the man who did not speak Sepedi'
Subjunctive		
	(gore) monna a bolele Sepedi '(so that) the man must speak Sepedi'	(gore) monna a se bolele Sepedi '(so that) the man must not speak Sepedi'
Habitual		
	monna a bolele Sepedi 'the man (usually) speaks Sepedi'	monna a se bolele Sepedi 'the man (usually) does not speak Sepedi'
Consecutive		
	monna a bolela Sepedi '(then) the man spoke Sepedi'	monna a se bolele Sepedi '(then) the man did not speak Sepedi'
Infinitive		
	go bolela Sepedi 'to speak Sepedi'	go se bolele Sepedi 'not to speak Sepedi'
Imperative		
	bolela Sepedi! 'speak Sepedi!'	se bolele Sepedi! 'do not speak Sepedi!'

Appendix B: Static and dynamic copulatives in Sepedi

STATIC COPULATIVE	Identifying			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke moruti	ga ke moruti		1 monna ke moruti	monna ga se moruti
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke le moruti	(ge) ke se moruti		1 (ge) monna e le moruti	(ge) monna e se moruti
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke lego moruti	yo ke sego moruti		1 monna yo e lego moruti	monna yo e sego moruti
STATIC COPULATIVE	Descriptive			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke bohlae	ga ke bohlae		1 monna o bohlae	monna ga a bohlae
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke le bohlae	(ge) ke se bohlae		1 (ge) monna a le bohlae	(ge) monna a sa se bohlae
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke lego bohlae	yo ke sego bohlae		1 monna yo a lego bohlae	monna yo a sa sego bohlae
STATIC COPULATIVE	Associative			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke na le bohlae	ga ke na (le) bohlae		1 monna o na le bohlae	monna ga a na (le) bohlae
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke na le bohlae	(ge) ke se na (le) bohlae		1 (ge) monna a na le bohlae	(ge) monna a se na (le) bohlae
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke nago le bohlae	yo ke se nago (le) bohlae		1 monna yo a nago le bohlae	monna yo a se nago (le) bohlae

DYNAMIC COPULATIVE	Identifying			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke ba moruti	ga ke be moruti		1 monna e ba moruti	monna ga e be moruti
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke eba moruti	(ge) ke sa be moruti		1 (ge) monna e eba moruti	(ge) monna e sa be moruti
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke bago moruti	yo ke sa bego moruti		1 monna yo e bago moruti	monna yo e sa bego moruti
		4	Subjunctive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(gore) ke be moruti	(gore) ke se be moruti		1 (gore) monna e be moruti	(gore) monna e se be moruti
		5	Habitual	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ke be moruti	ke se be moruti		1 monna e be moruti	monna e se be moruti
		6	Consecutive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ka ba moruti	ka se be moruti		1 monna ya ba moruti	monna ya se be moruti
		7	Infinitive	
Pos	Neg			
go ba moruti	go se be moruti			
		8	Imperative	
Pos	Neg			
eba moruti!	se be moruti!			

DYNAMIC COPULATIVE	Descriptive			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke ba bohlale	ga ke be bohlale	1	monna o ba bohlale	monna ga a be bohlale
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke eba bohlale	(ge) ke sa be bohlale	1	(ge) monna a eba bohlale	(ge) monna a sa be bohlale
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke bago bohlale	yo ke sa bego bohlale	1	monna yo a bago bohlale	monna yo a sa bego bohlale
		4	Subjunctive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(gore) ke be bohlale	(gore) ke se be bohlale	1	(gore) monna a be bohlale	(gore) monna a se be bohlale
		5	Habitual	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ke be bohlale	ke se be bohlale	1	monna a be bohlale	monna a se be moruti
		6	Consecutive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ka ba bohlale	ka se be bohlale	1	monna a ba bohlale	monna a se be bohlale
		7	Infinitive	
Pos	Neg			
go ba bohlale	go se be bohlale			
		8	Imperative	
Pos	Neg			
eba bohlale!	se be bohlale!			

DYNAMIC COPULATIVE	Associative			
		1	Indicative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
Ke ba le bohlale	ga ke be le bohlale	1	monna o ba le bohlale	monna ga a be le bohlale
		2	Situative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(ge) ke eba le bohlale	(ge) ke sa be le bohlale	1	(ge) monna a eba le bohlale	(ge) monna a sa be le bohlale
		3	Relative	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
yo ke bago le bohlale	yo ke sa bego le bohlale	1	monna yo a bago le bohlale	monna yo a sa bego le bohlale
		4	Subjunctive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
(gore) ke be le bohlale	(gore) ke se be le bohlale	1	(gore) monna a be le bohlale	(gore) monna a se be le bohlale
		5	Habitual	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ke be le bohlale	ke se be le bohlale	1	monna a be le bohlale	monna a se be le moruti
		6	Consecutive	
Person: 1PS - 2PP				Classes 1-18
Pos	Neg		Pos	Neg
ka ba le bohlale	ka se be le bohlale	1	monna a ba le bohlale	monna a se be le bohlale
		7	Infinitive	
Pos	Neg			
go ba le bohlale	go se be le bohlale			
		8	Imperative	
Pos	Neg			
eba le bohlale!	se be le bohlale!			

Lexicographic Treatment of Zero Equivalence in isiZulu Dictionaries

D.J. Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (danie.prinsloo@up.ac.za)*

and

Nompumelelo Zondi, *Department of African Languages, University of
Pretoria, Pretoria, South Africa (mpume.zondi@up.ac.za)*

Abstract: One of the main tasks of compilers of bilingual dictionaries is to find suitable translation equivalents for source language lemmas in the target language. It could be expected that to a large degree one should find full equivalence, or at least partial equivalents in the target language and that there might not be many instances where such translation equivalents are not available. Typically common words such as *table, chair, man, woman* come to mind and they are most likely to have equivalents in the target language. This article focuses on lexical and referential gaps between English and isiZulu, and their treatment in English and isiZulu paper dictionaries. The aim is to determine to what extent suitable translation equivalents are available for English and isiZulu lemmas and what the nature, extent and treatment strategies are in cases where such equivalents are not available. It will be shown that the extent of zero equivalence is much higher for this language pair than expected in general literature on zero equivalents. In some cases a specific concept is known in the target language but the target language has no word for it but in many instances the concept itself is unknown in the target language which implies that the language will also not have a word for such a concept.

Keywords: LEXICOGRAPHY, BILINGUAL DICTIONARIES, EQUIVALENCE, TRANSLATION EQUIVALENTS, ZERO EQUIVALENCE, LEXICAL GAPS, REFERENTIAL GAPS, ISIZULU, PARAPHRASE OF MEANING, SURROGATE EQUIVALENTS

Opsomming: Leksikografiese bewerking van zero-ekwivalensie in isiZulu woordeboeke. Een van die hoofake van samestellers van tweetalige woordeboeke is om geskikte vertaalekwivalente vir brontaal-lemmas in die doeltaal te vind. Daar kan verwag word dat 'n mens in 'n groot mate volle ekwivalensie of ten minste gedeeltelike ekwivalente in die doeltaal sal vind, en dat daar moontlik nie baie gevalle sal wees waar sulke vertaalekwivalente nie beskikbaar is nie. Gewoonlik word eerste gedink aan gewone woorde soos *tafel, stoel, man, vrou* en is dit waarskynlik dat hulle ekwivalente in die doeltaal sal hê. Hierdie artikel fokus op leksikale en verwysingsgapings tussen Engels en isiZulu en hulle bewerking in papierwoordeboeke vir Engels en isiZulu. Die doel is om vas te stel tot watter mate geskikte vertaalekwivalente beskikbaar is vir Engelse en isiZulu lemmas en wat die aard, omvang en behandelingstrategieë is in gevalle waar sulke ekwivalente nie beskikbaar is nie. Daar sal aangetoon word dat die omvang van zero-ekwiva-

lensie vir hierdie taalpaar veel groter is as wat in die algemene literatuur verwag word oor zero-ekwivalente vir beide leksikale en verwysingsgapings. In sommige gevalle is 'n spesifieke begrip in die doeltaal bekend, maar die doeltaal het geen woord daarvoor nie, maar in baie gevalle is die begrip self onbekend in die doeltaal, wat impliseer dat die taal ook nie 'n woord vir so 'n begrip sal hê nie.

Sleutelwoorde: LEKSIKOGRAFIE, TWEETALIGE WOORDEBOEKE, EKWIVALENSIE, VERTAALEKWIVALENTE, ZERO-EKWIVALENSIE, LEKSIKALE GAPINGS, REFERENSIËLE GAPINGS, ISIZULU, BETEKENISOMSKRYWING, SURROGAATEKWIVALENTE

1. Introduction

In an ideal world suitable translation equivalents for each lemma in a bilingual dictionary should be available in the target language. The prospective compiler of a bilingual dictionary might assume that one can expect to find translation equivalents for all lemmas and that there might not be many instances where such translation equivalents are not available. Taken at face value, considering common objects and concepts, a first impression is that both the source and target languages have suitable translation equivalents. Such an assumption could be strengthened by Adamska-Salaciak's (2006: 117) view that "zero equivalence, while more frequent, is also relatively rare". Gouws and Prinsloo (2008: 869), however, state that there are frequent instances in any given language pair where suitable translation equivalents are not available. It also has to be kept in mind that English is a language with a deeply rooted lexicographic tradition and isiZulu, although having excellent dictionaries, is a language with a strong oral tradition, rich in cultural terminology. Adamska-Salaciak (2006: 24) says that "bilingual dictionaries have to try and do their job despite the fact that the lexicons of natural languages are resistant to pairwise matching" and that finding suitable translation equivalents is problematic because a precise one-to-one correlation between a word and a translation equivalent is rare. Adamska-Salaciak (2006: 99) goes as far as to state that "due to interlingual anisomorphism a bilingual dictionary is, strictly speaking, an impossibility" and that "all we can hope to produce are better or worse approximations".

The aim of this article is to determine for the language pair isiZulu and English to what extent suitable translation equivalents are available for English and isiZulu lemmas and what the nature, extent and treatment strategies are in cases where translation equivalents are not available.

2. Equivalence in bilingual dictionaries

The core of the task of the bilingual lexicographer is to find translation equivalents in the target language for lemmas in the source language. Gouws and Prinsloo (2005: 158) state that the lexicon of a language "does not necessarily

develop parallel to the lexicon of any other language". They emphasize that when a language acquires a word for a given concept it does not imply that other languages will do the same. The views of Gouws and Prinsloo (2005) and Adamska-Sałaciak (2006) are in line with Dagut's (1981) view that equivalence between languages are in principle characterized by a degree of incompatibility and the difficulty, if not impossibility to find translation equivalents in all cases.

... translation is viewed as an attempt to achieve a relationship of "equivalence" between two languages (the SL and the TL), an attempt which, by its very nature, focuses attention on all those incommensurabilities of the two languages concerned that render such "equivalence" difficult (if indeed possible) of attainment. ... The as yet undefined key-term, "equivalence," is to be understood in the Saussurean sense of equal linguistic "value," i.e., as the relationship existing between an item in SL and one in TL, when the TL item performs as nearly as possible the same semantic function in TL as the SL item in SL. (Dagut 1981: 61)

Ideal for the bilingual lexicographer are instances of full equivalence, also known as *congruence* where, in terms of Gouws and Prinsloo (2005: 154) "the source language item and the target language item have exactly the same meaning, function on the same stylistic level and represent the same register". This means that there are no restrictions on the use of the translation equivalent as in example (1).

(1)

dog *noun* inja
malaria *noun* uqhuhqho
talk *verb* khuluma

Gouws and Prinsloo (2008 and 2010) emphasize that in most cases, however, a relation of full equivalence does not exist. Gouws and Prinsloo (2008: 870) state "when comparing any two languages one soon becomes aware of the absence of certain words in a given language". They emphasize that "even highly used common words in a language could lack suitable translation equivalents in the target language". In the absence of full equivalence, the lexicographer tries to find *partial equivalents* as illustrated in example (2).

(2)

incasiso *noun* solution

In (2) the English word *solution* is a partial equivalent of *incasiso* only in the context of an explanation. *Solution* also has other senses e.g. *liquid*.

If it is not possible for the lexicographers to find translation equivalents, they are compelled to revert to *surrogate equivalents* to make up for the lack of full or partial equivalents. These are instances where lexical gaps exist between

source and target language. The absence of suitable translation equivalents is referred to as *zero equivalence*

The *Dictionary of Lexicography (DL)* does not offer treatment of the lemma *surrogate equivalent* but gives a cross-reference to *explanatory equivalent* and defines it as in figure 1.

explanatory equivalent
 In the translation of CULTURE-SPECIFIC VOCABULARY, the explanation of a word or phrase by means of a surrogate PARAPHRASE in the target language rather than a one-to-one EQUIVALENT, e.g. German *Dolchstoßlegende* – English ‘myth of the “stab in the back” (betrayal of Germany after the First World War by its own politicians)’.
 Zgusta 1984, Schnorr 1986.

Figure 1: *Explanatory equivalent* in *Dictionary of Lexicography*

Dagut (1981: 64) distinguishes two types of lexical gaps namely (a) linguistic gaps resulting from linguistic factors and (b) referential gaps caused by linguistic-external factors. A linguistic gap occurs when the concept is known in both the source and target languages but the target language does not have a specific word for it. A referential gap occurs when a source language concept is unknown in the target language.

The strategies employed in the case of zero equivalence include the use of glosses, paraphrase of meaning, pictures and illustrations and information boxes. Consider table 1 and examples (3) and (4) from the *English–isiZulu Dictionary* (henceforth referred to as *EID*), *Scholar’s Zulu Dictionary (SZD)* and *Compact Zulu Dictionary (CZD)*

Table 1: *-ngcuba* in *EID*, *SZD* and *CZD*

<i>EID</i>	<i>SZD</i>	<i>CZD</i>
<p>-ngcuba (ingcuba, 2.9.9, izingcuba) n.</p> <ol style="list-style-type: none"> 1. Meat of an animal that has died, not been slaughtered. 2. Lean meat. 3. Worthless person. 	<p>-ngcuba (i-) (n) meat of an animal which has died from natural causes.</p>	<p>-ngcuba (n) (ingcuba, izingcuba), meat of an animal which has died from natural causes.</p>

(3)

mistral ['mistral] *n.* uhlobo lomoya oyisiphepho obandayo.

In table 1 there is no English translation equivalent for the isiZulu word *ingcuba*. In this case paraphrase of meaning and two short descriptions are given. Table 1 reflects zero equivalence in the form of a lexical gap, i.e. the concepts *meat of an animal which died from natural causes*, *lean meat* and *worthless person* are known to speakers of English but there are no English translation equivalents. Likewise, in (3) there is no isiZulu translation equivalent for the English word *mistral*. In such cases the challenge for the lexicographer is to treat the lemma *mistral* in such a way that users get the required information from the dictionary. The Collins Dictionary gives the meaning of *mistral* as 'a strong, cold north-westerly wind that blows through the Rhône valley and southern France into the Mediterranean, mainly in winter' (<https://www.collinsdictionary.com/dictionary/english/mistral>). The entire concept of a wind blowing through the Rhône valley and southern France into the Mediterranean is unknown in isiZulu, thus representing a referential gap in addition to the language not having a word for it. Thus in terms of Dagut (1981: 63) "there is no single designator ... which provides the required equivalent encapsulation" in isiZulu of "strong, cold north-westerly wind" + "Rhône valley" + "southern France" + "into the Mediterranean". It is not a matter of an inability in the target language to express the meaning of the source language item but whether equivalents are available in the target language.

The question under investigation here is not whether anything said in language A can also be said, in some way, in language B, but rather what happens (in terms of translation equivalence) when A is lexically better equipped than B to say a certain thing. And "better equipped" refers to A's possession of the compressed power of a designator which B lacks. (Dagut 1981: 64)

In (3) the lexicographer opted to give a paraphrase of meaning using a possessive construction *uhlobo lomoya*. 'type of wind', and verbal relative *oyisiphepho obandayo* 'a strong cold wind' (a strong wind which is cold). This is still a far cry from a description of all the meaning connotations of *mistral* conveyed in the Collins Dictionary entry but at least captures the core semantic elements.

The concept *hlonipha* is unknown in English. English therefore lacks a translation equivalent. This means that the lexicographer first has to introduce the concept to the users and then find a way to tell them what it means and not surprisingly leads to lengthier articles, cf. the article for *hlonipha* in figure 2.

hlonipha (3.2.9) v. [denom. < amahloni. > perf. -hloniphile; pass. hlonishwa; neut. hlonipheka; ap. hloniphela; rec. hloniphana; caus. hloniphisa; int. hloniphisisa; dim. hlonihlonipha; inhlonipho; isihlonipho; umhlonishwa.]

1. Respect, reverence, regard with awe, honour; fight shy of. [cf. *hlomuza*.] *Kuhlonishwana kabili* (Respect is mutual, i.e. If you want respect, you must pay respect). *Ubohlonipha uyihlo nonyoko* (Honour your father and your mother).
2. Act respectfully, modestly; cover the breasts or head (according to Native rules of modesty for married or engaged women). *Ihlonipha lapho ingayugana khona* (A bride acts with modesty even where she will not go to marry).
3. Avoid in conversation the use of words which contain the radical of the name of certain persons to whom such respect is due; substitute a fresh term for a word avoided due to respect according to Native custom. *Amakhangelo yizwi lokuhlonipha amehlo* ('Lookers' is the hlonipha term for 'eyes').

Figure 2: Article for *hlonipha* in EID

Consider also cultural words such as *isidanga* 'a kind of necklace', *umhlaba* 'red arrow' and *ukhuzo* 'customary ceremony' where the treatment in the central text by *The Greater Dictionary of (isi)Xhosa* volumes 1, 2 and 3 (GDX1, GDX2 and GDX3) is supplemented by lengthy addenda in the back matter of the dictionaries. As far as *necklace* is concerned, dictionaries such as the *English Xhosa Dictionary* (EXD) and *Xhosa Live Dictionary* (<https://www.gononda.com/xhosa/>) give a surrogate equivalent *intsimbi yomqala* 'iron of the neck' or a loan word *ineklesi* 'necklace' respectively.

(4)

a. EXD

neck

...

~ **lace** n. *intsimbi yomqala* (in- iin-)

b. *Xhosa Live Dictionary*

necklace *ineklesi* (ii-)

GXD1, however, gives a detailed treatment for *isidanga* in figure 3 in the central text of the dictionary supplemented by more detailed information given in the form of an addendum (12) in isiXhosa, English and Afrikaans in the back matter as given in figure 4 for isiXhosa and English.

<p>isi-dānga¹ b/n 7/8:</p> <p>1 intambo yobuhlalu obusinxibo senkosi;</p> <p>2 intsimbi enemingqi emininzi ethe yahlolhwa intsimbi yexabiso, kumaXhosa sinxitywa sithiwe gatya emidudweni ngamadoda nasemigidini nakuzo zonke iindawo zemgqobo yabantu abakhulu; kwaisidanga esi yinto ebonisa umfazi odudileyo, oye ngoduli waduda emzini wakhe, kuba unxiba sona mhla waduda enesidabane, aye exhantini ukuya kuhlaba umkhonto enxibe sona; esi sidanga ke uya kuhlala naso ade aluphale emana xa aya emidudweni esithi gatya; kgl ne- <i>Add 12</i>:</p>	<p>1 traditional royal necklet of red beads;</p> <p>2 an ornament consisting of many strings of blue beads worn around the neck and across the upper part of the body; it is a proud possession of an adult Xhosa, worn by men on all ceremonial occasions, eg at weddings and other festivities; a woman wears hers at her wedding (<i>umduko</i>, qv) when she thrusts the assegai into the ground at the kraal gate (see <i>ukuhlaba umkhonto</i>); thereafter she wears it at all festivities as a proud token of her having been married by the full traditional rites: see also <i>Add 12</i>.</p>	<p>1 koninklike halsnoer van rooi krale;</p> <p>2 ornament van blou krale bestaande uit veelvuldige stringe wat om die nek en gekruis oor die bors gedra word; dit is 'n trotse besitting van 'n volwasse Xhosa; mans dra dit by alle seremoniële geleenthede soos huwelike en ander feeste; 'n vrou dra hare op haar huweliksdag (<i>umduko</i> q v) wanneer sy die assegai by die kraalhek in die grond steek (sien <i>ukuhlaba umkhonto</i>); daarna dra sy dit by alle geestelike geleenthede as trots bewys dat sy ten volle volgens tradisionele ritusse gehuud is; sien ook <i>Add 12</i>.</p>
--	--	---

Figure 3: Article for *isidanga* in *GDX1*

<p>IADDENDAM (ISIHLOMELo SE-) YE-12</p> <p>ISIDANGA</p> <p>Yintsimbi enemingqi emininzi ethe yahlolhwa kwaye inombala ozuba oki oblofu. Yintsimbi yexabiso kumaXhosa kwaye sinxitywa sithiwe gatya emidudweni ngamadoda nasemigidini nakuzo zonke iindawo zemgqobo yabantu abakhulu. Kwaisidanga esi yinto ebonisa umfazi odudileyo, oye ngoduli waduda emzini wakhe, kuba unxiba sona mhla waduda enesidabane, aye exhantini ukuya kuhlaba umkhonto enxibe sona, esi sidanga ke uya kuhlala naso ade aluphale emana xa aya emidudweni esithi gatya. Uyata wekhaya okanye intlabi usinxiba nentsimbi yentloko apha the umkhonto abhinqe ingubo yakwaXhosa iblankethe kuba izixhobo zomsebenzi wekhaya, esalatha ukumela amanyange.</p>	<p>ADDENDUM 12</p> <p>It is a necklace consisting of many strings of blue beads worn around the neck and across the upper part of the body. It is a proud possession of an adult Xhosa, worn by men on all ceremonial occasions, eg at weddings and other festivities. A woman wears hers at her wedding (<i>umduko</i>) when she thrusts the assegai into the ground at the kraal gate (see <i>ukuhlaba umkhonto</i>). Thereafter she wears it all festivities as a proud token of her having been married by the full traditional rites. The head of the house or the official butcher of the sacrificial animal wears this necklace with a head band made of beads together with a blanket. This is the required regalia during a ritual ceremony which also facilitates easy communication with the ancestors.</p>
--	---

Figure 4: Addendum for *isidanga* in *GDX1*

Gouws and Prinsloo (2008: 873) propose a hierarchy of surrogate equivalents where taboo, culture specific or sensitive values represent the top tier. They state that in such cases treatment will go further than short paraphrases of meaning — the lexicographer has to revert to more elaborate descriptions, pictorial illustrations, etc.

As a second example consider a taboo word in isiZulu in example (5).

- (5)
- Oxford Zulu Dictionary (OZD)*
- umfana** *** *noun* 1/2 (pl. **abafana**) ■ boy
- ◆ Umfana wakwamakhelwane usikwe ibhodlela.
- The boy from next door was cut by a bottle.*

Umfana is a top frequency word in isiZulu, appropriately marked as such by a three star (***) rating in OZD. It is furthermore unmarked, suggesting that it can be used in all contexts to refer to a boy. There is, however, a taboo word *umkhapheyana* for *umfana* for which English has a referential gap and which will require a much more detailed treatment and a taboo-alert as in (6).

(6)

umkhapheyana *noun 1/2 (pl. abakhapheyana)*

Hlonipha term for **umfana** boy / **abafana** boys.

A woman married into polygamy will avoid mentioning/uttering/saying the words *umfana/abafana* since the husband was also first a boy before being a husband. So it is taboo in this setting to ever utter such a word, be it in the singular or plural

The value of a pictorial illustration as surrogate equivalent is clear in cases where a mere description is insufficient. So, for example, is a Yorkshire terrier described in the *Macmillan Dictionary (MD)* as *a small dog with long shiny pale brown hair*, supplemented by a picture. The description given is fine but without a picture, e.g. as in figure 5, the user will still not know how a Yorkshire terrier really looks like in contrast to other small dogs with long brown hair. The isiZulu lexicographer can even coin the term *iYorkshire terrier* and give it as a translation equivalent for *Yorkshire terrier* but that won't be of much help to the user. A combination of a definition and a pictorial illustration as in example (7) and figure 5 is much more informative.

(7)

iYorkshire terrier

Inja encane, eyazalwa ngekhulu le -19, eYorkshire eNgilandi, ubude bayo buthi abube ngamasentimitha angama-20 futhi ingaphila isikhathi esithi masibe yi-15. Ngokuvamile iba nombala omnyama okuthi ngokukhula kwayo ugquke ube mpunga.

A small dog, bred in the 19th century, Yorkshire England, height approx 20 cm, life expectancy approx 15 years, typical colour black turning to grey colour as it gets older.



Figure 5: Picture of a typical Yorkshire terrier

The question to be addressed in the following section is to what extent does the lexicographer in a bidirectional bilingual English–isiZulu dictionary has to deal with zero equivalence. Will it be a matter of finding surrogate equivalents for a few rare cases according to the view of Adamska-Salaciak or that instances of zero equivalence between English and isiZulu are so frequent that it presents a major challenge to the lexicographer?

3. Zero equivalence in English–isiZulu dictionaries

Consider the following analysis of a single, randomly selected page from the English–isiZulu side of *EID* covering the lemma stretch *misdealings* to *mitigate* and a random page from the isiZulu–English side of the same dictionary covering the lemma stretch *-ngcuba* to *-ngeke* in the following discussion.

<p>misdealings ['mis'di:lɪŋz] <i>n.</i> izeleleso (4). misdeed ['mis'di:d] <i>n.</i> iseleso (4), impambuko, isiphosiso, i(l)cala, isono (4); isenzo esibi. misdeemeanour ['misdi'mi:nə] <i>n.</i> iseleso (4); ukuziphatha kabi. misdirect ['misdai'rekt] <i>v.t.</i> qondisa kabi, dukisa; faka ikheli elingeyilo. misdirection ['misdai'rekʃən] <i>n.</i> ukuqondiswa kabi. misdoubt ['mis'daʊt] <i>v.t.</i> -ngathembi, sola. miser ['maɪzə] <i>n.</i> umuntu ombanyileyo, unqodoyi (1a), i(l)qongqela.</p>	<p><i>catch</i>: ehluleka ukwengaka; <i>m. out</i>: shiya, sulazela. (ii) (<i>feel absence of</i>) swela, ntula; <i>They m. him very much</i>: Bamntula kakhulu. <i>2. n.</i> (i) ukugeja, ukuphutha. (ii) *umisi (1a); inkosazana, intombazana. (iii) <i>Miss</i>: uNkosazana; <i>Miss Jones</i>: uNkosazana Jones. missal ['misal] <i>n.</i> incwadi equkethe inkonzo yemisa. misshapen ['mis'ʃeɪpən] <i>a.</i> -bunjiwe kabi, -magodoloz. missile ['misɪl or 'misail] <i>n.</i> isilabi, umlandelisele (2); into ejukujelwayo.</p>
<p>...</p> <p>...</p>	
<p>misread ['mis'ri:d] <i>v.t.</i> funda kabi, -ngafundisisi. misreport ['misri'pɔ:t] <i>v.t.</i> bika ngokungeyikho. misrepresent ['misre'pri:zənt] <i>v.t.</i> sonta; phendukezela iqiniso lento; hlanekezela. misrepresentation ['misre'prɪzən'teɪʃən] <i>n.</i> inhlanekezela. misrule ['mis'ru:l] <i>1. v.t.</i> phatha kabi, busa kabi. <i>2. n.</i> ukuphatha kabi. miss [mis] <i>1. v.t.</i> (i) (<i>fail</i>) duka, phutha; <i>m. the mark</i>: geja, phaphalaza; <i>m. the train</i>: shiywa yisitimela; <i>m. a</i></p>	<p>misunderstanding ['misʌndə'stændɪŋ] <i>n.</i> (i) ukungaqondi kahle; inzwakabi. (ii) ukungezwani, ukungabonani. misuse ['mis'ju:s] <i>n.</i> ukuphatha kabi. misuse ['mis'ju:z] <i>v.t.</i> sebenzisa ngendlela engalungile. mite [maɪt] <i>n.</i> (i) (<i>insect</i>) ubuvunya, isibungu. (ii) (<i>child</i>) umntwanyana omncane. (iii) (<i>coin</i>) uhlamvanyana lwemali. (iv) (<i>tiny thing</i>) u(lu)chokwana, u(lu)hoyizana. mitigate ['mitigeɪt] <i>v.t.</i> thambisa, nciphisa; <i>m. a sentence</i>: nciphisa isigwebo.</p>

Figure 6: Extracts from the stretch *misdealings* to *mitigate* in *EID*

This page is part of the alphabetical stretch M. Seventy-five lemmas are presented on this page and 74 were treated. For 53 lemmas isiZulu translation equivalents are offered of which 29 lemmas were treated using only translation equivalents. For 45 lemmas surrogate equivalents were given of which 21 lemmas were treated using only surrogate equivalents. Twenty four lemmas were treated using both translation equivalents and surrogate equivalents. What is important to note is the *huge number* of lemmas for which surrogate equivalents were used in the treatment, i.e. 45 from 74 = 60,8%, in contrast to translation equivalents 53 out of 74 = 71.6% — thus, even though more lemmas were

treated with translation equivalents, a substantial number of lemmas were treated by surrogate equivalents. Consider table 2 and figure 7.

Table 2: The use of translation equivalents and surrogate equivalents in the English–isiZulu side of *EID*

English–isiZulu	Counts	%
Lemmas treated with translation equivalents	53	71,6
Lemmas treated with surrogate equivalents	45	60,8
Lemmas treated with translation equivalents only	29	39,2
Lemmas treated with surrogate equivalents only	21	28,4
Lemmas treated with both translation equivalents and surrogate equivalents	24	32,4



Figure 7: Translation equivalents and surrogate equivalents in *-misdealings* to *-mitigate* in *EID*

In total 194 equivalents were offered for the 74 lemmas treated of which 58, i.e. 29,9% were surrogate equivalents. Consider (8)–(10) as examples of treatment.

(8)

Translation equivalents only

misdealings ... *n.* izezeleso ...

misgiving ... *n.* ukusola, u(lu)khonono.

misguided ... *a.* -dukile, -dukisiwe, -khohlisiwe.

(9)

Surrogate equivalents only

misspell ... *v.t.* bhala ngokungalingile, pela kabi.

misspend ... *v.t.* chitha imali.

misprint ... 1. *v.t.* cindezela isiphosiso.

2. *n.* isiphosiso ekucindezelweni.

(10)

Both translation and surrogate equivalents

misdirect ... *v.t.* qondisa kabi, dukisa; faka ikheli elingeyilo.

mistress ... *n.* (i) (*woman in authority*) *umesisi (la), inkosikazi;

umaqumbane (la); *school m.:* *uthishakazi; *singing m.:*

umculisi wesifazane. (ii) (*paramour*) i(li)shende. (iii) (= Mrs.)

uNkosikazi.

mitigate [*mitigeit*] *v.t.* thambisa, nciphisa; *m. a sentence:*

nciphisa isigwebo.

Note that in many instances where a suitable translation equivalent does exist, the compilers nevertheless give a number of surrogate equivalents in addition as in the articles of *misdirect*, *mistress* and *mitigate* in (10).

For the opposite side of *EID*, consider extracts from the stretch *-ngcuba* to *-ngeke* in figure 8.

<p>-ngcuba (<i>ingcuba</i>, 2.9.9, <i>izingcuba</i>) <i>n.</i> 1. Meat of an animal that has died, not been slaughtered. 2. Lean meat. 3. Worthless person.</p> <p>-ngcubangcono <i>rel.</i> [<i><</i> <i>ingcuba</i>+<i>ngcono</i>, <i>lit.</i> improved dead meat.] Slightly improved, on the mend. <i>Kufiké yena umhlangano waba ngcubangcono</i> (When he arrived the tone of the meeting improved).</p> <p>-ngcubula (<i>ingcubula</i>, 6.6.3-8.9, <i>izingcubula</i>) <i>n.</i> Fibre girdle. [<i>cf.</i> <i>umkhwindi</i>.]</p> <p>-ngcugcu (<i>ingcugcu</i>, 2.6.9, <i>izingcugcu</i>) <i>n.</i> Vessel narrowing at its mouth.</p>	<p>*-ngcwele (<i>ingcwele</i>, 2.6-3-8.9, <i>sg. only</i>) <i>n.</i> [<i><</i> Xh. <i>-ngcwele</i>. <i>></i> <i>-ngcwele</i>.] Pure, holy person or thing (rarely used as a <i>n.</i>).</p> <p>-ngcweleha (<i>ingcweleha</i>, 2.6.3.9, <i>izingcweleha</i>) <i>n.</i> Murderer. [<i>cf.</i> <i>umbulali</i>, <i>inxeleha</i>.]</p> <p>ngcwelisa (3.2.9) <i>v.</i> [<i><</i> <i>-ngcwele</i>. <i>></i> <i>perf.</i> <i>ngcwelisile</i>; <i>pass.</i> <i>ngcweliswa</i>; <i>neut.</i> <i>ngcweliseka</i>; <i>ap.</i> <i>ngcwelisela</i>; <i>rec.</i> <i>ngcwelisana</i>; <i>caus.</i> <i>ngcwelisisa</i>.] Purify, make holy, hallow. <i>Malingcwelisive igama lakho</i> (Hallowed be Thy name).</p> <p>-ngcwembesi (<i>ingcwembesi</i>, 2.4.6.3, <i>izingcwembesi</i>) <i>n.</i> Clever person, expert. [<i>cf.</i> <i>ingcweti</i>.]</p> <p>-ngcwenge (<i>ingcwenge</i>, 2.4.9, <i>izingcwenge</i>) <i>n.</i> <i>hlonipha</i> term for <i>!(I)gquma</i>, hill.</p>
<p>.....</p>	
<p>-ngcwece (<i>ingcwece</i>, 2.6.3, <i>ongcwece</i>) <i>n.</i> [<i><</i> <i>cwecwe</i>.] 1. Thin stone or iron sheet; slate. 2. [<i>mod.</i>] Sheet of corrugated iron or roofing material.</p> <p>ngcweka (6.3) <i>v.</i> [<i>></i> <i>perf.</i> <i>ngcwekile</i>; <i>pass.</i> <i>ngcwekwa</i>; <i>ap.</i> <i>ngcwekela</i>; <i>caus.</i> <i>ngcwekisa</i>; <i>umngcweko</i>.] Fence, fight with sticks.</p> <p>-ngcweko (<i>umngcweko</i>, 3.2.9.9, <i>sg. only</i>) <i>n.</i> [<i><</i> <i>ngcweka</i>.] A fencing with sticks, <i>ukudlala umngcweko</i> (to play at fencing).</p> <p>-ngcwele <i>rel.</i>; [<i><</i> <i>ingcwele</i>. <i>></i> <i>ngcwelisa</i>.] Pure, holy. <i>UMoya oNgcwele</i> (Holy Spirit); <i>amanzi angcwele</i> (holy water); <i>abangcwele</i> (holy ones).</p>	<p>-ngebeza (<i>amangebeza</i>, 6.6.6.3-8.9, <i>pl. only</i>) <i>n.</i> 1. Pleasant woodland glade with rocks and trees. 2. Dancing arena. 3. <i>Al-fresco</i> feast. <i>ukudla amangebeza</i> (to feast in the open).</p> <p>-ngebezane (<i>ingebezane</i>, 2.4.6.8-3-8.3, <i>izingebezane</i>) <i>n.</i> 1. Morsel. 2. Sickness which weakens the neck muscles, causing the head to flop about.</p> <p>ngedwa (6-3.9) <i>pron. quant.</i> 1st pers. <i>sg.</i> [<i><</i> <i>-dwa</i>.] I alone, only me. <i>mina ngedwa</i> (I alone). <i>Bengihlezi ngedwa</i> (I was sitting by myself).</p> <p>-ngeke <i>defic. v.</i> (<i>fol.</i> by the subjunct. mood).</p>

Figure 8: Extracts from the stretch *-ngcuba* to *-ngeke* in *EID*

This page is part of the alphabetical stretch N dealing with the alphabetical range *-ngcuba – -ngeke*. Forty-two lemmas are presented on this page. For 24 lemmas isiZulu translation equivalents are offered of which 15 lemmas were treated using only translation equivalents. For 27 lemmas surrogate equivalents were given of which 18 lemmas were treated using only surrogate equivalents. Nine lemmas were treated using both translation equivalents and surrogate equivalents. What is important to note for the isiZulu–English side of the dictionary is the *huge number* of lemmas for which surrogate equivalents were used, i.e. 27 out of 42 = 64,3%, compared to translation equivalents 24 out of 42 = 57.1% — thus more lemmas treated with surrogate equivalents than translation equivalents. Consider table 3 and figure 9.

Table 3: The use of translation equivalents and surrogate equivalents in the isiZulu–English side of *EID*

IsiZulu–English	Counts	%
Lemmas treated with translation equivalents	24	57,1
Lemmas treated with surrogate equivalents	27	64,3
Lemmas treated with translation equivalents only	15	35,7
Lemmas treated with surrogate equivalents only	18	42,9
Lemmas treated with both translation equivalents and surrogate equivalents	9	21,4

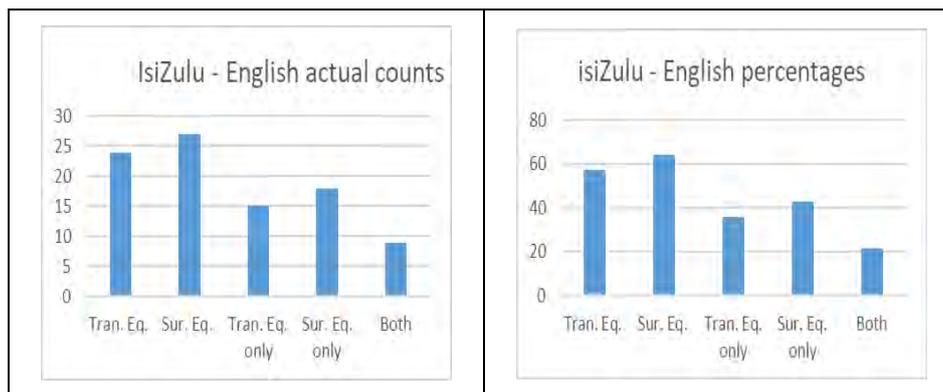


Figure 9: Translation equivalents and surrogate equivalents in *-ngcuba* to *-ngeke* in *EID*

A total number of 78 equivalents were offered for the 42 lemmas treated of which 43, i.e. 55,1% were surrogate equivalents and 35, i.e. 44,9% were translation equivalents. Consider (11) to (13) as examples of treatment.

(11)

Translation equivalents only

-ngcwabi (umngcwabi, 2.6.6-3-8.9, abangcwabi) n.

[< ngcwaba.]

Undertaker.

-ngcumaza (3.2.9) v.

hlonipha term for *dumaza*, disgrace, [cf. ^o*wangaza*.]

-ngcweleha (ingcweleha, 2.6.3.9, izingcweleha) n.

Murderer, [cf. *umbulali, inxeleha*.]

(12)

Surrogate equivalents only

-ngcugcu (ingcugcu, 2.6.9, izingcugcu) n.

Vessel narrowing at its mouth.

-ngcwayo (ingcwayo, 2.9.9, izingcwayo) n.

Garment worn by women to cover the breasts and abdomen when first married, [cf. *ingcayi*.]

-nge (umunge, 2.6.S-3, iminge) n. [< nge.]

1. Hole in the bottom of a sour-milk calabash to let out the whey.
2. (sg. only) Qualms of fear; sudden flash of fear.

(13)

Both translation and surrogate equivalents

ngcwaba (6-3.9) v. [> perf. ngcwabile; pass, ngcwatshwa; neut. ngcwabeka; ap. ngcwabela; rec. ngcwabana; caus. ngcwabisa; int. ngcwabisisa; umngcwabo; i(li)ngcwaba; umngcwabi.]

1. Bury, *ukungcwaba isidumbu* (to bury a corpse).
2. Forget, bury in oblivion.
3. hlonipha term for *gcwala*, be full.

ngcweka (6.3) v. [> perf. ngcwekile; pass, ngcwekwa; ap. ngcwekela; caus. ngcwekisa; umngcweko.]
Fence, fight with sticks.

ngcwelisa (3.2.9) v. [< -ngcwele. > perf. ngcwelisile; pass, ngcweliswa; neut. ngcweliseka; ap. ngcwelisela; rec. ngcwelisana; caus. ngcwelisisa.]

Purify, make holy, hallow. *Malingcweliswe igama lakho* (Hallowed be Thy name).

Treatment strategies for zero equivalents in isiZulu

Typical strategies utilised include adjective and passive constructions, adverbs, relative, possessive and negation strategies, etc. as in examples (14) to (19).

(14)

adjective

isenzo esibi 'a bad deed' (*isenzo* 'deed' + *esibi* 'bad') (noun + adjective)

(15)

passive

shiywa yisitimela 'missed a train' (*shiywa* 'be left behind' + *yisitimela* 'by the train') (verbstem + passive + copulative formative + noun)

(16)

adverbs

ukuziphatha kabi 'misbehave' (*ukuziphatha* 'to handle yourself' + *kabi* 'badly') (verb + adverb)

(17)

relative

faka ikheli elingeyilo 'misdirect' (*faka* 'put in' + *ikheli* 'address, location' *elingeyilo* + 'which is incorrect' (verb + noun + relative)

(18)

possessive

ukunxapha komshini wesibhamu 'misfire' (*ukunxapha* 'vexation' + *komshini* 'of the machine' + *wesibhamu* 'of a rifle')

(19)

negation

humusha ngokungaqondile 'misinterpret' (*humusha* 'mislead' + *ngokungaqondile* 'with not to understand') (verb + negative verb)

Comparison between the isiZulu to English alphabetical stretch *-ngcuba* to *-ngeke* and the English to isiZulu stretch *misdealings* to *mitigate* in EID in table 4 and figure 10 indicate that the use of translation equivalents only is only 3.5% more in the English to isiZulu side than in the isiZulu to English side and 11% more in the English to isiZulu side for cases where both translation equivalents and surrogate equivalents were used. However, instances where the compilers had surrogate equivalents as the only option for treatment were much more frequent, 14.5% more in the isiZulu to English side than in the English to isiZulu side of the dictionary.

Table 4: The use of translation equivalents and surrogate equivalents in the isiZulu–English side compared to the isiZulu–English side

IsiZulu to English	%	English to isiZulu	%
Translation equivalents only	35,7	Translation equivalents only	39,2
Surrogate equivalents only	42,9	Surrogate equivalents only	28,4
Both Translation equivalents and surrogate equivalents	21,4	Both Translation equivalents and surrogate equivalents	32,4

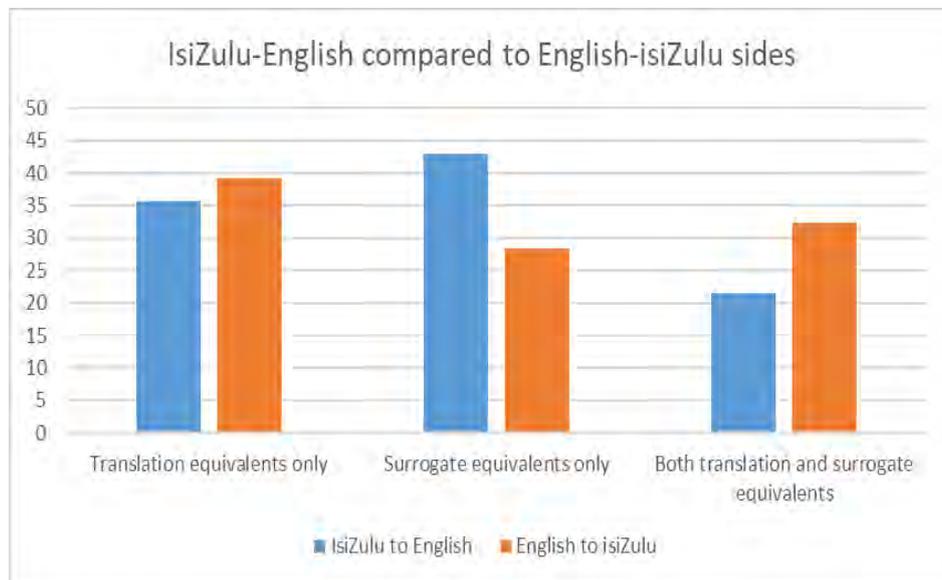


Figure 10: Comparison between isiZulu to English and English to isiZulu

A total number of 101 *surrogate* equivalents were used in the treatment of lemmas in the two randomly selected pages from the isiZulu to English and the English to isiZulu sides of the dictionary. Extrapolation for the entire dictionary gives a total of 63,000 instances of the use of surrogate equivalents in *EID*. A total number of 171 *translation* equivalents were used in these two pages. Extrapolation for the entire dictionary gives a total of 107,000. It is clear that the occurrence of zero equivalence is not rare in English–isiZulu dictionaries.

4. Conclusion

A fundamental task of the lexicographer is to act as the mediator between the

language(s) treated in the dictionary and the target user. In modern lexicography users require a user friendly reference work in which they, in terms of Haas (1962) could find the word preferably in the first place they are looking and in terms of Laufer (1992) enabled by the lookup to "know the word". In respect of lexical gaps the lexicographer should utilise all treatment options available such as loan words, paraphrase of meaning and pictorial illustrations to guide the users to understand the meaning of a word. In this article it has been indicated that lexical gaps are very frequent in the language pair isiZulu/English and subsequently poses a big challenge to the lexicographer to treat them in a satisfactory way in English–isiZulu / isiZulu–English bilingual dictionaries. Lexical gaps in isiZulu (English as source language and isiZulu as target language) are less frequent than lexical gaps in English (isiZulu as source language and English as target language). The research results indicate that *EID* did well in the treatment of the many lexical gaps in both directions of the dictionary. IsiZulu is rich in cultural terms for which there are no English equivalents and the compilers put much effort into finding ways to convey the meaning of such words. Likewise, for the lexical gaps in isiZulu, users are well-provided with surrogate equivalents.

Acknowledgements

This research is supported by the South African Centre for Digital Language Resources (SADiLaR). Findings and conclusions are those of the authors.

Bibliography

Dictionaries

- (CZD) Dent, G.R. (Compiler) and C.L.S. Nyembezi (Ed.). 1961. *Compact Zulu Dictionary English–Zulu, Zulu–English*. Pietermaritzburg: Shuter & Shooter.
- (DL) Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- (EID) Doke, C.M., D.M. Malcolm, J.M.A. Sikakana and B.W. Vilakazi. 2016. *English–IsiZulu, IsiZulu–English Dictionary*. Johannesburg: Wits University Press and South African Heritage Publishers.
- (EXD) Fischer, A., E. Weiss, S. Tshabe and E. Mdala. 1993. *English–Xhosa Dictionary*. Cape Town: OUP Southern Africa.
- (GDX1) Tshabe, S.L. (Editor-in-Chief). 2006. *The Greater Dictionary of isiXhosa. Volume 1: A to J*. Alice: IsiXhosa National Lexicography Unit, University of Fort Hare.
- (GDX2) Mini, B.M. (Editor-in-Chief). 2003. *The Greater Dictionary of isiXhosa. Volume 2: K to P*. Alice: IsiXhosa National Lexicography Unit, University of Fort Hare.
- (GDX3) Pahl, H.W. (Editor-in-Chief). 1989. *The Greater Dictionary of Xhosa Volume 3: Q to Z*. Alice: University of Fort Hare.

- (MD) *Macmillan Dictionary*. Available online at <https://www.macmillandictionary.com>. [Accessed on 20 August 2020]
- (OZD) De Schryver, G.-M. (Ed.). 2010. *Oxford Bilingual School Dictionary: Zulu and English*. Cape Town: OUP Southern Africa.
- (SZD) Dent, G.R. and C.L.S. Nyembezi. 1988. *Scholar's Zulu Dictionary*. Second edition. Pietermaritzburg: Shuter & Shooter.

Other

- Adamska-Salaciak, Arleta.** 2006. *Meaning and the Bilingual Dictionary. The Case of English and Polish*. Frankfurt am Main: Peter Lang.
- Dagut, M.** 1981. Semantic Voids as a Problem in the Translation Process. *Poetics Today* 2(4). 61-71.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. and D.J. Prinsloo.** 2008. What to Say about "mañana", "totems" and "dragons" in a Bilingual Dictionary? The Case of Surrogate Equivalence. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 869-877. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Gouws, R.H. and D.J. Prinsloo.** 2010. Surrogaatekwivalensie in tweetalige woordeboeke met spesifieke verwysing na zero-ekwivalensie in Afrikataalwoordeboeke. [Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Zero Equivalence in Dictionaries for African Languages]. *Tydskrif vir Geesteswetenskappe* 50(4): 502-519.
- Haas, M.R.** 1962. What Belongs in the Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Laufer, B.** 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, H. et al (Eds.). 1992. *EURALEX '92 Proceedings I–II: Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*: 71-76. Tampere: University of Tampere.

Which Defining Model Contributes to More Successful Extraction of Syntactic Class Information and Translation Accuracy?

Bartosz Ptasznik, *University of Warmia and Mazury, Olsztyn, Poland*
(bartosz.ptaszniak@uwm.edu.pl)

Abstract: Definitions in English monolingual learners' dictionaries are the central focus of the paper. Metalexigraphers have had a consuming interest in the following three types of definitions: analytical definitions (or classical definitions), full-sentence definitions (also called contextual definitions) and single-clause *when*-definitions. The use of *when*-definitions, the role of which is to define abstract nouns, has raised questions and doubts as to their efficacy on correct part of speech recognition of the *definiendum*, or item being defined, in light of the problems related to the substitutability of headwords and parts of definitions (lack of general category words in this definition format). By and large, existing research has substantiated the superiority of the classical definition-type over single-clause *when*-definitions with respect to the accuracy of word class identification. The current experiment attempts to further delve into the subject of part of speech recognition with regard to the three aforementioned defining formats — in previous studies only data from analytical and single-clause *when*-definitions were collated, since contextual definitions were not included in the study design. The study was conducted on a group of 120 advanced-level Polish university students of English. The subjects were tested on their ability of correct extraction of syntactic class information and translation accuracy of abstract noun headwords as regards the three predominant definition-types in English lexicographic practice.

Keywords: LEARNERS' DICTIONARIES, DEFINITIONS, MONOLINGUAL DICTIONARIES, ANALYTICAL DEFINITIONS, FULL-SENTENCE DEFINITIONS, SINGLE-CLAUSE *WHEN*-DEFINITIONS

Opsomming: Watter definiëringsmodel dra by tot meer geslaagde onttrekking van inligting rakende sintaktiese kategorie en vertaalakkuraatheid? Die hoofokus van hierdie artikel is definisies in Engelse eentalige aanleerderswoordeboeke. Metaleksikograwe het 'n intense belangstelling in die volgende drie tipes definisies gehad: analitiese definisies (of klassieke definisies), volsindefiniesies (ook genoem kontekstuele definisies) en enkel-bysin *when*-definiesies. Die gebruik van *when*-definiesies, wat die definiëring van abstrakte selfstandige naamwoorde ten doel het, het vroeë en twyfel laat ontstaan oor hul effektiwiteit in die herkenning van die korrekte woordsoort van die *definiendum*, oftewel die item wat gedefinieer word, met inagneming van die probleme wat verband hou met die vervangbaarheid van trefwoorde en dele van definisies ('n gebrek aan algemeenekategoriewoorde in hierdie definisieformaat). Met betrekking tot

die akkuraatheid van woordsoortherkenning, het bestaande navorsing in hoofsaak die groter geslaagdheid van die klassieke definisie-tipe bo die enkel-bysin *when*-definisies bevestig. Die huidige eksperiment poog ook om verder ondersoek in te stel na woordsoortherkenning met betrekking tot die drie bogenoemde definisietipes — in vorige studies is slegs data van analitiese en enkel-bysin *when*-definisies met mekaar vergelyk, aangesien kontekstuele definisies nie deel gevorm het van die studiedoelwit nie. Hierdie studie is uitgevoer op 'n groep van 120 gevorderde vlak Poolse universiteitsstudente van Engels. Die proefpersone is met betrekking tot die drie hoofdefinisetipes in die Engelse leksikografiese praktyk getoets op hul vermoëns om inligting rakende sintaktiese kategorie korrek te onttrek asook op hul vertaalakkuraatheid van abstrakte selfstandige naamtrekwoorde.

Sleutelwoorde: AANLEERDERSWOORDEBOEKE, DEFINISIES, EENTALIGE WOORDEBOEKE, ANALITIESE DEFINISIES, VOLSINDEFINISIES, ENKEL-BYSIN *WHEN*-DEFINISIES

1. How can words be defined?

It goes without saying that meaning is the main reason why people decide to consult dictionaries. This rather unsurprising fact had been long since established in one of the earliest dictionary use questionnaires, which were conducted on Polish and American students (Tomaszczyk 1979), as well as French students of English (Béjoint 1981). In brief, people use dictionaries with a view to learning or acquiring newly encountered lexical items. Significantly, the meanings of words are conveyed to dictionary users through the use of definitions in dictionaries. Consequently, this means that definitions are one of the key features of dictionaries, as without them dictionaries would simply not be able to serve their primary function — which is providing English learners with pertinent information about word meanings.

So how can words be defined? According to Richards and Taylor (1992), there are various strategies that can be adopted. Synonyms, antonyms¹, taxonomic² definitions (for example, when we define the word "rugby" as "a sport"), as well as definitions by exemplification or function are only a handful of defining strategies employed by those of us who attempt to explain to someone the meaning of a given word. But there is no denying that the implementation of such simple and basic defining techniques by lexicographers might not be a satisfactory method in most cases — the meanings of words have to be explained more scrupulously if an average student is to fully grasp the meaning of a word. Hence, dictionary compilers need to strive to enhance the quality of their dictionary's definitions (the present context here applies to English monolingual learners' dictionaries) if users are not to become discontented with the level of the dictionaries.

The most important types of definitions that are commonly applied in professional lexicographic practice as regards English monolingual learners' dictionaries have been discussed in the following section.

2. Types of definitions in English monolingual learner's dictionaries

The analytical definition (also called a classical definition, or Aristotelian definition) is the most basic and standard type of definition in lexicographic practice; hence, "traditional definition" is yet another common term for this specific definition-type. The following constituent parts form an analytical definition: 1) *definiendum*; 2) *definiens*; 3) *genus proximum*; and 4) *differentiae specificae*. In simple terms, the *definiendum* is the word (or term) that is being defined, the *definiens* is the "right-hand side, defining, part of the definition" (Adamska-Salaciak 2012: 324), the *genus proximum* is the general category (superordinate) under which the item being defined can be classified, while the term *differentiae specificae* apply to the specific or distinguishing features of the *definiendum*. An example of the analytical⁴ definition has been demonstrated below:

house⁵ — a building⁶ that someone lives⁷ in, especially one that has more than one level and is intended to be used by one family (*Longman Dictionary of Contemporary English*, 6th edition)

Dziemianko and Lew (2013: 156) rightly notice that substitutability is an essential and inherent characteristic of the traditional definition-type. Put another way, the word class of the superordinate (general category word) needs to match the syntactic class of the *definiendum* — as can be seen from the example above, both "building" and "house" are nouns. Last but not least, classical definitions lend themselves to defining especially concrete nouns, as well as, for example, verbs of motion⁸ and verbs of making or creating (Atkins and Rundell 2008: 415).

In general, substitutability can be perceived as an advantageous defining strategy. Some metalexigraphers are of the opinion that substitution seems to be the right strategy that allows one to grasp the meaning of definitions (Fischer 1991) and also this definition-type makes it possible to describe "meanings (...) with precision and economy" (Atkins and Rundell 2008: 439). From a different perspective, research (Deese 1967; Miller 1985; Miller and Gildea 1985; Fischer 1991; Nesi and Meara 1994, Nesi 2000; Nesi and Hail 2002) has demonstrated that analytical definitions tend to result in incorrect word class recognition of the items that are being defined. One possible explanation for this phenomenon is the so-called "kidrule⁹ strategy" (Miller and Gildea 1985) — a simple strategy based on the reasoning that specific fragments (words) of definitions can replace the *definiendum* in different contexts. Not surprisingly, adhering to this strategy may prove to be an erroneous choice on numerous occasions.

The full-sentence definition is a signature feature¹⁰ of the *Collins Cobuild Advanced Learner's Dictionary* and has become a defining characteristic of the dictionary since the 1980s. An example of this type of definition is shown below:

resurgence — if¹¹ there is a resurgence of an attitude or activity, it reappears and grows (*Collins Cobuild Advanced Learner's Dictionary*, 8th edition)

The full-sentence definition (also called a contextual definition) has the following structure (Lew and Dziemianko 2006a: 228): 1) hinge (*if*); 2) co-text1 (*there is a*); 3) topic (*resurgence*); 4) co-text2 (*of an attitude or activity*); 5) matching framework (*it*); and 6) gloss (*reappears and grows*). In short, a full-sentence definition consists of two parts: left-hand side and right-hand side parts of the whole definition. The left-hand side part of the definition most importantly has a *hinge* (*if/when*), the item being defined (*topic*) and the surrounding environment or context in which the word being defined most typically appears (*co-text*). As for the right-hand side of the definition, there is a *matching framework* (*it* — this specific pronoun refers to the *topic*) and a *gloss* (it explains the meaning of the *topic*) — this suggests that the right-hand contextual frame of the definition provides learners with an explanatory comment as to what the word being defined specifically means. As for the initial part of the definition, one discovers more about the context of the item being defined; as an example, in this case we learn that the abstract noun *resurgence* most probably frequently occurs with the preposition *of*¹² (colligational preferences). All things considered, this is a real advantage of contextual definitions — as the name suggests, these definitions provide us with some context in which the given word routinely functions (we are provided with grammatical and collocational information of lexical items) and consequently we hone our linguistic production skills. An added advantage would be that dictionary users learn what words mean and how they are used in a more naturalistic setting — the words being defined are incorporated into sentences. As for the drawbacks, a few have been noted. One problem is that full-sentence definitions are lengthy — they tend to be longer than analytical definitions. Furthermore, Piotrowski (1994: 127) contends that on the contrary to traditional definitions, contextual definitions hinder the process of substitutability of full-sentence definitions for the item being defined. Rundell (2006: 326) contributes to the discussion by emphasizing the restrictiveness¹³ of full-sentence definitions, as they primarily describe only the most frequent instances of word use, or in other words they demonstrate only the most common contexts in which a given item appears, excluding the less common ones despite their undisputed significance and relevance in various different situations.

The third type of definition format which has left its mark in English monolingual learners' dictionaries is the so-called single-clause *when-definition*¹⁴, which can be characterized as being shorter than its double clause counterpart (full-sentence definition), it is used in folk-defining and conversation, and also sporadically in spontaneous defining as demonstrated in Fabiszewski-Jaworski's research (2011). Most importantly, however, it is perceived by lexicographers as an appealing and straightforward method of defining abstract nouns in dictionaries (for example, *Longman Dictionary of Contemporary English*, Cambridge

Advanced Learner's Dictionary, Cambridge Learner's Dictionary). More specifically, Lew and Dziemianko (2012: 997) describe the single-clause *when*-defining style as "a stand-alone relative clause introduced with the relative word *when*". An example of the single-clause *when*-definition for the entry "rebirth" is illustrated below:

rebirth — when something or someone becomes alive again after dying (*Longman Dictionary of Contemporary English*, 6th edition)

The schematic structure of the single-clause definition format above is the following (Lew and Dziemianko 2006a): 1) the word *when*; 2) co-text² (*something or someone*); and 3) gloss (*becomes alive again after dying*). The question which perhaps ought to be addressed¹⁵ is how does the single-clause *when*-definition of a word affect one's ability to correctly identify the syntactic category of this word. Put another way, the analytical definition-type is structured in a convenient way for dictionary users — learners can easily discover the word class of the item being defined through a simple and uncomplicated analysis of the general category word in the classical definition. However, the single-clause *when*-definition does not permit one to extract this type of information. Of course, the more aware language learners would most probably contrive to find this information in the entry itself, as dictionary entries are designed in a user-friendly way, and it is not rocket science to know that information about an item's part of speech is located at the very beginning of an entry. Nevertheless, the definition itself does not contain pertinent information about a given word's syntactic category and hence it would be interesting to see how this peculiarity of the single-clause *when*-definition influences learners' accuracy of part of speech recognition.

The following section elaborates on the topic of part of speech recognition and single-clause *when*-definitions in the context of empirical research that has been conducted in this field (mainly Dziemianko and Lew's studies). Also, the subject of syntactic class recognition and the single-clause *when*-defining format is the area of interest of this paper (the present study).

3. Empirical studies on syntactic category recognition of nominal head-words

Dziemianko and Lew conducted a series of studies (2006a; 2006b; 2013) dealing with the effect of definition-type on part of speech recognition in English monolingual learners' dictionaries. 129 native speakers of Polish representing an upper intermediate and advanced English level participated in their first experiment (2006a). The design of the study included 20 test items (10 abstract nouns and 10 distracters¹⁶ — 5 low-frequency verbs and 5 adjectives) which appeared with their definitions, half¹⁷ of which were analytical definitions and the remaining half were single-clause *when*-definitions (there were two different test versions, each test had a different assignment of definition-types to the

target items, each subject was given only one test version). Instead of using the actual headwords selected, pseudo words replaced the original items in the study in order to remove information about a word's part of speech that could have possibly been obtained from the morphological structure of words. The verbs and adjectives were included in the design of the experiment with a view to achieving some variation as to the part of speech of the items tested. As for the procedure, the subjects had to provide the Polish equivalents of the target items (the items appeared only with their definitions) and form English sentences with these items. Each student had 45 minutes to complete the tasks. Importantly, statistical significance was achieved in both types of tasks — the students managed to correctly identify a target item's part of speech 66.7% of the time in the case of analytical definitions and 33.2% of the time when dealing with single-clause *when*-definitions, while in the sentence formation task, the subjects met with success in 53.6% and 26.6% of the cases when being assisted by analytical and single-clause *when*-definitions respectively. Despite a definite advantage of the analytical definition format, Lew and Dziemianko stressed the need for more research¹⁸ as the microstructure adopted for their study lacked example sentences and, most importantly, grammatical information, which normally is incorporated into dictionary entries and creates an opportunity for dictionary users to discover an unknown word's syntactic class (part-of-speech labels).

A follow-up study (Dziemianko and Lew 2006b) was done on 238 Polish students of English who primarily had an intermediate level of proficiency in English. There were minor differences between this experiment and the previous one with respect to study design. First of all, in the present study a richer microstructure was incorporated into the dictionary entries — part-of-speech labels were included, as well as example sentences, syntactic codes and usage labels. Existing research (Bogaards and Van der Kloot 2002; Dziemianko 2006) has shown that grammar codes and examples have some significance when it comes to acquiring part-of-speech information from dictionary entries. Second, the task was different — this time the subjects were to complete a 45-minute multiple-choice task and they were provided with the possible answers (three Polish equivalents¹⁹ which were of a different part of speech — adjectives, nouns, verbs), which means that the subjects were given semantic information. In addition to this, there was no compose-sentence task, this meaning that the students were supposed to devote all their attention to syntactic class identification only. On the whole, the present subjects were exposed to a more naturalistic environment in light of having access to a more elaborate microstructure (however, there was no information about phonetic transcription), nevertheless, the task focusing solely on syntactic class recognition could also be perceived as an artificial one, not having much in common with natural and everyday dictionary consultation. Interestingly, the effect of defining-style format was statistically nonsignificant. Regardless of this finding, it is important to mention that analytical definitions slightly outperformed single-clause *when*-

definitions — word class identification scores amounted to 86.1% for the classical defining model and 85.4% for the single-clause *when*-defining model. These results suggest that Polish intermediate-level students can remain boastful about their dictionary reference skills — they possess the ability to successfully extract syntactic information from part-of-speech labels found in the microstructure of entries, and this skill allows them to dexterously compensate for *when*-definitions lacking in this specific type of information. Moreover, although there was only a slight difference, analytical definitions once again proved to be superior to the less common single-clause *when*-definitions with respect to the accuracy of correct syntactic class identification. In conclusion, Dziemianko and Lew felt the need for conducting yet another study — only this time they wanted to enrich the microstructure of entries by including information about phonetic transcription, as well as create a "less syntax-focused task²⁰" (Dziemianko and Lew 2006b: 862).

In their third study (2013), Dziemianko and Lew departed from adhering to identical microstructure designs that were applied in previous studies. To be more precise, they incorporated information about phonetic transcription in-between the lemma sign and syntactic class label, the aim of which was to minimize the salience of part-of-speech labels that was present in the second experiment. Also, more improvements were introduced in comparison with their first two studies through the implementation of "explicit grammatical information, style labels and examples of usage" (Dziemianko and Lew 2013: 164), all of this being done with a view to exposing the subjects to the most naturalistic environment of dictionary consultation possible. Furthermore, the subjects were asked to complete a 30-minute meaning-based task — provide the translation (single-word equivalent) of the English item into their native language. In the second study, the students were given the answers — they were provided with three Polish equivalents which were all of a different part of speech. The Polish equivalents were derivatives from the same root, which meant that the task was explicitly syntax-based. In other words, the subjects could have easily discovered that the task was grammar-oriented, or focusing on one's ability of part of speech recognition. The present meaning-based task was different in this respect. Apart from the modifications mentioned above, the study design did not differ much from the paradigm selected for the earlier studies. 134 subjects participated in the experiment who were Polish learners of English (upper-intermediate — advanced level of proficiency in the English language). Once again, the analytical definition-type achieved a higher score (90.1%) than the single-clause *when*-definition (87%) and this difference was found to be statistically significant²¹, however, in reality this was only a marginal difference of three percentage points and hence Dziemianko and Lew contend that the effect of defining style on part of speech recognition of abstract nouns is rather small. In their view, advanced dictionary users have enough reference skills to acquire syntactic class information from part-of-speech labels located within entries, rather than from definitions themselves.

Needless to say, a more complete microstructure plays an "important compensatory role²² (...) in POS identification" (Lew and Dziemianko 2012: 1002). Nevertheless, the findings also suggest that not only part-of-speech labels but also example sentences are a reliable source of information about the syntactic class of words — "syntactic class labels and examples obviously attract users' attention and offset the apparent syntactic emptiness of *when*-definitions (Dziemianko and Lew 2013: 169). On the whole, Dziemianko and Lew²³ are of the opinion that the use of single-clause *when*-definitions with a richer microstructure in English monolingual learners' dictionaries is reasonable, but perhaps excluding this definition-type from dictionaries might be a more shrewd decision in light of the fact that the subjects in the experiment were advanced students of English whose reference skills could have been rated as above average — this meaning that less advanced students could still perhaps encounter some difficulty with respect to the extraction of part-of-speech information from single-clause *when*-definitions embedded in abstract noun entries of richer microstructures. This view is in line with Atkins and Rundell's stance on single-clause *when*-definitions (this defining format ought not to be applied in English monolingual learners' dictionaries).

In the present context, perhaps the results of one more study should be briefly discussed. Fabiszewski-Jaworski and Grochocka (2010) experimented the effect of definition-type on part of speech recognition on 150 upper-intermediate-level (level of proficiency in English) native speakers of Polish. The task was to provide Polish equivalents of English target items appearing with either analytical or single-clause *when*-definitions. Not surprisingly, analytical definitions scored significantly higher (33.3%) than single-clause *when*-definitions (26.2%) with respect to the accuracy of part of speech recognition. Most importantly, Fabiszewski-Jaworski and Grochocka explain that the inclusion of *when*-definitions in dictionaries seems to be a logical solution especially when dictionary compilers encounter problems with finding the right general category words required for defining abstract concepts.

In the following section, the author discusses the current study's research design.

4. Research design

Dziemianko and Lew's studies inspired the present paper and served as a paradigm for the research design in the current experiment. The aim was to test the usefulness of the most common definition-types in English monolingual pedagogical dictionaries for advanced learners. Two research questions were addressed:

- (1) Which defining-model (analytical definitions vs. full-sentence definitions vs. single-clause *when*-definitions) contributes to more successful extraction of syntactic class information from abstract noun entries?

- (2) Which defining model (analytical definitions vs. full-sentence definitions vs. single-clause *when*-definitions) contributes to higher translation accuracy with respect to abstract noun entries?

Similarly to previous studies, the analytical and single-clause *when* definition-types were selected for the analysis as intuition suggests that the genus and differentia model of the classical definition is conducive to leading to more effective extraction of part-of-speech information from abstract noun entries rather than the single-clause *when*-defining format. As for two-clause²⁴ *when*-definitions, it is possible that this specific definition-type can be beneficial to dictionary users with respect to the accuracy of part-of-speech recognition as its inherent nature provides learners not only with some general contextual information but also grammatical information, which is a key factor in deriving pertinent part-of-speech information from dictionary entries. However, as no attempt had been previously made to test the effectiveness of the full-sentence definition format on the accuracy of part of speech recognition, the contextual definition-type was added to the design of the present study and a null hypothesis that there would be no relationship or association among the three groups (analytical definitions/full-sentence definitions/single-clause *when*-definitions) was assumed to be true. Put another way, the effect of definition-type (independent variable) on both syntactic class recognition and also translation accuracy would turn out to be nonsignificant — a separate one-way GLM ANOVA was run for each dependent variable — (1) syntactic class recognition; and (2) translation accuracy. The data were calculated in SPSS (version 25).

24 test items were selected for the study — 12 abstract nouns, 6 low-frequency verbs and 6 low-frequency adjectives. The verbs (*ensnare, glisten, lash, wring, yank, devour*) and adjectives (*obnoxious, adamant, comely, concomitant, egregious, fecund*) which were used in the study were the distracters (data were not collected from these items) and their role was to avoid having subjects focus solely on the part of speech of the items (grammar-oriented tasks), as this was primarily supposed to be a semantic-based²⁵ task. Also, another function of the distracters was to reduce the saliency of the target items in the whole study. The 12 nouns (*dexterity, disturbance, omen, deceit, disquiet, peculiarity, quirk, abstraction, compliance, legislation, apprehension, infatuation*) used in the experiment were replaced with pseudo words (*tiezon, menave, conluse, sardy, reprive, tortex, overlar, arouch, vargin, mortap, sharpeg, barrex*) — the aim of such a study design being that any morphological information carried by nouns had to be removed from context, so that the subjects would not take advantage of their existing knowledge about the language, or derivational information about words, which would most probably allow them to easily identify the part of speech of the test items. Nonexistent words, however, did not replace the verbs and adjectives used in the study. The pseudo words (nouns) were generated by a nonword-generating program, called WordGen²⁶ (Duyck, Desmet, Verbeke and Brysbaert 2004). All of the test items (24 items) were assigned random loca-

tions within specific test versions thanks to the Random Integer Generator tool available online and free of charge. There were three different test versions. Each test version²⁷ consisted of 12 nouns (pseudo words), 6 verbs and 6 adjectives. Four nouns in one test version appeared with analytical definitions, the other four nouns with full-sentence definitions and the remaining four with single-clause *when*-definitions. The assignment of specific definition-types to nouns was rotated across different test versions and as a result a cross-balanced design was achieved (confounding effect of item and subject were reduced to a minimum, each subject was exposed to a specific defining style of the target items $\frac{1}{3}$ of the time). Each single test item (headword) in the test formed an entry with its own microstructure. In order for the aim of the study to be met, a richer microstructure²⁸, which was based on the one applied in Dziemianko and Lew's most recent study (2013), was adopted for the experiment. To be more precise, the headword/test item (appearing in boldface font) was followed by pronunciation information (prepared by the author of the study), part-of-speech labels, grammar codes, usage labels, definitions and example sentences. The aim of the incorporation of such a microstructure was to create a naturalistic²⁹ environment for dictionary use consultation, which of course can never be fully achieved under artificial experimental conditions. The lexicographic data for the definitions and examples used in the study were taken from various English monolingual learners' dictionaries: LDOCE online, COBUILD8, OALD9, CALD4, MED2 and MWALED. Sporadically, the author had to slightly modify some of the definitions³⁰ and example sentences. Moreover, the pronunciation and grammatical information in entries was based on lexicographic information from LDOCE online.

There were 120 subjects (males and females) who were native speakers of Polish. Their English proficiency level had been assessed as upper-intermediate or advanced, and they were third-year and fourth-year students of English at a Polish university (University of Warmia and Mazury in Olsztyn). The subjects were asked to provide a one-word Polish equivalent of the English items in the spaces provided (this was not a multiple-choice task). The subjects had 45 minutes to complete the task during their regular class at the university. Prior to the experiment, they were briefly instructed by the experimenter (the subjects were instructed orally and they were provided with a Polish instruction in written form) and the subjects were also told how much time they would have for the completion of the whole task.

As for the grading system, the subjects' answers were assessed separately for syntactic class recognition and translation accuracy. In the case of the former, subjects would receive a score of "1" only when being able to provide the correct part of speech of the target item — the word written down in the answer sheet had to be a noun. The meaning of this noun was not taken into consideration but only its grammatical category. In the case of the latter, this time not the part of speech but the meaning of the word was most important. The answers that were considered to be correct were not only the Polish

equivalents of the target item found in the dictionary, but also other answers had to be taken into account as being possibly correct in the present context, as the subjects were supposed to infer what the correct equivalent was only on the basis of the type of information which they had access to in the task given — more specifically, the microstructure of entries. The Polish equivalents of the test items, which were marked as either correct or incorrect in the experiment (part of speech recognition and translation accuracy), are listed in Table 1 in the following section.

5. Results

The mean results for syntactic class identification accuracy by definition-type are illustrated below in Figure 1.

Figure 1: Mean syntactic class identification accuracy by definition-type

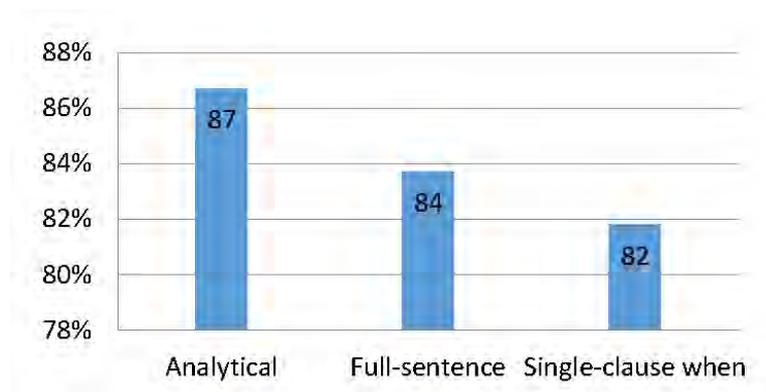


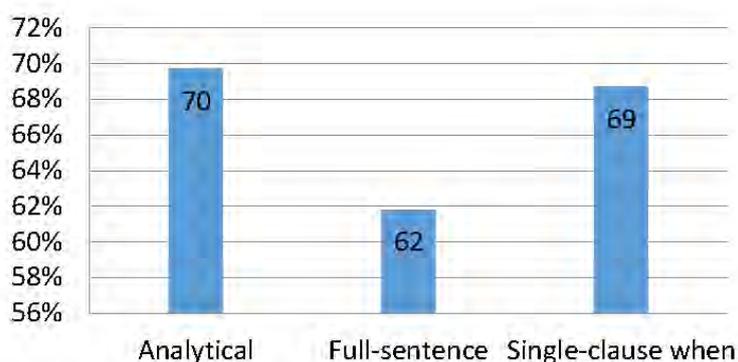
Figure 1 confirms the superiority of the analytical definition-type over the single-clause *when*-defining format. The subjects managed to correctly select the part of speech of a headword when dealing with analytical definitions in 86.6% of the cases, while when being exposed to single-clause *when*-definitions the accuracy rate amounted to 81.8%, which means that there was a difference of almost 5 percentage points (4.8%) between the two defining styles. As for full-sentence definitions, altogether the subjects achieved a score of 83.7%, which indicates that contextual definitions only slightly outperformed single-clause *when*-definitions by approximately 2 percentage points (1.9%), but fared worse than the classical definition by almost 3 percentage points (2.9%). The success rate for individual items ranged between 66%–95% — the test items *compliance* (*vargin*) and *dexterity* (*tiezon*) having an average of 66% and 95% respectively.

The statistical analysis revealed that the effect of definition-type on syntactic class identification accuracy was nonsignificant³¹ (one-way GLM ANOVA,

$F_{(2,1437)}=2.09$, $p=0.124$). Moreover, there was very little practical significance and the effect size was very small ($\eta^2 = 0.003$). Hence, the null hypothesis was not rejected.

The mean results for translation accuracy by definition-type are demonstrated in Figure 2.

Figure 2: Mean translation accuracy by definition-type



Once again it was the analytical definition which contributed to the highest scores. Figure 2 reveals that the analytical definition-type helped the subjects in the translation tasks (English into Polish) more than the remaining two types of definitions. Exposure to analytical definitions led to successful translations in 69.7% of the cases, which was only a marginally better score than the 68.7% success rate of single clause *when*-definitions. Interestingly, the single-clause *when*-defining style proved more beneficial than the two-clause *when*-definition format in the translation tasks, outperforming full-sentence definitions by almost 7 percentage points (translation accuracy amounted to 68.7% for analytical definitions, whereas for full-sentence definitions it amounted to only 61.8%). Overall, translation accuracy ranged between 42%–91% for individual test items — the test items *abstraction (arouch)* and *apprehension (sharpeg)* had a mean of 42% and 91% respectively.

In this case, the null hypothesis was rejected. A separate ANOVA was run for translation accuracy and the analysis showed that there was a statistically significance³² difference among the groups (one-way GLM ANOVA, $F_{(2,1437)}=4.02$, $p=0.018$), which strongly suggested that one or more pairs of treatments were significantly different. Once again, there was little practical significance and the effect size proved to be very small ($\eta^2 = 0.006$). The Tukey HSD test revealed that there was a statistically significant difference between analytical and full-sentence definitions with respect to translation accuracy ($p=0.025<.05$), while the difference between full-sentence and single-clause *when*-definitions approached statistical significance ($p=0.061>.05$). No statistical significance was noted

between the analytical and single-clause *when*-definition types ($p=0.899>.05$).

Given the subjects' answers in the test, the most frequent Polish equivalents of the test items that were provided by the subjects have been listed in Table 1.

Table 1: Subjects' answers — Polish equivalents of the test items

PSEUDO WORD (TEST ITEM)	ANSWERS — POLISH EQUIVALENTS OF TEST ITEMS
CONLUSE (OMEN)	ZNAK, ZAPOWIEDŹ, OZNAKA, ZWIASTUN, PRZEPOWIEDNIA, OMEN, OBJAWIENIE, PROGNOZYK, WRÓŻBA
VARGIN (COMPLIANCE)	ZGODNOŚĆ, POSŁUSZEŃSTWO, PODPORZĄDKOWANIE, PRZESTRZEGAĆ, NASTĘPSTWO, PRYZWOLENIE, PRZECIWKO, PRZYSIĘGAĆ, WYKONAĆ, PRZYMUS, POSZANOWANIE, PRAWIDŁOWOŚĆ
SHARPEG (APPREHENSION)	LĘK, PRZECZUCIE, OBAWA, NIEPOKÓJ, ZŁE PRZECZUCIE, ZANIEPOKOJENIE, PRZECZUWAĆ, PRZERAŻENIE
REPRIVE (DISQUIET)	NIEZADOWOLENIE, NIEPOKÓJ, ZANIEPOKOJENIE, ZAWÓD, PRZYGNĘBIENIE, WĄTLIWOŚĆ, LĘK, ZMARTWIĆ SIĘ, PRZEJĘTY
AROUC (ABSTRACTION)	ODDZIELENIE, SEPARACJA, ODŁĄCZENIE, ODEBRANIE, ZNIESIENIE, POZBAWIĆ, ROZŁAM, USUWAĆ
TORTEX (PECULIARITY)	DZIWACTION, ODMIENNOŚĆ, EKSCENTRYZM, PRZYPADŁOŚĆ, ODCHYLENIE, BZIK, PRYZYWYCZAJENIE, FETYSZ, DZIWNY
MORTAP (LEGISLATION)	UCHWAŁA, USTAWA, USTAWODAWSTWO, LEGISLACJA, ZASADA, KODEKS, ZBIÓR, POPRAWKA, UCHWALIĆ
BARREX (INFATUATION)	ZAUROCZENIE, ZADURZENIE, ZAANGAŻOWANIE, FASCYNACJA, POCIĄG, OBSESJA, ZAURCZYĆ SIĘ
OVERLAR (QUIRK)	PRZYPADEK, KAPRYS, ANOMALIA, ZJAWISKO, CUD, RZADKOŚĆ, NADZWYCZAJNE
SARDY (DECEIT)	OSZUSTWO, PODSTĘP, MANIPULACJA, OSZUKAĆ, SABOTAŻ, MANIPULOWAĆ, FAŁSZERSTWO, PROPAGANDA, KŁAMSTWO
TIEZON (DEXTERITY)	SPRAWNOŚĆ, ZRĘCZNOŚĆ, UMIEJĘTNOŚĆ, SPRYT, KOORDYNACJA, ZDOLNOŚĆ
MENAVE (DISTURBANCE)	ZAKŁÓCENIE, PRZERWANIE, ZMIANA, PRZERYWNIK, ZANIECHANIE, ROZPROSZENIE, ZAKŁÓCAĆ, NARUSZAĆ

6. Discussion and conclusions

To reiterate, Dziemianko and Lew's series of studies (2006a; 2006b; 2013) functioned as the paradigm for the current research. In brief, each consecutive experiment endeavored to correct the previous study design and as a result adopt a more satisfactory and practical design, one which would eliminate the artificiality of the tasks at hand and at the same time provide a more natural dictionary use environment. Innovations involved including a richer micro-structure with not only the definitions of words but also examples, syntactic

class labels, grammar codes, usage labels and information about pronunciation. Phonetic transcription was located in-between the headwords and part-of-speech labels, which increased the level of naturalness of the tasks by minimizing the saliency of syntactic class labels. Moreover, Dziemianko and Lew shifted away from applying grammar-oriented tasks and opted for using meaning-based tasks in their most recent experimental design. Hence, the aim of the present study was to test the general usefulness of the most common defining styles in pedagogical monolingual dictionaries for learners of English with respect to syntactic class identification and translation accuracy, by drawing from Dziemianko and Lew's conclusions. Significantly, as no significant existing study had previously incorporated full-sentence definitions into its design, contextual definitions were included in the present experiment and were treated as one of the three levels of the independent variable — definition-type — the two other levels being the default analytical and single-clause *when*-definitions. Also, another modification in the study design involved the introduction of an additional dependent variable — translation accuracy. Therefore, the effect of definition-type on the subjects' accuracy of translation from English into Polish was tested.

The current study demonstrates that analytical definitions hold a clear advantage over full-sentence and single-clause *when*-definitions (Research question 1: Which defining-model contributes to more successful extraction of syntactic class information from abstract noun entries?). This finding is in line with Dziemianko and Lew's three studies (2006a; 2006b; 2013), as well as Fabiszewski-Jaworski and Grochocka's (2010) experiment, where in all four studies the classical definition format proved superior to the less common and newer in pedagogical dictionaries for learners of English single-clause *when*-definition. In the present study, in the context of part-of-speech identification accuracy, exposure to analytical definitions led to success in 86.6% of the cases, whereas headwords which were defined through full-sentence and single-clause *when*-definitions achieved a score of 83.7% and 81.8% respectively. Notwithstanding the lack of a statistically significant difference among the three groups, these numbers suggest the analytical definition's dominance, especially over single-clause *when*-definitions, bearing in mind that classical definitions outperformed single-clause *when*-definitions with regard to syntactic class identification accuracy in all of the earlier studies. It seems, then, that the oldest and most traditional way of defining words is most beneficial to dictionary users. The genus and differentia paradigm allows for the substitutability of the genus expression and the headword (when the *genus proximum* and the headword are of the same word class), which most apparently suits the user more than any other existing and practiced-in-lexicography defining model. The data also suggest that two-clause *when*-definitions are slightly more useful when it comes to the extraction of part-of-speech information from entries than their single-clause counterparts, this finding perhaps being unsurprising as full-sentence definitions are lengthier definition-types which contain more contex-

tual information (also grammatical information) about the word being defined.

The mean part-of-speech identification accuracy by definition-type of Dziemianko and Lew's studies (2006a; 2006b; 2013) and the present study is summarized in Table 2.

Table 2: Mean syntactic class identification accuracy by definition-type — a comparison of Dziemianko and Lew's studies (2006a; 2006b; 2013) and the present study

STUDY/DEFINITION-TYPE	ANALYTICAL	SINGLE-CLAUSE <i>WHEN</i>	FULL-SENTENCE
STUDY 1 ³³	66.7%	33.2%	NO EMPIRICAL DATA
STUDY 2	86.1%	85.4%	NO EMPIRICAL DATA
STUDY 3	90.1%	87.0%	NO EMPIRICAL DATA
PRESENT STUDY	86.6%	81.8%	83.7%

Similarly to Dziemianko and Lew's study (2013), the study participants in the current experiment were not asked to underline any fragments of the entry during entry consultation and hence it was not possible to draw any conclusions or infer which specific parts of entries were most useful for the part-of-speech recognition task. However, a detailed analysis of Table 2 indicates that the syntactic class recognition scores ranged between approximately 33%–67% in the first study and 82%–90% in the second, third and present study. Consequently, the evidence supports the conclusion that a more sophisticated micro-structure enhances the extraction of syntactic class information from entries regardless of the given definition-type at hand, in light of the fact that studies 2 and 3, as well as the current study, all incorporated a more elaborate micro-structure into the experimental design. As a result, the data demonstrate that it is highly likely that the presence of, for example, syntactic class labels and example sentences in dictionary entries enhances the process of extracting correct part-of-speech information from abstract nominal headwords. This finding confirms the conclusion from Dziemianko and Lew's study (2013) that complete entries allow more proficient dictionary users to acquire pertinent information about a word's part of speech even when being exposed to the syntactically impoverished single-clause *when*-definitions. As for full-sentence definitions, there is some likelihood that a higher amount of contextual information in this definition format, for example, grammatical information, compensates for the absence of the abundance of such information from single-clause *when*-definitions and perhaps for this reason success rates for the contextual definition-type slightly exceeded the scores for the single-clause *when*-definition by barely 1.9%. However, the lack of such empirical evidence in other studies does not allow to arrive at any correct conclusions. On balance, despite everything said, it would seem wise to agree with Fabiszewski-Jaworski and Grochocka (2010)

that perhaps the analytical definition is the safer option for lexicographers while single-clause *when*-definitions can be helpful when the genus and differentia model of the classical definition does not stand up to its expectations. The same should possibly apply to full-sentence definitions, nevertheless, it must be admitted that there were only marginal differences in the experiment among all three means (range of 81.8%–86.6%), which in turn implies that the three aforementioned definition-types are comparably effective with respect to the accuracy of part of speech recognition.

As far as translation accuracy is concerned (Research question 2: Which defining model contributes to higher translation accuracy with respect to abstract noun entries?), exposure to full-sentence definitions (61.8%) led to the lowest scores of all three types of definitions (69.7% — analytical definitions; 68.7% — single clause *when*-definitions) analyzed in the study. First and foremost, single-clause *when*-definitions fared worse than the analytical definition by only 1 percentage point. There is no doubt that such a minor difference cannot have any significance at all in the present context. This finding perhaps is indicative of the fact that both analytical and *when*-definitions are equally beneficial defining styles in relation to meaning-oriented tasks such as translating from the target language to one's mother tongue. In other words, traditional definitions are more valuable for learners than single-clause *when*-definitions when it comes to identifying the correct part of speech of a headword, however, this definite advantage of the classical definition seems to be counterbalanced in more meaning-related tasks. Second, it is possible that the non-substitutability of full-sentence definitions hinders the process of correct translation (despite the fact that the word class of the Polish equivalents in the present experiment was not taken into account when assessing and assigning scores to the Polish equivalents in the translation task), a drawback generally (but only by principle) non-existent in analytical definitions, which heavily rely on the genus and differentia model. Nevertheless, *when*-definitions also suffer from this inconvenience and yet this defining style contrived to achieve comparable if not almost identical results in comparison with the Aristotelian defining style. One possible explanation is that Polish learners of English do not really tend to define words in their native language through the use of contextual definitions (Mikołajczak-Matyja 1998), and also this specific definition-type is not widely applied in Polish lexicography — hence the problems that the subjects in the present study might have encountered in the target language. Another possibility is that full-sentence definitions could be perceived as too long or wordy, whereas analytical definitions and single-clause *when*-definitions seem to be shorter and concise, or simply more straightforward, lacking in more complex syntactic structures. Moreover, we cannot count out the learner variable. English learners differ from one another in many respects — cultural background, intelligence levels, motivation, linguistic abilities, etc. Some of these factors may have influenced the scores achieved in the translation task. Finally, the lexicographic information at hand might have played its

part in the present experiment. Some concepts expressed in the subjects' second language could have been either easier or more difficult to render into Polish by the study participants.

One of the findings from Dziemianko and Lew's study (2013) was that single-clause *when*-definitions which fit the pattern *when + indefinite pronoun (someone/something)* had a rather negative effect on part-of-speech recognition accuracy of abstract nouns, especially when being compared to *when*-definitions which take the *when + personal pronoun/noun phrase* structure. Of the two types of definitions mentioned above, it was the latter that contributed to higher syntactic class identification success rates, or results that were nearly as successful as the scores of the analytical definition-type in the experiment. In the current study, the item analysis of single-clause *when*-definition scores reveals a marked tendency for the *when + personal pronoun/noun phrase when*-definitions to outperform their *when + indefinite pronoun* counterpart. The data³⁴ are gathered in Table 3.

Table 3: Mean part of speech identification and translation accuracy by test item — results of single-clause *when*-definitions

PSEUDO WORD (TEST ITEM)	TEST VERSION	POS IDENTIFICATION ACCURACY	TRANSLATION ACCURACY	TYPE OF SINGLE-CLAUSE <i>WHEN</i> -DEFINITION
CONLUSE (OMEN)	V1	100.0%	77.5%	WHEN + THERE IS + NP
VARGIN (COMPLIANCE)	V1	65.0%	37.5%	WHEN + INDEFINITE PRONOUN
SHARPEG (APPREHENSION)	V1	77.5%	82.5%	WHEN + PERSONAL PRONOUN/NP
REPRIVE (DISQUIET)	V1	72.5%	70.0%	WHEN + PERSONAL PRONOUN/NP
AROUCH (ABSTRACTION)	V2	77.5%	55.0%	WHEN + INDEFINITE PRONOUN
TORTEX (PECULIARITY)	V2	82.5%	70.0%	WHEN + INDEFINITE PRONOUN
MORTAP (LEGISLATION)	V2	85.0%	67.5%	WHEN + PERSONAL PRONOUN/NP
BARREX (INFATUATION)	V2	95.0%	90.0%	WHEN + PERSONAL PRONOUN/NP
OVERLAR (QUIRK)	V3	85.0%	57.5%	WHEN + INDEFINITE PRONOUN
SARDY (DECEIT)	V3	70.0%	57.5%	WHEN + INDEFINITE PRONOUN
TIEZON (DEXTERITY)	V3	95.0%	85.0%	WHEN + PERSONAL PRONOUN/NP
MENAVE (DISTURBANCE)	V3	77.5%	75.0%	WHEN + PERSONAL PRONOUN/NP

The mean for part of speech identification accuracy of the *when + personal pronoun/noun phrase* definition-type was by approximately 7.8 percentage points higher (83.8%) than the mean of the *when + indefinite pronoun* definition (76%). Likewise, taking into consideration translation accuracy, the *when + personal pronoun/noun phrase* definition pattern had better scores by as many as 22.8 percentage points (78.3%) than the *when + indefinite pronoun* single-clause *when*-definition (55.5%). As illustrated above in Table 3, the *when + personal pronoun/noun phrase* defining style clearly dominated the *when + indefinite pronoun* definition format within specific test versions (there were three different test versions). In the case of part of speech identification accuracy, the item *overlar* (*quirk*) was one exception, as this item achieved an accuracy of 85%, while the item *menawe* (*disturbance*) had an accuracy of only 77.5% in test version no. 3. A similar pattern occurred with respect to the accuracy of translating target items from English into Polish, only this time it occurred in test version no. 2 — the test item *tortex* (*peculiarity*) had an accuracy of 70%, while the item *mortap* (*legislation*) performed worse by about 2.5% (67.5%). By and large, these observations indicate that the *when + indefinite pronoun* definition-type may indeed be the inherently weaker defining style of the two defining single-clause *when*-formats discussed above. In order to see whether this is actually true, an additional study would be needed — one with an experimental design suited for testing and comparing the effectiveness of the *when + indefinite pronoun* and *when + personal pronoun/noun phrase* defining styles.

To sum up, more empirical research is required if any right conclusions are to be reached about the role of full-sentence definitions on syntactic class recognition or translation accuracy. At the present moment, it seems like it would be worthwhile to undertake research into the usefulness of the single-clause *when*-defining style in the context of part-of-speech extraction of information from entries. More specifically, the single-clause *when + personal pronoun/noun phrase* and *when + indefinite pronoun* (*someone/something*) defining models could be tested experimentally and contrasted with one another.

Endnotes

1. Synonyms and antonyms — definitions which are one-word equivalents of the item being defined.
2. Taxonomic definitions involve classifying words into classes or subclasses.
3. The term *genus proximum* is also called the "genus expression" (Atkins and Rundell 2008: 414).
4. For more information about analytical definitions see Adamska-Salaciak (2012).
5. *Definiendum*.
6. *Genus proximum*.
7. *Differentiae specificaе* — the distinguishing features of this specific building are: 1) it is a building that someone lives in; 2) it is a building that has more than one level; 3) it is intended to be used by one family.

8. Atkins and Rundell (2008) mention that the genus expression "walk" can be used for defining words such as "stroll" or "tiptoe" (verbs of motion), while the genus expression "copy" can be the superordinate for the words "reproduce" or "photocopy" (verbs of making or creating).
9. "Kidrule" — a rule applied by children — hence the term "kidrule".
10. From a historical point of view, the *Collins Cobuild Dictionary* is also known for: (1) being the first corpus-based English learners' dictionary; and (2) its grammatical column.
11. Hinge — usually *if/when* form the hinge in full-sentence definitions.
12. Another common grammatical pattern is *resurgence in*.
13. This problem has been termed by Rundell as "overspecification".
14. The single-clause *when*-definition begins with the word "when".
15. This subject has been the focus of a few studies conducted by Anna Dziemianko and Robert Lew (this topic is also the primary focus of the present paper).
16. The distracters appeared only with analytical definitions and they were not included in the analysis.
17. Half of the definitions of the ten target items (abstract nouns).
18. One suggestion was to include a richer microstructure in the next study, while the other one was that it seemed to "be worthwhile to further extend the scope of the study and compare *when*-definitions not only with analytical, but also contextual ones" (Lew and Dziemianko 2006a: 237). The current study extends the scope of Dziemianko and Lew's studies (2006a; 2006b; 2013) by introducing the full-sentence definition into the design of the study.
19. These equivalents were derivatives from the same root.
20. The intention was to reduce the salience of part of speech information (syntactic labels) by having this type of information separated from the lemma sign through the inclusion of phonetic transcription information.
21. However, the effect size was small and hence there was little practical significance.
22. This is especially true of single-clause *when*-definitions (Dziemianko and Lew 2013: 169) and not necessarily analytical definitions. When comparing Dziemianko and Lew's second (2006b) and third study (2013), the success rate tripled for *when*-definitions in a rich-microstructure environment.
23. Also, Dziemianko and Lew (2013) notice that single-clause *when*-definitions can especially be misleading when the following structure of these definitions is applied: *when* + indefinite pronoun (*someone/something*). They imply that whenever possible single-clause *when*-definitions should perhaps adopt a different structure: *when* + personal pronoun/nominal phrase. Importantly, single-clause *when*-definitions which take the *when* + indefinite pronoun (*someone/something*) structure decrease part-of-speech recognition scores even in more elaborate microstructures.
24. Full-sentence definitions in other words.
25. In this regard, the study was no different from Dziemianko and Lew's contribution (2013) to the topic of study.
26. WordGen is a simple tool which is based on the CELEX and Lexique lexical databases. Its main function is to select words and generate nonwords not only in the English language, but also in Dutch, German and French.
27. By contrast, in Dziemianko and Lew's three studies there were always 20 test items: 10 abstract nouns, 5 verbs, 5 adjectives.

28. See Appendix at the end of the paper to acquire more information about the microstructure applied in the present experiment.
29. The order of specific types of information appearing in the microstructure of entries that was adopted was also based on Dziemianko and Lew's study (2013) with a view to avoiding having a typical syntax-based task. Hence, pronunciation information separated the lemma sign from the part-of-speech labels and in this respect a more natural setting of dictionary use was created for the subjects in the current study.
30. For example, due to the lack of a full-sentence definition for a specific word in various dictionaries, the author invented his own example of such a definition, on the basis of the lexicographic data in the dictionary. One example includes the test item *disturbance* (*menave*). The following definition for the noun entry *disturbance* can be found in the *Oxford Advanced Learner's Dictionary of Current English* (9th edition): "actions that make you stop what you are doing, or that upset the normal state that something is in". This definition was changed into: "a *menave* (*disturbance*) is an action that makes you stop what you are doing, or that upsets the normal state that something is in".
31. The effect was nonsignificant at the significance level of 0.05.
32. The effect was statistically significant at the significance level of 0.05.
33. Study 1 — Lew and Dziemianko (2006a); Study 2 — Dziemianko and Lew (2006b); Study 3 — Dziemianko and Lew (2013).
34. As this analysis was not the primary aim of the paper, it has been mentioned and elaborated on in the Discussion and Conclusions section.

References

Dictionaries

- Delacroix, L. (Ed.).** 2014. *Longman Dictionary of Contemporary English*. Sixth edition. Harlow: Longman. [LDOCE6]
- Delacroix, L. (Ed.).** 2014. *Longman Dictionary of Contemporary English*. Sixth edition (free online version). Harlow: Longman = <https://www.ldoceonline.com/>. [LDOCE online]
- Deuter, M. (Ed.).** 2015. *Oxford Advanced Learner's Dictionary of Current English*. Ninth edition. Oxford: Oxford University Press.
- Gillard, P. (Ed.).** 2003. *Cambridge Advanced Learner's Dictionary*. First edition. Cambridge: Cambridge University Press. [CALD1]
- Gillard, P. (Ed.).** 2004. *Cambridge Learner's Dictionary*. Second edition. Cambridge: Cambridge University Press. [CALD2]
- McIntosh, C. (Ed.).** 2013. *Cambridge Advanced Learner's Dictionary*. Fourth edition. Cambridge: Cambridge University Press. [CALD4]
- Perrault, S.J. (Ed.).** 2008. *Merriam-Webster's Advanced Learner's English Dictionary*. Springfield: Merriam-Webster. [MWALED]
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan Education. [MED2]
- Sinclair, J. (Ed.).** 2014. *Collins COBUILD Advanced Learner's Dictionary*. Eighth edition. London: HarperCollins. [COBUILD8]

- Summers, D. (Ed.).** 2003. *Longman Dictionary of Contemporary English*. Fourth edition. Harlow: Longman. [LDOCE4]
- Walter, E. (Ed.).** 2005. *Cambridge Advanced Learner's Dictionary*. Second edition. Cambridge: Cambridge University Press. [CALD2]

Other literature

- Adamska-Salaciak, A.** 2012. Dictionary Definitions: Problems and Solutions. *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 129: 323-339.
- Atkins, S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Béjoint, H.** 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.
- Bogaards, P. and W.A. Van der Kloot.** 2002. Verb Constructions in Learners' Dictionaries. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*. Vol. II: 747-757. Copenhagen: Center for Sprogteknologi, Copenhagen University.
- Deese, J.** 1967. Meaning and Change of Meaning. *American Psychologist* 22(8): 641-651.
- Duyck, W., T. Desmet, L. Verbeke and M. Brysbaert.** 2004. WordGen: A Tool for Word Selection and Nonword Generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers* 36(3): 488-499.
- Dziemianko, A.** 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Tübingen: Max Niemeyer Verlag.
- Dziemianko, A. and R. Lew.** 2006b. When You are Explaining the Meaning of a Word: The Effect of Abstract Noun Definition Format on Syntactic Class Identification. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, Torino, Italy, September 6–9, 2006*: 857-863. Allessandria: Edizioni dell'Orso.
- Dziemianko, A. and R. Lew.** 2013. When-Definitions Revisited. *International Journal of Lexicography* 26(2): 154-175.
- Fabiszewski-Jaworski, M.** 2011. Spontaneous Defining by Native Speakers of English. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto Terra, Kyoto, Japan, August 22–24, 2011*: 102-109. Kyoto: The Asian Association for Lexicography.
- Fabiszewski-Jaworski, M. and M. Grochocka.** 2010. Folk Defining Strategies vs. Comprehension of Dictionary Definitions: An Empirical Study. Kernerman, I. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DNSA 2009*: 89-105. Tel Aviv: K Dictionaries.
- Fischer, U.** 1991. How Students Misunderstand Definitions: Some Evidence for Contextual Representations of Word Meanings. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, 7–10 August 1991, Chicago, Illinois*: [no pagination]. Hillsdale: Lawrence Erlbaum.
- IBM SPSS Statistics Package.** Version 25.
- Lew, R. and A. Dziemianko.** 2006a. A New Type of Folk-inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography* 19(3): 225-242.

- Lew, R. and A. Dziemiątko.** 2012. Single-clause *when*-definitions: Take Three. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August 2012, Oslo*: 997-1002. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Mikołajczak-Matyja, N.** 1998. *Definiowanie pojęć przez przeciętnych użytkowników języka i przez leksykografów*. Poznań: Sorus.
- Miller, G.** 1985. Dictionaries of the Mind. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*: 305-314. Chicago: University of Chicago.
- Miller, G. and P. Gildea.** 1985. How to Misread a Dictionary. *AILA Bulletin*, 1985: 13-26.
- Nesi, H.** 2000. *The Use and Abuse of EFL Dictionaries: How Learners of English as a Foreign Language Read and Interpret Dictionary Entries*. Tübingen: Max Niemeyer Verlag.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
- Nesi, H. and P. Meara.** 1994. Patterns of Misinterpretation in the Productive Use of EFL Dictionary Definitions. *System* 22(1): 1-15.
- Piotrowski, T.** 1994. *Z Zagadnień Leksykografii*. Warsaw: Państwowe Wydawnictwo Naukowe.
- Random Integer Generator.** <https://www.random.org/integers/>.
- Richards, J. and A. Taylor.** 1992. Defining Strategies in Folk Definitions. *Perspectives (City University of Hong Kong)* 4: 1-8.
- Rundell, M.** 2006. More than One Way to Skin a Cat: Why Full-sentence Definitions have not been Universally Adopted. Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, Torino, Italy, September 6–9, 2006*: 323-338. Allessandria: Edizioni dell'Orso.
- Tomaszczyk, J.** 1979. Dictionaries: Users and Uses. *Glottodidactica* 12: 103-119.

APPENDIX: THE STUDY — SAMPLE 5 TEST ITEMS (VERSION 1)

INSTRUKCJA: Test zawiera 24 wyrazy w języku angielskim (są to raczej wyrazy trudne). Każdy wyraz wraz z podaną informacją o tym wyrazie tworzy tzw. hasło słownikowe. Twoim zadaniem jest przetłumaczyć wyrazy angielskie na język polski. Uwaga – polskie odpowiedniki angielskich wyrazów muszą być wyrazami jednowyrazowymi!

1	tiezon ▶ /'taɪzən/ <i>noun</i> [U] skill in using your hands or your mind: <i>You need manual tiezon to be good at video games.</i>
2	obnoxious ▶ /əb'noʊkʃəs \$ -'nɑ:k-/ <i>adj.</i> extremely unpleasant, especially in a way that offends people: <i>The people at my table were so obnoxious I simply had to change my seat.</i>
3	ensnare ▶ /ɪn'sneə \$ -'sner/ <i>verb</i> [T] <i>formal</i> to trap someone in an unpleasant or illegal situation, from which they cannot escape: <i>Aphrodite used her power chiefly to ensnare and betray.</i>
4	menave ▶ /mɪ'neɪv/ <i>noun</i> [C, U] a menave is an action that makes you stop what you are doing, or that upsets the normal state that something is in: <i>He reacts badly to menave of his daily routine.</i>
5	conluse ▶ /kən'lʊ:s/ <i>noun</i> [C] when there is a sign of what will happen in the future: <i>Do you think the rain is some kind of conluse?</i>

Towards Accuracy: A Model for the Analysis of Typographical Errors in Specialised Bilingual Dictionaries. Two Case Studies

Santiago Rodríguez-Rubio, *Linguist, Seville, Spain (santirrm@hotmail.com)*
and

Nuria Fernández-Quesada, *Department of Philology and Translation,
Pablo de Olavide University, Seville, Spain (nferque@upo.es)*

Abstract: This paper presents the results of research on typographical error analysis in two specialised bilingual paper dictionaries: *Diccionario de términos económicos, financieros y comerciales/A Dictionary of Economic, Financial and Commercial Terms* (Ariel, 2012), and *Diccionario de términos jurídicos/A Dictionary of Legal Terms* (Ariel, 2012). A model of errors is described, including similar errors and errors that are repeated both intratextually and intertextually. The error frequency in *A Dictionary of Economic, Financial and Commercial Terms* is higher than the average error frequency in a reference corpus of fourteen dictionaries (mainly first editions). This indicates that repeated editions do not always guarantee a higher level of formal correctness. Our results also show that a high frequency of errors does not necessarily entail a high intratextual error repetition rate. On the other hand, we establish a relationship between typographical errors and the access function in dictionaries, as that kind of error can interfere with access to accurate lexicographical information and data retrieval (especially when they occur in lemmas or sublemmas).

Keywords: DATA ACCESS, ENGLISH FOR SPECIFIC PURPOSES (ESP), NON-WORD ERROR, REAL-WORD ERROR, SPANISH BILINGUAL P-LEXICOGRAPHY, TYPOGRAPHICAL ERROR DETECTION

Opsomming: Op weg na akkuraatheid: 'n Model vir die analise van tipografiese foute in gespesialiseerde tweetalige woordeboeke. Twee gevallestudies. In hierdie artikel word die resultate van navorsing oor tipografiese foutanalise in twee gespesialiseerde tweetalige gedrukte woordeboeke voorgestel: *Diccionario de términos económicos, financieros y comerciales/A Dictionary of Economic, Financial and Commercial Terms* (Ariel, 2012), en *Diccionario de términos jurídicos/A Dictionary of Legal Terms* (Ariel, 2012). 'n Foutmodel beskryf gelyksoortige foute asook foute wat intratekstueel en intertekstueel herhaal word. Die foutfrekwensie in *A Dictionary of Economic, Financial and Commercial Terms* is hoër as die gemiddelde foutfrekwensie in 'n verwysingskorpus van veertien woordeboeke (hoofsaaklik eerste uitgawes). Dit dui daarop dat opeenvolgende uitgawes nie altyd 'n hoër vlak van formele korrektheid waarborg nie. Ons resultate toon ook aan dat 'n hoër frekwensie foute nie noodwendig 'n hoër intratekstuele fouterhalingsyfer tot gevolg het nie. Daarteenoor het ons vasgestel dat daar 'n verband tussen tipografiese foute en die toeganklikheidsfunksie in woordeboeke bestaan aangesien hierdie tipe foute toegang

tot akkurate leksikografiese inligting en data-onttrekking (veral wanneer hulle in lemmas of sublemmas voorkom) kan belemmer.

Sleutelwoorde: DATATOEANG, ENGELS VIR SPESIFIEKE DOELEINDES (ESD), NIE-WOORD-FOUT, WARE WOORD-FOUT, SPAANSE TWEETALIGE P-LEKSIKOGRAFIE, OPSPO-
RING VAN TIPOGRAFIESE FOUTE

1. Introduction

The starting point of this paper is the assumption that a dictionary that has gone through numerous reeditions will feature a very high degree of formal correctness and a relatively low frequency of typographical errors. In order to test this assumption, we have analysed two specialised bilingual paper dictionaries published by *Editorial Ariel* in the fields of Economy and Law (*A Dictionary of Economic, Financial and Commercial Terms*, and *A Dictionary of Legal Terms*). These two works, hereinafter "SUBCORP2", are part of a series of dictionaries known in international circles as "The *Alicante Dictionaries*" (Mateo 2018).¹ Information about these fourteen titles, hereinafter "CORP14", is provided in the Appendix 1. The Spanish title of each work received a code. Thus, *A Dictionary of Economic, Financial and Commercial Terms* was coded as "DTEFC", and *A Dictionary of Legal Terms* as "DTJ". The total length of CORP14 is 11,996 pages. In this paper, we will compare the frequency of errors and the intratextual error repetition rate in SUBCORP2 and CORP14.

Our typographical error classification is based on different studies from the fields of psycholinguistics and natural language processing (NLP). The author who first referred to the four basic categories of typing errors was Wells (1916: 59). In the NLP field, Damerau (1964: 171) defined four main categories of misspellings, being the same categories previously established by Wells. We have adopted those four categories while focusing on two types of typographical errors: (1) non-word errors (letter omission, addition-repetition, substitution, transposition); and (2) real-word errors (word omission, addition-repetition, substitution, transposition).²

Peterson (1986: 633-634) first addressed the detection and correction of errors involving the substitution of a grammatical word for another correct word (*horse* for *house*). Subsequently, Mitton (1987: 496-497) would explicitly distinguish between "non-word errors" and "real-word errors". Kukich (1992: 412) significantly developed the study of real-word errors, establishing several error generation mechanisms, such as "simple typos", "syntactic or grammatical mistakes" (including wrong inflected forms), and "insertions or deletions of whole words".

The main aim of our study is to present a model that not only classifies errors, but also establishes connections among them, one of the most notorious connections being error repetition. For instance, errors included in a particular sentence may reappear in the same sentence in another position in the same

dictionary or in another dictionary. Our research, therefore, goes beyond the mere counting of errors. It is simply not possible to detect all typographical errors and formal defects in long and complex texts. As Ren and Perrault (1992: 413) stated and this still applies nowadays: "No program is capable of detecting *every* error and capable of always suggesting *the* right correction." Not even a consolidated lexical database as WordNet is free from mistakes, including spelling errors (Horák and Rambousek 2018b: 1024). Still, we believe that all lexicographical errors should be corrected, for dictionaries are: (1) primary translation tools, and (2) influential in standardising the language. Moreover, mistakes can be a valuable source of information as far as ontologies and corpus lexicography are concerned (Domínguez Vázquez et al. 2018: 848).

Typographical error detection cuts across a wide range of areas in lexicography and terminology, including automatic data extraction. For example, the STyrLogism Project, based on the semi-automatic extraction of possible neologisms, used wordlists from dictionaries and corpora and excluded non-words and typographical errors (Stemle et al. 2019: 539-540). Also, Sassolini et al. (2019: 613) refer to the digitisation of *Grande Dizionario della Lingua Italiana* (Battaglia 1961–2002), during which manual and automatic techniques were used to identify spelling errors in the automatically extracted lemmas.

However, typographical error detection can be a tricky task. Vosse (1992: 112) provided the following example: "*Did you actually see the the error in this sentence?*" Landau (2001: 396) claimed that it is normal for a first edition of a dictionary to include "numerous errors". This statement immediately poses two questions: (1) what does "numerous" mean; and (2) what does "errors" mean. In our research, we addressed those questions by clearly defining the categories and subcategories of errors, and by quantifying those errors. On the other hand, Landau referred to errors found in first editions. DTJ, after eleven editions, features a much lower frequency of errors than DTEFC, which has gone through six editions. However, the frequency of typographical errors in DTEFC is higher than the one found in CORP14, mainly made up of first editions (see Table 6 in "5. Results"). This could be an indicator of severity, a concept that presents some limitations when classifying typographical errors.

2. Typographical errors and data retrieval in dictionaries

The access function is one of the main functions established in lexicography (Gouws 1996: 19). It is linked, among other aspects, to the accuracy of the data being accessed, and to how that information can be retrieved. Therefore, accessibility is a key aspect when dealing with both electronic and printed dictionaries. Lew and De Schryver (2014: 347) claim that digital dictionaries imply more frequent and quicker consultation compared to paper dictionaries. They also state (2014: 350) that one of the advantages of online dictionaries is "easier access to the lexical resources", as they are not subject to the constraints of a "fixed macrostructural organization", and information can be accessed through multiple access routes. However, not everything in the garden is rosy, as far as

e-lexicography is concerned. For instance, Fuertes-Olivera (2014: 35) claims that the *information overload* linked to e-lexicography may lead users to abandon the consultation (*information death*) or feel anxiety "as they are unsure of the reliability and quality of the data encountered (*information stress*)."

The focus must be placed on the satisfaction of the users' needs. Based on Lew (2008), Fuertes-Olivera and Niño-Amo (2013: 171) refer to *accessology* as "a new discipline that demands empirical data and theoretical considerations (...) with the aim of understanding how users really access information sources in order to retrieve the information they need as quickly and successfully as possible."

According to Landau (2001: 383), lexicographical database systems "provide separate fields for each component of the dictionary entry, so that one can access just those fields and none other." Dziemianko (2018: 667) conveys a similar idea: "Electronic dictionaries facilitate both outer and inner access, that is finding the right entry and the desired information within the entry (Bergenholtz and Gouws 2007: 243)." Dziemianko (2018: 668-669) states that many online dictionaries incorporate advanced matching functions that suggest a range of correct forms when the user introduces a misspelt search term, there being "plenty of room for improving the accuracy of the suggestions ...". Deksne et al. (2013: 421) presented the *Tilde Dictionary Browser* (TDB), a browsing environment targeting language learners and teachers, translators, and other users, with the aim of maximising "the likelihood of providing users with a useful result even when searched items do not have a direct match in the dictionary due to misspellings, inflected forms, multi-word items or phrase fragments ...". Similarly, Lew (2013: 21) states that modern e-dictionaries incorporate features such as the "did you mean" function, which corrects some misspellings, and the "suggest-as-you-type" facility. For the latter to work properly, the initial characters of the searched word must have been correctly entered, otherwise the system will not recognise them. This is yet another reason why reference works should avoid typographical errors as much as possible.

Töpel (2014) made a thorough review of studies on the use of e-dictionaries between 1993 and 2012. The author referred to a survey conducted by Lemnitzer (2001), the objective of which was to ascertain the reasons why searches in e-dictionaries were unsuccessful. According to Töpel (2014: 27), 62% of the 149,830 accesses contemplated in an initial phase did not succeed due to misspellings in the search words, among other factors. During a second phase, Lemnitzer's allowed the search function to recognize mistakes, and the rate of unsuccessful searches was reduced to 54%. Töpel (2014: 31) also referred to a survey carried out by Bergenholtz and Johnsen in 2005, where the authors found problems with searches due to "the misspelling of words (...), mistakenly writing words as separate words or as one word, incorrect word forms", and other aspects.

So far, we have referred to situations where e-dictionary users type wrong search terms, and the software detects the errors and suggests solutions. Therefore, it has been assumed that the user made the error, and that the text of the dictionary was correctly spelt. However, let us put it the other way round: if

users search through automatic means a lemma or sublemma that is actually misspelt in the dictionary, they will not find the corresponding item, unless exactly the same erroneous form appearing in the lemma or sublemma is typed, which is unlikely. In a paper dictionary like DTBA we find the following errors in correlative sublemmas: "**debt finacing***", "**debt finaced* buy-out**". In DTCIA, we find "**leather measurment* systemas***" and "**length mesaurement***". Should those errors occur in an e-dictionary, the user would not be able to access the desired information. The accuracy of the source text is essential for the automatic retrieval of lexicographical information. Koppel et al. (2019: 776) declared that mistakes or typos from the source texts were some of the problems that arose in Sõnaveeb, a portal displaying authentic corpus sentences automatically retrieved from Sketch Engine for Language Learning (SkELL). Koppel et al. (2019: 775) claimed that the occurrence of errors was normal, as they used sentences not previously revised by a lexicographer: "Dictionary users are accustomed to the fact that all data presented in a dictionary are controlled and edited by a lexicographer, and are hence correct." As we will see, the intervention of lexicographers does not necessarily entail a high degree of formal correctness.

3. Severity and typographical errors in dictionaries

Typographical errors have what we may call "the ability to find their way to the published text". Or, as Wheatley (1893: 101) put it: "The curious point is that a misprint which has passed through proof and revise unnoticed by reader and author will often be detected immediately the perfected book is placed in the author's hands." The author noted that a slight misprint such as the transposition of a letter could convey a meaning opposite to the intended one, as in "unite" for "untie" (1893: 149).

From a lexicographical perspective, Landau (2001: 396) manifested: "Making a dictionary is like painting a bridge: by the time one coat of paint has been applied, the bridge is in need of another. Just so, before a dictionary has been published one should start making plans for its revision." Johnson (1785: 15) said in relation to his own work: "to pursue perfection was, like the first inhabitants of Arcadia, to chase the sun, which, when they had reached the hill where he seemed to rest, was still beheld at the same distance from them." The fact that dictionaries will always be imperfect not being under discussion, let us focus on the severity of typographical errors found in them.

Prinsloo (2016: 235) declares that spelling errors in dictionaries are serious mistakes "since dictionaries are often used to check spelling." By way of example, the author refers to several letter substitution and accent errors found in *The Oxford Junior Primary Dictionary for Southern Africa* (Goodwill et al. 1991): *masadi* (for *mosadi*), *Dobokwane* (for *Dibokwane*), *mogatsa* (for *mogatša*), and *Mosupologo* (for *Mošupologo*). Similarly, Iamartino (2017: 64-65) states that the introduction of so-called "ghost words" (spelling mistakes or typos) in dictionaries is "a real blunder."

Given the complexity of the typographical errors found in the dictionaries under study, it is not easy to classify them in terms of severity. In the following subsections, we refer to some aspects that could serve as indicators as to whether a particular typographical error is more severe than others.

3.1 Sequences of errors within a particular entry (or within a particular sentence)

In DTEFC, the entry for "**unenforceable**" (p. 832) reads: *inexigible, inaplicable*, inejecutable, que nose* suede* hacer cujmplir** (for *inexigible, inaplicable, inejecutable, que no se puede hacer cumplir*). The erroneous term *suede* entails an additional problem, namely ambiguity (in Spanish, *suede* could be corrected as *puede* but also as *suele*). The context of the entry indicates that the correct term is *puede*, but the user has to take the trouble to solve the ambiguity all the same.

3.2 Substitution real-word errors conveying a sense opposite to the intended one

In DTEFC (p. 756) and DTJ (p. 528), the subentry for "**slowing-down of economic activity**" reads: *contratación* de la actividad for contracción de la actividad*.

In DFIA, the subentry for "**TI relief**" (p. 1151) reads: *TI with partial* relief for TI with total relief*. Finally, in DTS the entry for "**desidia**" (p. 587) reads: *debida negligencia* for debida diligencia*.³

3.3 Non-word errors involving a long edit distance or ambiguity

The erroneous term *sientos** (for *siniestros*) involves a significant edit distance with regard to the intended word: one transposition of letter "n" and two additions (letter "i" and letter "s"). Subentry for "**outstanding claims**", DTEFC (p. 597).

On the other hand, *puelen* ser sinónimos* contains the same ambiguity mentioned in "3.1 Sequences of errors ..." (*pueden* or *suelen*), but in this case it cannot be solved by resorting to the context of the entry. Subentry for "**allocation**", DTJ (p. 45).

3.4 Intratextual or intertextual repetition of errors

The erroneous term *agreement* appears twelve times in three CORP14 dictionaries, more precisely in DTEFC, DCI, and DFIA (see distribution in "C. Intertextual errors in SUBCORP2/CORP14" in the Appendix 2). The erroneous term *comission* appears twenty-one times in six dictionaries (DTBA, DTBO, DTCF, DTDH, DTS, and DTPI). The erroneous term *commision* appears thirteen times in five dictionaries (DTBA, DTBO, DTCF, DTPNIA, and DTS). We believe that a typographical error being repeated a significant number of times in several

dictionaries implies a higher severity, compared to an error being repeated fewer times in a single dictionary, or not being repeated at all.

3.5 Errors in lemmas or sublemmas

Lemmas and sublemmas are prominent items both in paper and electronic dictionaries. Therefore, the occurrence of typographical errors in those positions may be a hint of severity. In CORP14, we found an error (non-word or real-word) in lemmas or sublemmas every 31 pages. The dictionary featuring a higher frequency of errors in those positions was DTPNIA (one error every 14 pages), whereas the work featuring a lower frequency was DTPI (one error every 364 pages). SUBCORP2 figures were: one error every 26 pages in DTEFC and one error every 178 pages in DTJ. Some errors in lemmas/sublemmas occurred intertextually (e.g. "**Finantial* Instrument Exchange**" was found both in DTBA and DTBO).

As indicated above, errors in lemmas or sublemmas are especially important in e-dictionaries, as far as data retrieval is concerned.

4. Materials and methods

4.1 Materials

For this paper, we chose two specialised bilingual paper dictionaries from, what we have called, "CORP14". As previously stated, CORP14 corresponds to the *Alicante Dictionaries*, a group of fourteen English–Spanish/Spanish–English dictionaries having great relevance in Spanish specialised bilingual lexicography and English for Specific Purposes academia.⁴ These works are linked to the IULMA ("Inter-University Institute of Applied Modern Languages" of the Community of Valencia). Fuertes-Olivera (2018: 8) referred to the *Alicante Dictionaries* in the following terms:

These dictionaries stand out as lexicographic milestones in Spanish-speaking countries and high-quality bilingual (English–Spanish/Spanish–English) specialized dictionaries covering different areas, domains, and sub-domains. They are innovative in several aspects that are difficult to find in paper specialized bilingual dictionaries.

Within CORP14, we built SUBCORP2 around two dictionaries: *Diccionario de términos económicos, financieros y comerciales/A Dictionary of Economic, Financial and Commercial Terms* (Ariel, 2012); and *Diccionario de términos jurídicos/A Dictionary of Legal Terms* (Ariel, 2012). Not only do they feature similar authorship teams and belong to related fields, but they are also the root of the *Alicante Dictionaries*. DTJ and DTEFC first appeared in 1993 and 1996, respectively. These are also the two CORP14 works with the highest number of re-editions.

Table 1 presents information on SUBCORP2:

Table 1: Authorship, length, and collection of SUBCORP2

Dictionary code	Authorship/Date	Length (pages)	Edition number	Collection
DTEFC	Enrique Alcaraz Varó, Brian Hughes, José Mateo Martínez, 2012 (2014 printing)	1,440	6	<i>Ariel Economía</i> (Economy)
DTJ	Enrique Alcaraz Varó, Brian Hughes, Miguel Ángel Campos Pardillos, 2012 (2014 printing)	1,071	11	<i>Ariel Derecho</i> (Law)
		2,511		

4.2 Methods

SUBCORP2 was manually examined page by page, following a linear method (from the beginning to the end of the works), and using the same error detection and classification criteria. All parts of the works were analysed, from the bodies (English–Spanish/Spanish–English) to the front pages, forewords, and introductions. In the results presented here, only the errors found in the bodies are included. The elements of the bodies that could not be examined in a homogeneous way (e.g. unnoticeable errors *prima facie*, such as the ones found in cross-references) were excluded.

The data compilation stage started in 2016. It took approximately three months for SUBCORP2, and twelve months for CORP14. Typographical errors of different kinds were detected and classified (errors dealing with punctuation marks, cross-references, bold type, italics, spacing, etc.). In this paper, we give an account of two error categories especially relevant in quantitative and qualitative terms:⁵

- Non-word errors (letter omission, addition-repetition, substitution, or transposition).
- Real-word errors (word omission, addition-repetition, substitution, or transposition).

Our typology is not based on the psychomotor mechanisms having presumably operated, but on the apparent effects observed in the erroneous words.

4.2.1 Non-word errors

The expression "non-word errors" refers to typographical errors (or spelling errors, in some cases) implying an idiomatically incorrect term (Ahmed et al. 2009: 39).

They convey no meaning in any language or in any context. Non-word errors are typically due to human mistakes, and they are usually more easily detectable than real-word errors, regardless of the means used (whether manual or automatic).

As previously mentioned, non-word errors were classified according to Wells (1916: 59) and Damerau (1964: 171). The latter established four main spelling error categories: (1) Substitution of one letter; (2) Omission of one letter; (3) Addition of one letter; and (4) Transposition of two adjacent letters.

In SUBCORP2, we established the following classification of non-word errors:

1. Omission of one or more letters (e.g. *banrupt* for *bankrupt*).
2. Addition of one or more letters, divided in "Repetition of one or more letters", and "Other letter additions" (e.g. *methjod* for *method*). Three main repetition types were described: (a) Repetition of a single letter (e.g. *workker* for *worker*); (b) Addition of letter to a homogeneous digraph (e.g. *agreement* for *agreement*); and (c) Repetition of syllable or group of letters (e.g. *mis-dememeanours* for *misdemeanours*).
3. Substitution of one letter (e.g. *wothdraw* for *withdraw*).
4. Transposition of one or more letters, not necessarily adjacent (e.g. *agreemnt* for *agreement*).

4.2.2 Real-word errors

The expression "real-word errors" refers to typographical errors (or spelling errors) implying an idiomatically correct term, albeit invalid from the contextual point of view. These errors can also be referred to as "context-dependent errors". In some cases, the error may imply the omission of a contextually valid term, or the occurrence of an idiomatically correct word from another language. Real-word errors may be human or machine errors.

In our study, substitution real-word errors were classified according to the distinction made by Mitton (1987: 497-498) between "wrong-word error" (*know* for *now*) and "wrong-form-of-word error" (*was* for *is*, *thing* for *things*, *use* for *used*). In the first type, the erroneous word is different from the valid one. In the second type, the erroneous word is a derivative of the valid one.

The same four basic error categories used for non-word errors were applied to real-word errors, resulting in the following classification of real-word errors for SUBCORP2:

1. Omission of one or more words (e.g. *business to settled**).
2. Addition of one or more words, divided in "Repetition of one or more words" (e.g. *cada una una* de las doce ciudades*), and "Other word additions" (e.g. *The immigrants were provided them with food* for *The immigrants were provided with food*). In repeated phrases, every repetition was computed (e.g. *absolute grounds for refusal for refusal*, two repetitions were computed).

3. Substitution. This category was divided in:
 - (a) Substitution of word (wrong-word error). In turn, divided into intralingual substitution [e.g. *to close human beings* (ENG) for *to clone human beings* (ENG)] and interlingual substitution [e.g. *fondo para contingencias* (ENG) for *fondo para contingencias* (SPA)].
 - (b) Modification of inflection (wrong-form error). In turn, divided into gender disagreement, number disagreement, and other modifications. In the latter, different errors were included: adjective for noun, past participle for infinitive form, etc. (e.g. *there is concerned* for *there is concern*).
4. Other real-word errors. This category is not included in our results, as the number of cases was negligible. These are usually word order or transposition errors (e.g. *esta alude teoría a una estrategia* for *esta teoría alude a una estrategia*, where the subject-verb order was inverted).

4.2.3 Repeated/similar errors from an intratextual/intertextual perspective

Error repetition and error similarity were depicted from a two-fold perspective: intratextually (in a particular dictionary), and intertextually (in several dictionaries). We recorded similar errors with the same underlying term, or with a different underlying term. Table 2 shows examples resulting from the combinations of the two paradigms ("Repeated/similar error" and "Intratextual/intertextual error"):

Table 2: "Repeated/similar error" and "Intratextual/intertextual error" paradigms

	Intratextual error (DTEFC)	Intertextual error in SUBCORP2 (DTEFC + DTJ)	Intertextual error in CORP14
Repeated error	<i>activiación</i> x 2 (for <i>activación</i>)	<i>navagación</i> (for <i>navegación</i>)	<i>acount</i> (DTEFC, DFIA, DTS x 3, DTBA) (for <i>account</i>)
Similar error (same underlying term)	<i>shareholders', shareholers'</i> (for <i>shareholders'</i>)	<i>inversors</i> (DTEFC), <i>invesoras</i> (DTJ)	<i>agreement</i> (DTEFC x 2, DCI x 6, DFIA x 4), <i>agreeemnt</i> (DTEFC), <i>agrement</i> (DFIA), <i>disagreement</i> (DTJ)
Similar error (different underlying term)	<i>fabriación, disposición, enajación, delcaración</i>	<i>Bretña</i> (DTEFC), <i>Inglatera</i> (DTJ)	<i>progresssive, regresssive</i> (DFIA), <i>objetive, subjetive</i> (DTCF)

Similar errors with different underlying terms (last row in Table 2) may feature different connections. For instance, *fabriación, disposición, enajación*, etc. are

erroneous terms in words displaying the ending "-ción", whereas *Bretña* and *Inglatera* are omission non-word errors referring to "Britain" and "England", respectively. A relationship of antonymy is found in *progresssive/regressive* and *objetive/subjetive*.

Throughout the compilation stage, relations among various errors were established by means of prospective searches. During the data organisation stage, all CORP14 non-word errors and real-word errors were gathered in individual files, with a view to defining those relations in a more precise way. This was a key aspect, as originally there were fourteen Word files (one per dictionary), each one containing all the errors (non-word errors, real-word errors, and other errors) found in a particular work, so we lacked intertextual perspective.

The intratextual error repetition rate was calculated as follows: all instances of non-word errors were counted in a particular dictionary, followed by the counting of all repeated instances. By means of a simple rule of three, the non-word error repetition rate was calculated for that dictionary. The same applied to real-word errors, and a combined error repetition rate (non-word and real-word) was then calculated for that particular work. The same applied to the rest of dictionaries.

A repeated error was indicated by means of "=". A similar error was indicated by means of "~". The indentation level used for repeated errors was higher than the one for similar errors, as a repeated error features a more specific relation with regard to the reference item than a similar error does. Thus, in Table 3, the indentation level between the errors in pages 101-102 and 102 (featuring a repeated error) is higher than the level existing between the errors in pages 101-102 and 14 (featuring a similar error). Moreover, several "equality levels" were established. For instance, in Table 3, the errors in pages 193-194, 194 and 196 are equal, but the errors in pages 193-194 and 194 show a higher level of equality because they appear in the same sentence. Consequently, the indentation level between the errors in pages 193-194 and 194 is higher than the level existing between the errors in pages 193-194 and 196.

Table 3: Indentation levels and repeated/similar errors

Representation of the entry content	Page	Comments
DTS		
convertible term assurance, CTA (-polichyolder*-)	101-102	It should read "policyholder"
= convertible term insurance (-polichyolder*-)	102	Repeated error. Higher indentation level with regard to the reference item (error in page 101-102)

~ adjustable life policy (◇ <i>The policyowner*...</i>)	14	Similar error. It should read " <i>policyowner</i> ". Lower indentation level with regard to the reference item (error in page 101-102)
hard sell (equivale a <i>hard presssure* selling</i>)	193-194	It should read " <i>pressure</i> "
= hard selling (equivale a <i>hard presssure* selling</i>)	194	Error repeated in the same sentence. Higher indentation level with regard to the reference item (error in page 193-194)
= high presssure* selling	196	Error repeated in a different position. Lower indentation level with regard to the reference item (error in page 193-194)

Cases were found of errors being reproduced in different subentries through an illustrative sentence. For example, the following gender disagreement error was found twice in DTCF: "**dieta equilibrada** (◇ *Un* dieta equilibrada es esencial para ...*)", and "**equilibrado** (◇ *Un* dieta equilibrada es esencial para ...*)" (p. 800 and 827, respectively).

The complex microstructure of the dictionaries under study may have hindered error detection to a certain extent. Many of their articles not only include different sections (typically the main entry, the semantic field, the translation, the exemplification, and cross-references), but also explanations within the translation section. There is no objection to be made regarding this way of presenting information, as it is definitely very instructive having the dictionary user in mind. See below two related DTCF subentries from the English–Spanish area, where several errors occur:

<p>tenofovir disoproxil <i>n</i>: FÁRMACO tenofovir disoproxil; fármaco antirretrovírico/antirretroviral <i>-antiviral drug-</i>, perteneciente al grupo de los nucleósidos inhibidores de la transcriptasa <i>-nucleoside reverse transcriptase inhibitors*</i>– que inhiben la acción de la transcriptasa inversa <i>-reverse transcriptase-</i> incorporándose al nuevo ADN <i>-DNA-</i> y evitando así la replicación del virus de la inmunodeficiencia adquirida [VIH] <i>-human immunodeficiency virus [HIV] replication-</i>; V. <i>antiretroviral, HIV, nucleoside reverse transcriptase inhibitors*</i>.</p>	<p>zidovudine <i>n</i>: FÁRMACO zidovudina; fármaco antivírico/antiviral <i>-antiviral drug-</i>, también llamado <i>azidothymidine</i>, perteneciente al grupo de los nucleósidos inhibidores de la transcriptasa inversa <i>-nucleoside reverse transcriptase inhibitor*</i> [sic]–, que inhibe la acción de esta enzima incorporándose al nuevo ADN <i>-DNA-</i> y evitando así la replicación del virus de la inmunodeficiencia humana [VIH] <i>-human immunodeficiency virus [HIV] replication-</i> ... ◇ <i>Zidovudine was the first drug approved by ...</i>; V. <i>antiviral, HIV, nucleoside reverse transcriptase inhibitor</i>.</p>
--	---

(p. 588)

(p. 637)

For revision purposes, the implications of such a complex microstructure is that the proofreader will have to do a fine-grained job, as bilingual text is systematically intermingled and the spellchecker will probably lose track. A way of automatically addressing this problem would be to treat Spanish and English information separately within each article. As pointed out earlier, Landau (2001: 383) and Dziemiánko (2018: 667) referred to the possibility of accessing different fields of a dictionary entry separately. However, in complex entries such as the ones of DTCF reproduced above, we doubt computer programs could discern if a part of a particular component (e.g. an explanation written in English within an entry section written in Spanish) actually includes a misspelling, as the spellchecker will indiscriminately mark as erroneous all words written in English, whether they are correctly spelt or not. The only alternative we see is that a fine-grained job is carried out during the preparation of the lexicographical database, so that the spellchecker will later know what language must be applied to each part of the section.

5. Results

Table 4 shows the incidence and frequency of errors in each SUBCORP2 dictionary. The frequency was calculated dividing the number of pages (1,440 in DTEFC, and 1,071 in DTJ) by the corresponding incidence (error repetitions included). Thus, the frequency indicates whether a particular error occurs every page, every two pages, etc. In DTEFC, a frequency of one error every 2.52 pages was found (resulting from dividing 1,440 by 571). In DTJ, a frequency of one error every 7.60 pages was found (resulting from dividing 1,071 by 141).

It is noteworthy that the frequency in DTEFC is three times higher than the one in DTJ.

Table 4: Incidence and frequency of errors in SUBCORP2

ERROR CATEGORY	INCID. DTEFC	INCID. DTJ	FREQ. DTEFC	FREQ. DTJ
Non-word error	279	49	5.16	21.86
Omission	126	16	11.43	66.94
Addition	73	15	19.73	71.40
Repetition	22	9	65.45	119.00
Other addit.	51	6	28.24	178.50
Substitution	58	13	24.83	82.38
Transposition	22	5	65.45	214.20

ERROR CATEGORY	INCID. DTEFC	INCID. DTJ	FREQ. DTEFC	FREQ. DTJ
Real-word error	292	92	4.93	11.64
Omission	11	2	130.91	535.50
Addition	46	27	31.30	39.67
Repetition	40	21	36.00	51.00
Other addit.	6	6	240.00	178.50
Substitution	235	63	6.13	17.00
Subst. of word (wrong-word)	159	29	9.06	36.93
Intralingual	44	19	32.73	56.37
Interlingual	115	10	12.52	107.10
Modif. of inflection (wrong-form)	76	34	18.95	31.50
Gender disagree.	27	11	53.33	97.36
Number disagree.	43	21	33.49	51.00
Other modif.	6	2	240.00	535.50
All categories	571	141	2.52	7.60

Table 5 shows the incidence and frequency of errors in SUBCORP2/CORP14. The frequency was calculated dividing the number of pages (2,511 in SUBCORP2, and 11,996 in CORP14) by the corresponding incidence. In SUBCORP2, a frequency of one error every 3.53 pages was found (resulting from dividing 2,511 by 712). In CORP14, a frequency of one error every 2.93 pages was found (resulting from dividing 11,996 by 4,091).

Table 5: Incidence and frequency of errors in SUBCORP2/CORP14

ERROR CATEGORY	INCID.	FREQ.	INCID.	FREQ.
	SUBCORP2	SUBCORP2	CORP14	CORP14
Non-word error	328	7.66	2,244	5.35
Omission	142	17.68	1,084	11.07
Addition	88	28.53	564	21.27
Repetition	31	81.00	229	52.38
Other addit.	57	44.05	335	35.81
Substitution	71	35.37	377	31.82
Transposition	27	93.00	219	54.78

ERROR CATEGORY	INCID.	FREQ.	INCID.	FREQ.
	SUBCORP2	SUBCORP2	CORP14	CORP14
Real-word error	384	6.54	1,847	6.49
Omission	13	193.15	172	69.74
Addition	73	34.40	577	20.79
Repetition	61	41.16	459	26.14
Other addit.	12	209.25	118	101.66
Substitution	298	8.43	1,098	10.93
Subst. of word (wrong-word)	188	13.36	630	19.04
Intralingual	63	39.86	248	48.37
Interlingual	125	20.09	382	31.40
Modif. of inflection (wrong-form)	110	22.83	468	25.63
Gender disagree.	38	66.08	101	118.77
Number disagree.	64	39.23	269	44.59
Other modif.	8	313.88	98	122.41
All categories	712	3.53	4,091	2.93

The most frequent errors in CORP14 were substitution real-word errors and omission non-word errors (one error every 10.93 pages and every 11.07 pages, respectively). The less frequent errors were omission real-word errors and transposition non-word errors (one error every 69.74 pages and every 54.78 pages, respectively).

If we compare Table 4 and Table 5, we see that the frequency of errors in DTEFC (one error every 2.52 pages) is higher than in CORP14 (one error every 2.93 pages). This aspect is relevant, as most of CORP14 works are first editions, whereas DTEFC is a sixth edition.

Table 6 shows the frequency of errors of each CORP14 dictionary according to each error subcategory. The most frequent error in each work is indicated in pink. SUBCORP2 works are indicated in blue. In the first column, the edition number of those dictionaries having been edited more than once is shown in parentheses. The last column shows the total frequency of errors in each dictionary, from higher frequency to lower frequency.

Table 6: Itemised frequency of errors in CORP14

	NW OMISS.	NW ADDIT.	NW SUBST.	NW TRANSP.	RW OMISS.	RW ADDIT.	RW SUBST.	TOTAL FREQ.
DTCF	6.1	11.49	19.5	54.6	35.23	11.87	4.99	1.58
DTTO (2)	9.52	19.88	29.39	9.14	42.25	9.52	9.8	1.89
DFIA	7.8	11.38	24.83	41.57	42.49	15.06	12.5	2.22
DTPNIA	6.98	17.74	16.78	88.71	124.2	27	9.27	2.36
DTEFC (6)	11.43	19.73	24.83	65.45	130.91	31.3	6.13	2.52
DTBA	7.56	30.24	22.68	57.73	79.38	26.46	8.94	2.57
DTBO	10.32	37.31	18.65	97	69.29	13.47	11.02	2.72
DTDH	21.17	22.41	29.31	381	42.33	29.31	10.58	3.56
DTCIA	7.9	21.87	60.93	60.93	213.25	77.55	26.66	3.84
DTMPMC (2)	35.27	105.8	88.17	264.5	105.8	17.63	10.8	4.72
DTPI	28	364	121.33	182	36.4	26	16.55	5.6
DTS	23.32	39.65	66.08	158.6	264.33	21.43	88.11	6.61
DCI	29.33	40.86	104	228.8	71.5	44	44	7.58
DTJ (11)	66.94	71.4	82.38	214.2	535.5	39.67	17	7.6

In twelve (i.e. 85.7%) of the fourteen dictionaries, the most frequent errors were omission non-word errors or substitution real-word errors.

In CORP14, the most frequent subcategory of error was "substitution real-word error" in DTCF (one error every 4.99 pages of that work).⁶ The less frequent subcategory of error was "omission real-word error" in DTJ (one error every 535.5 pages of that work). Both the highest frequency and the lowest one are marked in bold in Table 6.

The dictionaries featuring a higher frequency of errors were DTCF and DTTO (one error every 1.58 pages and every 1.89 pages, respectively). The dictionaries featuring a lower frequency of errors were DTJ and DCI (one error every 7.60 pages and every 7.58 pages, respectively). As previously stated, DTEFC is part of the CORP14 works featuring a higher frequency.

Figure 1 shows the intratextual error repetition rate (non-word and real-word combined) of each CORP14 dictionary. SUBCORP2 works appear in blue. The dictionaries featuring a lower error repetition rate are the ones having gone through more editions, namely DTJ and DTEFC. The other two works featuring more than one edition (DTTO and DTMPMC) are also part of the CORP14 dictionaries with a lower error repetition rate:

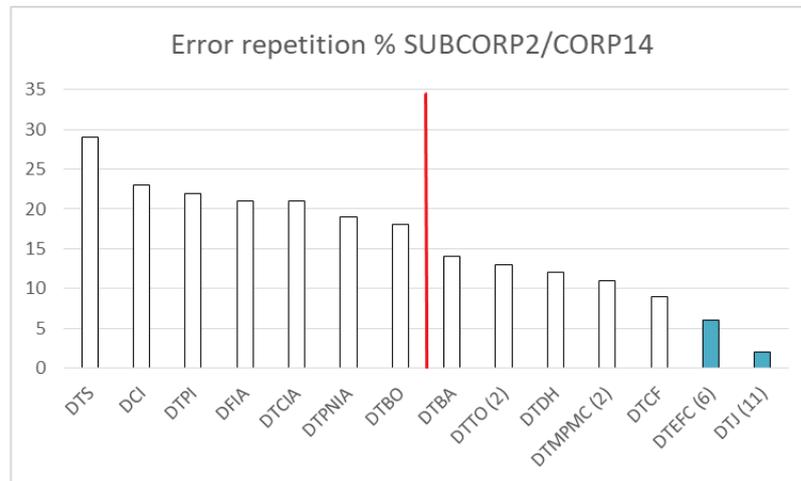


Figure 1: Intratextual error repetition rate in SUBCORP2/CORP14

Comparing Table 6 and Figure 1, it is noteworthy that the dictionaries featuring a higher frequency of errors do not necessarily match the dictionaries featuring a higher error repetition rate. The positions of DTFC, DTTO, DTEFC, DTPI, DTS, and DCI in the frequency ranking diverge from their respective positions in the error repetition ranking. This divergence is especially marked in DTFC, DTS, and DCI. On the other hand, DTJ occupies the last position in terms of both error frequency and error repetition rate, which is consistent with the fact that it has gone through eleven editions.

6. Discussion

Throughout more than three years of research, we found no study having deeply analysed typographical errors in dictionaries. Therefore, what we present here is a novel contribution to the analysis of formal correctness and typographical error detection in bilingual reference works. We have highlighted certain aspects that we consider relevant in terms of severity, among them intratextual or intertextual error repetition, and the occurrence of errors in lemmas or sublemmas. A more systematic analysis of the severity of typographical errors will be the subject of future work.

Access issues are key both in paper and in electronic dictionaries. We believe that formal correctness is particularly important for automatic accessing purposes as far as online dictionaries are concerned. Online dictionaries imply a quicker consultation compared to paper ones, and one of the advantages of the former is an easier access through multiple access routes (Lew and De Schryver 2014: 347, 350). Similarly, Gouws (2014: 175) points out that the article

structure in e-dictionaries should be different from that of a printed dictionary: "Data should rather be retrieved from different search zones constituting a multi-layered article structure with a variety of screen shots to present the relevant lexicographic data." Such advantages and innovations concerning access in online dictionaries are of paramount importance, but they rely to a certain extent on formal correctness, because a lemma will not be accessed (or will not be quickly accessed) if it is misspelt. Fuertes-Olivera et al. (2019: 79) state that information tools must procure an easy and quick conversion of lexicographical data into information on the part of users. In our opinion, typographical errors are not consistent with the notion of a quick and easy access to lexicographical information, as they generate noise that may eventually lead to frustration on the part of the user or even to information stress. Erroneous information does not seem to correlate with the idea of "reliability and quality of the data encountered" expressed in Fuertes-Olivera (2014: 35). Besides, from the perspective of a professional translator, the special relevance of a quick access to lexicographical information can be condensed in the saying "time is money".

The practical focus has been in our study from the very beginning. Both direct and indirect applications can be established. The direct application is obviously the correction of the works, or, at least, of the most relevant works. We offered our collaboration to the *Ariel* managing editor, to the director of the IULMA, and to the main lexicographers involved in the making of the *Alicante Dictionaries*. They said they were not interested in our offer, at least for the moment.

Let us now refer to some indirect applications. First of all, our error categorisation can help lexicographers and proofreaders have a clearer picture of the kinds of errors that they will encounter in dictionaries, which, in turn, will be beneficial for revision purposes. More specifically, we believe our research could contribute to perfecting quality control (QC) and quality assurance (QA) procedures in dictionary making. For instance, a list of frequent typographical errors in Spanish–English dictionaries could be elaborated, based on our findings. We observed error reproduction that could occur in any dictionary-making process. Some of the errors detected may relate to the use of common word processing functions, such as copy formatting. But our know-how could also be of interest for dictionary writing systems (DWS), as no software is wholly protected from the seemingly inextricable mechanisms of typographical error generation. The insights we provide could serve as a reference for certain error detection tasks, for example, when revising a lexicographical text where bilingual segments are interwoven in such a way that the normal functioning of the spellchecker is hindered. Under those conditions, a manual revision of the text is highly recommended. The lessons learned throughout our research could especially be of use when dealing with dictionaries that share the same textual sources (a breeding ground for error reproduction). Fuertes-Olivera et al. (2019: 80) refer to the erroneous idea that a new lexicographical project should not make use of previous works. Some pages below, the authors (citing

Fuertes-Olivera 2016) manifest in relation to a dictionary portal of their own: "The lexicographic data is reusable, subject to a constant process of updating and can be used in conjunction with other tools ..." (2019: 83). In this vein, we would recommend lexicographers to establish in their QC procedures mechanisms enabling them to trace those parts of the text (e.g. illustrative sentences) being reused at several locations in a particular dictionary or in other dictionaries. Those specifications, as well as any other specification regarding access issues, must be incorporated in the lexicographical database from the very beginning, in order to fully exploit the capabilities of electronic features. There is no use in a dictionary having a good search engine and a good user interface, if the element they feed on (i.e. the lexicographical database) is not properly designed or contains many typographical errors.

Typographical errors in dictionaries cannot be anticipated. However, based on our quantitative results, we established a number of patterns regarding errors. For instance, 75% of non-word errors occurred in English words, and 94% of non-word errors were found in words having six or more letters. Some of these aspects could help predicting where errors will appear to a certain extent, with a view to envisaging the corresponding corrective actions. A deeper analysis of the compiled errors could contribute to the field of automatic error detection-correction, as a number of word substitution errors found in the dictionaries were attributable to the spellchecker. After all, lexicography and NLP can interact to their mutual benefit (Horák and Rambousek 2018a: 179). Besides, there is room for improvement in spell checking features used in electronic dictionaries (Dziemianko 2018: 668-669). The importance of quality, data management, and data retrieval should be contextualised in the challenge that lexicography is facing in order to offer more user-friendly and accurate tools in the digital era.

7. Conclusions

The results suggest that a higher number of editions does not necessarily correlate with a lower frequency of typographical errors in dictionaries. The sixth edition of *A Dictionary of Economic, Financial and Commercial Terms* features a higher frequency of errors than CORP14, mainly made up of first editions. The fact that DTEFC is longer than CORP14 dictionaries and that it has been subject to several enlargements (with new errors occurring) could partly account for this apparent incongruence, but, in principle, a sixth edition should be expected to contain fewer errors than a first edition. In contrast, *A Dictionary of Legal Terms* shows the lowest frequency of errors found in CORP14 after eleven editions (see Table 6 in "5. Results"). These findings help to provide a base for future research, as all six and eleven editions of the SUBCORP2 dictionaries should be analysed from a diachronic perspective in order to determine if the same errors persist or if the frequency of errors change through several editions.

Similarly, there is no apparent relationship between a high frequency of errors and a high intratextual error repetition rate, as exemplified by the data obtained from DTCF, DCI and DTS.

Contrariwise, there seems to be a relationship between the number of editions and the intratextual error repetition rate: the error repetition percentage is 2.8 times lower in SUBCORP2 (sixth and eleventh editions) than in CORP14 (mainly first editions). In order to establish a more precise correlation between the number of editions and a dictionary's error repetition rate, again, all editions of DTEFC and DTJ should be analysed.

Typographical error detection and formal correctness are essential aspects affecting the quality of dictionaries. Moreover, typographical errors can hinder data access, notably errors in lemmas or sublemmas in e-dictionaries. This may even have a negative impact on translators, as they usually work with tight deadlines, and for them time is precious. Finally, our model of typographical errors could contribute to expanding knowledge and to offer new perspectives on natural language processing areas such as machine learning or data extraction, with the aim of minimising the occurrence of errors in texts. Not only have we depicted a universe of authentic errors in reference works, but we have also established relationships and observed patterns among those errors, which might be relevant to future studies.

We recommend that *Diccionario de términos económicos, financieros y comerciales/A Dictionary of Economic, Financial and Commercial Terms* (Ariel, 2012) be revised on the following grounds. First, a dictionary is a reference work, and as such, it should evince a high degree of formal correctness. Let us not forget that a dictionary is relevant not only for text production, reception and translation, but also for revision and correction purposes (Fuertes-Olivera and Tarp 2008: 79; Fuertes-Olivera 2009: 22). Second, the typographical error frequency in DTEFC is three times higher than it is in *Diccionario de términos jurídicos/A Dictionary of Legal Terms* (Ariel, 2012). It is even higher than the overall frequency in the fourteen works of the reference corpus. Third, DTEFC contains a significantly high frequency of errors in lemmas and sublemmas (as compared to the reference corpus). Last, but not least, the DTEFC is one of the cornerstones of the *Alicante Dictionaries*, a series of renowned Spanish works having been recently covered in a prominent international work in the metalexicographical field. It is reasonable to assume that the users of these dictionaries (in Spain and abroad) would expect the work to contain as few errors as possible. The making of these dictionaries (as of any other first-rate dictionary) must have been complex and costly, but formal correctness is a basic requirement that cannot be ignored.

Endnotes

1. We do not deem it necessary to go any deeper into the importance of the Spanish language. Apart from being one of the most widely used languages in the world, its relevance for lexicography has been noted, for instance, in Mairal-Usón and Fuertes-Olivera (2016: 25).

2. "Non-word errors" and "real-word errors" are generic expressions that can refer both to typographical errors and to spelling errors. The errors found in the dictionaries under study were considered typographical errors (we assumed that the authors and proofreaders of the works knew the correct spelling).
3. The Spanish expression "*contratación de la actividad*" suggests the idea of "economic upswing", whereas "*contracción de la actividad*" refers to "economic downturn". On the other hand, "*negligencia*" (meaning "negligence") is the opposite to "*diligencia*" (meaning "diligence"). Moreover, the expression "*debida negligencia*" is a *contradictio in terminis*.
4. For further information about these dictionaries, see Mateo (2018). In section "Historical Perspective", Mateo (2018: 422) states that this group of dictionaries is made up of fourteen works, whereas in section "Core issues and topics" (423), we can read: "The *Alicante* specialised dictionaries consist of thirteen specialised dictionaries ..." In section "References" (435-436), thirteen titles are listed, not including DCI. We think the correct number is fourteen (not thirteen), as DCI should definitely be considered part of the series.
5. It is worth noting that the subclassification of non-word errors and real-word errors shown here is an improvement on the one presented in Rodríguez-Rubio (2018: 80-81), where we used the same general categories ("non-word error" and "real-word error"), but less homogeneous subcategories.
6. See in Rodríguez-Rubio (2018) a complete analysis of typographical errors detected in DTCF (*Diccionario Terminológico de las Ciencias Farmacéuticas/A Terminological Dictionary of the Pharmaceutical Sciences*). The paper is written in Spanish. Please contact the author for further information.

References

Dictionaries

- Battaglia, S. (Ed.)**. 1961–2002. *Grande Dizionario della Lingua Italiana*. Torino: UTET.
- Goodwill, J.S. et al.** 1991. *The Oxford Junior Primary Dictionary for Southern Africa with North Sotho, South Sotho, Setswana and Afrikaans Words*. Cape Town: Oxford University Press.
- Johnson, S.** 1785. *A Dictionary of the English Language (Vol. I)*. Sixth edition (Preface). London: J.F. and C. Rivington.

Other Literature

- Ahmed, F., E.W. de Luca and A. Nürnberger.** 2009. Revised N-Gram based Automatic Spelling Correction Tool to Improve Retrieval Effectiveness. *Polibits* 40: 39-48.
- Bergenholtz, H. and R. Gouws.** 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica. International Annual for Lexicography* 23: 237-260.
- Bergenholtz, H. and M. Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes — Journal of Language and Communication in Business* 34: 117-141.
- Damerau, F.J.** 1964. A Technique for Computer Detection and Correction of Spelling Errors. *Communications of the ACM* 7(3): 171-176.

- Deksne, D., I. Skadiņa and A. Vasiljevs.** 2013. The Modern Electronic Dictionary that Always Provides an Answer. Kosem, I., J. Kallas, P. Gantar, S. Krek, M. Langemets and M. Tuulik (Eds.). 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*: 421-434. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- Domínguez Vázquez, M.J., C. Valcárcel and D. Lindemann.** 2018. Multilingual Generation of Noun Valency Patterns for Extracting Syntactic-Semantical Knowledge from Corpora (MultiGenera). Čibej, Jaka, Vojko Gorjanc, Iztok Kosem and Simon Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia*: 847-854. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Dziemianko, A.** 2018. Electronic Dictionaries. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*. Abingdon/New York: Routledge: 663-683.
- Fuertes-Olivera, P.A.** 2009. El *English–Spanish Accounting Dictionary*: un diccionario de internet para traductores. *puntoycoma, Boletín de los traductores españoles de las instituciones de la Unión Europea* 115-S: 22-28.
- Fuertes-Olivera, P.A.** 2014. Designing Online Dictionaries of Economics: Two Opposing Views. *Hermes — Journal of Language and Communication in Business* 52: 25-40.
- Fuertes-Olivera, P.A.** 2016. European Lexicography in the Era of the Internet: Present Situations and Future Trends. *Plenary Talk*, Beijing, 2 December 2016. Talk sponsored by the Commercial Press and the Chinese Association of Lexicography.
- Fuertes-Olivera, P.A.** 2018. Introduction: Lexicography in the Internet Era. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 1-15. Abingdon/New York: Routledge.
- Fuertes-Olivera, P.A. and S. Tarp.** 2008. La teoría [sic] Funcional de la Lexicografía y sus consecuencias para los diccionarios de economía del español. *Revista de Lexicografía* 14: 75-95.
- Fuertes-Olivera, P.A. and M. Niño-Amo.** 2013. Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés–Español de Contabilidad*. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2013. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*: 168-186. London/New York: Continuum.
- Fuertes-Olivera, P.A., M. Niño-Amo and A. Sastre.** 2019. Tecnología con fines lexicográficos: su aplicación en los Diccionarios Valladolid-UVa. *Revista Internacional de Lenguas Extranjeras* 10: 75-100.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Horák, A. and A. Rambousek.** 2018a. Lexicography and Natural Language Processing. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 179-196. Abingdon/New York: Routledge.
- Horák, A. and A. Rambousek.** 2018b. Wordnet [sic] Consistency Checking via Crowdsourcing. Čibej, Jaka, Vojko Gorjanc, Iztok Kosem and Simon Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia*: 1023-1029. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Iamartino, G.** 2017. Lexicography, or the Gentle Art of Making Mistakes. *Altre Modernità (Numero speciale — Errors: Communication and its Discontents)*: 48-78.
- Koppel, K., J. Kallas, M. Khokhlova, V. Suchomel, V. Baisa and J. Michelfeit.** 2019. SkELL Corpora as a Part of the Language Portal Sõnaveeb: Problems and Perspectives. Kosem, I. et al. (Eds.). 2019.

- Electronic Lexicography in the 21st Century: Smart Lexicography. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 763-782. Brno, Czech Republic: Lexical Computing CZ s.r.o.
- Kukich, K.** 1992. Techniques for Automatically Correcting Words in Text. *ACM Computing Surveys* 24(4): 377-439.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.
- Lemnitzer, L.** 2001. Das Internet als Medium für die Wörterbuchbenutzungsforschung. Lemberg, I. et al. (Eds.). 2001. *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*: 247-254. Tübingen: Max Niemeyer.
- Lew, R.** 2008. Lexicographic Functions and Pedagogical Lexicography: Some Critical Notes on Sven Tarp's *Lexicography in the Borderland between Knowledge and Non-knowledge*. Iwan, K. and I. Korpaczewska (Eds.). 2008. *Przegląd Humanistyczny. Pedagogika. Politologia. Filologia*: 114-123. Szczecin: Szczecińska Szkoła Wyższa Collegium Balticum.
- Lew, R.** 2013. Online Dictionary skills. Kosem, I., J. Kallas, P. Gantar, S. Krek, M. Langemets and M. Tuulik (Eds.). 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia*: 16-31. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- Lew, R. and G.-M. de Schryver.** 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.
- Mairal-Usón, R. and Fuertes-Olivera, P.A.** 2016. Recursos tecnológicos y digitales para la gestión del lenguaje científico en español. *Educación Médica* 17(2): 24-38.
- Mateo, J.** 2018. The *Alicante Dictionaries*. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 421-437. Abingdon/New York: Routledge.
- Mitton, R.** 1987. Spelling Checkers, Spelling Correctors and the Misspellings of Poor Spellers. *Information Processing and Management* 23(5): 495-505.
- Peterson, J.L.** 1986. A Note on Undetected Typing Errors. *Communications of the ACM* 29(7): 633-637.
- Prinsloo, D.J.** 2016. A Critical Analysis of Multilingual Dictionaries. *Lexikos* 26: 220-240.
- Ren, X. and F. Perrault.** 1992. The Typology of Unknown Words: An Experimental Study of Two Corpora. *Proceedings of the 14th International Conference on Computational Linguistics, COLING 1992, Nantes, France, 23–28 August, 1992. Volume 1*: 408-414.
- Rodríguez-Rubio, S.** 2018. Análisis cuantitativo de erratas del *Diccionario Terminológico de las Ciencias Farmacéuticas Inglés-Español/Spanish-English* (Ariel, 2007). *Panace@* 19(47): 76-88.
- Sassolini, E., A.F. Khan, M. Biffi, M. Monachini and S. Montemagni.** 2019. Converting and Structuring a Digital Historical Dictionary of Italian: A Case Study. Kosem, I. et al. (Eds.). 2019. *Electronic Lexicography in the 21st Century: Smart Lexicography. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 603-621. Brno, Czech Republic: Lexical Computing CZ s.r.o.
- Stemle, E.W., A. Abel and V. Lyding.** 2019. Language Varieties Meet One-Click Dictionary. Kosem, I. et al. (Eds.). 2019. *Electronic Lexicography in the 21st Century: Smart Lexicography. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 537-546. Brno, Czech Republic: Lexical Computing CZ s.r.o.
- Töpel, A.** 2014. Review of Research into the Use of Electronic Dictionaries. Müller-Spitzer, C. (Ed.). 2014. *Using Online Dictionaries*: 13-54. Lexicographica Series Maior. Berlin/Boston: De Gruyter.

- Vosse, T.** 1992. Detecting and Correcting Morpho-syntactic Errors in Real Texts. *Proceedings of the 3rd Conference on Applied Natural Language Processing, ANLC '92, Trento, Italy, 31 March–3 April 1992*: 111-118.
- Wells, F.L.** 1916. On the Psychomotor Mechanisms of Typewriting. *The American Journal of Psychology* 27(1): 47-70.
- Wheatley, H.B.** 1893. *Literary Blunders: A Chapter in the "History of Human Error"*. London: Elliot Stock. Available online at: <https://bit.ly/3csDhX5> (Last accessed on 7 October 2020).

Appendix 1: List of CORP14 dictionaries

N.B. All works are English–Spanish/Spanish–English, published by *Editorial Ariel*. SUBCORP2 works appear in blue.

Title	Code	Authorship/Date	Length (pages)	<i>Ariel</i> collection
1. <i>Diccionario de Términos de la Banca</i>	DTBA	José Mateo Martínez, 2009	635	Economy
2. <i>Diccionario de Términos de la Bolsa</i>	DTBO	José Mateo Martínez (ed. Enrique Alcaraz Varó), 2003	485	Law
3. <i>Diccionario de Términos del Calzado e Industrias Afines</i>	DTCIA	Enrique Alcaraz Varó, Brian Hughes, José Mateo Martínez, Chelo Vargas Sierra, Adelina Gómez González-Jover, 2006	853	Industry
4. <i>Diccionario Terminológico de las Ciencias Farmacéuticas/A Terminological Dictionary of the Pharmaceutical Sciences</i>	DTCF	Alfonso Domínguez-Gil Hurlé, Enrique Alcaraz Varó, Raquel Martínez Motos, 2007 (2nd print. 2011)	1,092	Medical Sciences
5. <i>Diccionario de comercio internacional</i>	DCI	José Castro Calvín (ed. Enrique Alcaraz Varó), 2007	1,144	Law
6. <i>Diccionario de Términos de Derechos Humanos/A Dictionary of Human Rights</i>	DTDH	Miguel Ángel Campos Pardillos (dir. Enrique Alcaraz Varó), 2008	381	Law
7. <i>Diccionario de términos económicos, financieros y comerciales/A Dictionary of Economic, Financial and Commercial Terms</i>	DTEFC	Enrique Alcaraz Varó, Brian Hughes, José Mateo Martínez, 2012 (6th ed., 2nd print. 2014)	1,440	Economy
8. <i>Diccionario de Fiscalidad Internacional y Aduanas</i>	DFIA	José Castro Calvín, 2009	1,912	Economy
9. <i>Diccionario de términos jurídicos/A Dictionary of Legal Terms</i>	DTJ	Enrique Alcaraz Varó, Brian Hughes, Miguel Ángel Campos Pardillos, 2012 (11th ed., 2nd print. 2014)	1,071	Law

Title	Code	Authorship/Date	Length (pages)	<i>Ariel</i> collection
10. <i>Diccionario de términos de marketing, publicidad y medios de comunicación</i>	DTMPMC	Enrique Alcaraz Varó, Brian Hughes, Miguel Ángel Campos Pardillos, 2005 (2nd ed.)	529	Economy
11. <i>Diccionario de Términos de la Piedra Natural e Industrias Afines</i>	DTPNIA	Enrique Alcaraz Varó, Brian Hughes, José Mateo Martínez, Chelo Vargas Sierra, Adelina Gómez González-Jover, 2005	621	Industry
12. <i>Diccionario de Términos de la Propiedad Inmobiliaria</i>	DTPI	Miguel Ángel Campos Pardillos (ed. Enrique Alcaraz Varó), 2003	364	Law
13. <i>Diccionario de Términos de Seguros</i>	DTS	José Castro Calvín (ed. Enrique Alcaraz Varó), 2003	793	Law
14. <i>Diccionario de términos de turismo y de ocio</i>	DTTO	Enrique Alcaraz Varó, Brian Hughes, Miguel Ángel Campos Pardillos, Víctor Manuel Pina Medina, M ^a Amparo Alesón Carbonell, 2006 (2nd ed.)	676	Tourism
			11,996	

Appendix 2: Examples of typographical errors in SUBCORP2/CORP14

A. Intratextual non-word errors in SUBCORP2

OMISSION NON-WORD ERRORS

Representation of the entry content	Page	Comments
DTEFC		
propiedad neta o valor de una propiedad o sociedad (shareholders* equity)	1312-313	It should read "shareholders"
~ valor patrimonial (shareholders* interest/equity)	1422	Similar error (same underlying term)
DTJ		
improcedencia (inappropriateness*, inopportuneness*)	852	Similar errors (different underlying terms). It should read "inappropriateness" and "inopportuneness", respectively

ADDITION NON-WORD ERRORS

Representation of the entry content	Page	Comments
DTEFC		
trigger level (nivel de activación*)	819	Other addition. It should read "activación"
= trigger price (precio de activación*)	820	

SUBSTITUTION NON-WORD ERRORS

Representation of the entry content	Page	Comments
DTEFC		
beneficio de inventario (... outcome of an inventory* on the estate ...)	941	It should read "inventory". Possible interference of the Spanish term "inventario"
~ ratio entre existencias y ventas (inventory*-sales ratio)	1325	

TRANSPOSITION NON-WORD ERRORS

Representation of the entry content	Page	Comments
DTJ		
confinar (restrcti*)	725	It should read "restrict"

B. Intratextual real-word errors in SUBCORP2

OMISSION REAL-WORD ERRORS

Representation of the entry content	Page	Comments
DTEFC		
client account (opera en nombre su [sic] cliente)	193	Preposition missing ("en nombre <u>de</u> su cliente")
DTJ		
pendiente ¹ (business to settled*)	942	It should read "to <u>be</u> settled"

ADDITION REAL-WORD ERRORS

Representation of the entry content	Page	Comments
DTJ		
motivos de denegación absolutos (absolute grounds for refusal for* refusal*)	910	Repetition of phrase

SUBSTITUTION REAL-WORD ERRORS

N.B. "WWE" is used for wrong-word errors, and "WFE" for wrong-form errors:

Representation of the entry content	Page	Comments
DTEFC		
balloon gas* gone up, the	97	WWE (intralingual). It should read "has". The exclamation mark is inverted
DTJ		
intellectual property (la propiedad intelectual [sic] se divide en dos categorías*)	328	WWE (interlingual). It should read "categorías". Besides, letter addition error in "intelectual" (it should read "intelectual")
DTEFC		
globalizar ¹ (◇ <i>En los primeros años del siglo XXI la economía está globalizar*</i>)	1150	WFE (infinitive for past participle). It should read " <i>globalizada</i> "

C. Intertextual errors in SUBCORP2/CORP14

REAL-WORD ERRORS IN SUBCORP2

Representation of the entry content	Page	Comments
DTEFC		
appropriate intellectual property (piratear la propiedad intelectual*)	66	It should read in Spanish ("intelectual")
= World Intellectual Property Organization, WIPO (organización mundial para la defensa de la propiedad intelectual*)	857	
= Organización Mundial para la Defensa de la Propiedad Intelectual* (World Intellectual Property Organization, WIPO)	1259-260	There is a discrepancy in the use of initial letter upper case, compared to the previous example
= derechos de la propiedad intelectual* , DPI (intellectual property rights, IPR)	1048	
DTJ		
intellectual (intelectual*)	328	It should read in Spanish ("intelectual")

NON-WORD ERRORS IN CORP14

Representation of the entry content	Page	Comments
DTEFC		
coefficient (V. <i>agreement* coefficient</i>)	199	It should read " <i>agreement</i> ". Addition of letter to a homogeneous digraph
= acuerdo sobre aumento de salarios según productividad (annual improvement agreement*)	889	In the previous lemma, there is an omission error in the corresponding Spanish word (" acurdo* sintético de tipos de cambio de divisas a plazo "). It should read " acuerdo "
~ venta a plazos (sale or agreement*)	1427	Similar error (same underlying term). Transposition
DTJ		
disconformidad (disagreement*)	783	It should read "disagreement"

Representation of the entry content	Page	Comments
DCI		
bilateral agreement (<i>V. tripartite agreement*</i>)	63-4	
= bilateral contract (<i>V. tripartite agreement*</i>)	64	In the same phrase
= bilateral treaty (<i>V. tripartite agreement*</i>)	64	
= sector-specific trade agreement*	532	
= sectoral agreement*	532	
= sectoral trade agreement*	532	
~ treaty (<i>V. international agreements*</i>)	606	It should read "agreements"
DFIA		
bilateral agreement (<i>V. tripartite agreement*</i>)	117	Same entry/phrase as DCI (63-4)
= bilateral arrangement or contract (<i>V. tripartite agreement*</i>)	117	
= bilateral contract (<i>V. tripartite agreement*</i>)	117	Same entry/phrase as DCI (64)
= bilateral treaty (<i>V. tripartite agreement*</i>)	118	Same entry/phrase as DCI (64)
~ limited power of attorney (<i>V. general agency agreement*</i>)	666	Letter omission

Lexicography and Language Planning in 18th Century Sweden*

Lena Rogström, *Department of Swedish, University of Gothenburg, Sweden (lena.rogstroem@svenska.gu.se)*

Abstract: The 18th century was important for Swedish linguistic development. Foreign lexical influence and orthographical standardization were intensely discussed, and the vocabulary was codified in several dictionaries, all bilingual. In this article, two questions of 18th century lexicography are studied in two influential dictionaries: Serenius (1741) and Sahlstedt (1773). The first question concerns the inclusion of Latin and Swedish legal lexical items in the lemma list; the second question examines the lexicographical treatment of the lexical item and the division into senses.

40 lexical items with a legal sense were extracted from the first two judicial handbooks written in Swedish (Rålamb 1674 and Kloot 1676). As a benchmark, Dalin (1850–55) was used; a monolingual dictionary representing the period when lexicography became fully developed in Sweden. Two modern dictionaries are also used as a comparison, SO (2009) and NSEOP (2018). The results indicate that both Serenius and Sahlstedt were loyal to the ideas of their time. They included only Latin lexical items that were already fully incorporated in Swedish and relevant for general dictionaries. The judicial senses are also discerned in the articles, but sense indicators are used in an inconsistent way and examples get mixed up. The lexicographers also lean heavily on Latin as meta language.

Keywords: 18TH CENTURY, SWEDEN, BILINGUAL LEXICOGRAPHY, LEGAL LEXICAL ITEMS, SERENIUS, SAHLSTEDT, LEXICOGRAPHICAL DEVELOPMENT, LANGUAGE PLANNING, LATIN, SWEDISH

Opsomming: *Leksikografie en taalbeplanning in 18de-eeuse Sweeds.* Die 18de eeu was 'n belangrike tydperk vir Sweedse taalkundige ontwikkeling. Vreemde leksikale invloed en ortografiese standaardisering is in erns bespreek, en die woordeskat is in verskeie woordeboeke, almal tweetalig, gekodifiseer. In hierdie artikel word twee vraagstukke van die 18de-eeuse leksikografie in twee invloedryke woordeboeke bestudeer: Serenius (1741) en Sahlstedt (1773). Die eerste vraagstuk hou verband met die insluiting van Latynse en Sweedse leksikale regsitems in die lemmalys; die tweede vraagstuk hou verband met die leksikografiese hantering van die leksikale item en die verdeling daarvan in betekenisse.

40 leksikale items met 'n regsbetekenis is onttrek uit die eerste twee regs-handboeke wat in Sweeds geskryf is (Rålamb 1674 en Kloot 1676). Dalin (1850–55), 'n eentalige woordeboek wat verteenwoordigend is van die tydperk waarin die leksikografie ten volle in Swede ontwikkel is, is as maatstaf gebruik. Twee moderne woordeboeke, SO (2009) en NSEOP (2018), word ook ter vergely-

* Thanks to Anna Hannesdóttir, Rhonwen Bowen, and the two anonymous reviewers for many valuable comments.

king gebruik. Die resultate dui daarop dat beide Serenius en Sahlstedt getrou aan die opvatting van hul tyd was. Hulle het slegs Latynse leksikale items wat reeds ten volle in Sweeds geïnkorporeer is en wat belangrik vir algemene woordeboeke was, ingesluit. Die regsbetekenis is ook in die artikels onderskei, maar betekenisaauiders is inkonsekwent gebruik en voorbeelde is deurmekaar aangegee. Die leksikograwe het ook sterk op Latyn as metataal gesteun.

Sleutelwoorde: 18DE EEU, SWEDE, TWEETALIGE LEKSIKOGRAFIE, LEKSIKALE REGS-ITEMS, SERENIUS, SAHLSTEDT, LEKSIKOGRAFIESE ONTWIKKELING, TAALBEPLANNING, LATYN, SWEEDS

1. Introduction and aims

Dictionaries play an important role in the process of language planning and have probably done so for as long as lexicography has existed (Nkomo 2018: 152; cf. Bergenholtz and Gouws 2006). Dictionaries provide a systematic framework for the description and codification of the vocabulary and they are often perceived and used as a linguistic standard for orthography and morphology. Since the most salient feature of dictionaries is the words and their semantic descriptions, the choice of headwords can function as a tool for language planning strategies. Today, Swedish dictionaries do not have an explicit function of being prescriptive regarding the vocabulary, but the general public still consider dictionaries to be normative (Josephson 2018: 180ff.). Traditionally, dictionaries are usually seen as representatives of a lexical norm; they still bring about emotional discussions on lexical change, which indicates their delicate function in the process of language cultivation.

In modern Swedish language planning, dictionaries and glossaries have a natural place in the continuous work of corpus planning. This function can be traced back to the 18th century, a significant period for Swedish lexicography when discussions on linguistic standardization were widespread and Swedish vocabulary was subject to a great expansion. One part of this important process was to replace Latin with Swedish in order to consolidate the Swedish vernacular as a useful tool, thus helping to increase and elaborate science and education in the considerably diminished kingdom of Sweden after a long period of devastating wars that ended with the death of king Charles XII in 1718. His death put an end to the era of the Great Power of Sweden and the final loss of the territories surrounding the Baltic Sea (cf. Teleman 2002). The domains of the natural sciences (with the biologist Carl Linnæus as the fix star) and the judicial domain in which a new national code of law was about to be published, were of particular interest for lexical development.

The overall aim of the study presented is to clarify how this lexical discussion is reflected in 18th century lexicography, thereby creating more knowledge about which role dictionaries played in early language planning in Sweden. Two research questions can be identified. The first question concerns the inclusion of Latin and Swedish legal lexical items in the lemma list, the second

question regards the lexicographical treatment of the lexical item and whether the judicial meaning could be clearly discerned in the division into senses (I use the simplified designations *Swedish* and *Latin* for lexical items of Germanic and Latin/Greek origin respectively.). The study focuses on the treatment of legal lexical items in two dictionaries, Serenius (1741) and Sahlstedt (1773), that both played an important role in the establishment of Swedish lexicography. As a benchmark, Dalin's dictionary from 1850–55 is used; a period when lexicography was fully developed in Sweden. A comparison to contemporary dictionaries is also made. The purpose of this study is mainly related to the role the dictionaries might have played in the strategic question of how Swedish vernacular should be strengthened by the removal of foreign (i.e. Latin) loanwords. As a consequence of this, certain lexicographic questions were brought to the forefront, above all, the question of which lexical items for specific purposes to include in a general dictionary, and how to explain their specific sense in the lexicographic structure during a period when Swedish lexicography was still finding its way.

The study consists of six sections. In section two, a brief view of research on the history of Swedish lexicography and Swedish language planning during the 18th century is presented. In section three, the difference between lexicography and terminography is discussed and some notational use is explained. Section four contains the description of the material used in the study, and section five encompasses the results of the study. Finally, section six gives a short conclusion of the results.

2. An overview of Swedish lexicography and language planning during the 18th century

The emergence and establishment of Swedish lexicography is very well mapped out (cf. Hannesdóttir 1998, Hannesdóttir and Ralph 1988, Holm and Jonsson 1991, Holmer 2016, Johansson 1997, Larsson 2003, Malmgren 1988, Norén 1991, Ralph 2012, Rogström 1998). The development of Swedish vocabulary from 1800 and onwards is thoroughly treated in the ORDAT project (Malmgren 2000). Other research on Scandinavian lexicography from a historical point of view is presented by Jacoby (1990), Haugen (1984) and Malmgren and Sköldbberg (2013). Tiisala (2018) describes the paratextual context of Serenius's English–Swedish–Latin dictionary from 1734, thereby introducing a new perspective on historical lexicography, which focuses on the context of dictionary making. This perspective elucidates the role of the dictionary in a cultural context. Studies such as Tiisala's can also tell us more about the lexicographer's view of codification and normative questions.

The roots of Swedish lexicography can be traced back to the 14th century, but its most profound development took place during the 18th century when lexicographers actively took part in the language planning process and bilingual lexicography set the foundations for linguistic standardization (cf. Han-

nesdóttir 1998, 2000, 2011; Rogström 1998, 2010, 2017). Haugen (1985) gives a brief description of how certain important Scandinavian dictionaries have contributed to the standardization of the vernaculars of the Scandinavian languages, relating to one of the basic processes in his model for language planning, i.e. the process of *codification* (Haugen 1987). This model is also used by Hannesdóttir (2000) and Rogström (2010, 2017) and functions as a basic theoretical onset also for the study at hand.

Teleman (2002, 2003, 2005) gives a detailed overview and analysis of the standardization of Swedish starting with the 17th century, and, like Haugen (1985), he also allots Swedish lexicographers an important role in creating the stance for a more systematic treatment of Swedish orthography and morphology. Teleman also emphasizes the increased possibilities of lexical choices that were offered by the newly coined Swedish equivalents to foreign headwords in bilingual dictionaries (2002: 59).

During the late 17th century, Sweden held a position in Europe as a strong and important country with provinces around the Baltic Sea. The geographical situation made the nation multilingual, with vernaculars like Finnish, German, Estonian, Latvian and Russian (Teleman 2002: 24). In order to reflect the glory of this vast nation, Swedish had to be strengthened and cultivated. During the 18th century, Swedish was the dominating language of the official administration but Latin was unthreatened in certain domains such as the fine arts, science, jurisprudence and higher education. Since Sweden confessed to the Protestant church, vernaculars were used in church and for religious practice (Teleman 2002: 25). The status of Latin and foreign lexical influence on the vernaculars was profoundly discussed in Sweden as well as in the rest of Europe, and in the late 17th and early 18th centuries, the position of Latin started to give way to Swedish, and the choice between the two languages in technical domains was not obvious any longer (cf. Gunnarsson 2011). In the legal domain, discussions on the use of Latin had started already in the 17th century. Legislation and legal processes had always been conducted in the vernacular whereas Latin was used for legal education and jurisprudence. In 1614, an important reform was carried out in the Swedish legal system when the Courts of Appeal were inaugurated, overtaking the role of the highest instance of appeal that, up until then, had been performed by the King. The new situation created a need for more professional lawyers and judges also in the lower courts (Korpiola 2014: 30; Modéer 2014: 402ff.), and the system for legal education was slowly improved. In 1674 and 1676, the first two legal handbooks in Swedish were published, which was perhaps as a response to the need for better legal education (Rogström and Landqvist 2018a, b). Not only were the changing conditions of the judicial system of vital importance for the improvement of judicial education; the reformation of the judicial system also had a certain linguistic impact. In 1686, King Charles XI appointed a committee to compile a common law for the whole country; this process was not completed until 1736 when a new Swedish code of law was ratified by the Swedish parliament. One of the guiding princi-

ples for the work of the committee was to avoid foreign loanwords, especially Latin ones (Westman 1912: 61). The new code of law turned out to be a linguistic role model already from the start (Teleman 2002: 79) and became important for the standardization of Swedish. A more comprehensive description of the linguistic efforts of the King's Law Committee of 1686 is given in Rogström (2009).

Law and language are closely connected. This is one reason for choosing legal language as material for the study at hand. Another reason is that legal vocabulary stands out in contrast to many other technical vocabularies and shares many similarities with general language. Legal language is rooted in national culture and used by laymen in a way that many other technical languages are not (Mattila 2013), which makes it interesting to analyse from a lexicographical point of view. Hoare (2009: 85) characterizes Law as a subject "of exceptional lexicographic interest [...]" and discusses the need to include both Latin terms and their vernacular equivalents in legal dictionaries. One reason for this is the two different kinds of targeted readers: the professional and the layman (2009: 86). Since judicial language is often used by ordinary target users in their daily lives, legal vocabulary could be expected to be a part of the lemma list of general dictionaries as well, as Svensén (2009: 71f.) points out. Nowadays, most people encounter legal matters at various points during their lifetime and there is no reason to believe that this was not the case also in the 18th century. From that perspective, the argument for including legal lexical items in 18th century general dictionaries would be the same as for modern lexicography although access to dictionaries was more scarce for the 18th century user than is the case today. Nielsen (2015) argues that the relationship between legal language and lexicography needs to be clarified in greater detail for modern lexicography, and according to the judicial and linguistic situation in 18th century Sweden, such a study seems to be important for that period as well.

3. Lexicography and terminography

Legal vocabulary is not a product of systematic terminological work as many other technical vocabularies are (Mattila 2013: 137ff.), and its connections to general vocabulary as well as to language for specific purposes make it interesting from a lexicographical point of view. Therefore, general vocabulary is usually treated in dictionaries whereas terminology is usually treated by terminologists in the area of terminography. Dictionaries of general language are expected to include frequently used technical terms, whereas dictionaries of language for specific purposes (LSP) hardly ever include any general words (Svensén 2009: 3, Vrbinc and Vrbinc 2013: 440). For a more detailed discussion on scientific and technical dictionaries, see Becker (2015). The dichotomy between lexicography and terminography has different theoretical onsets but in recent years a certain rapprochement is discerned between the two disciplines (Bowker 2018, Faber 2012, Fontenelle 2014, Geeraerts 2015). Today, more

technical terms tend to be included in dictionaries of language for general purposes, LGP-dictionaries, thus changing the boundaries between dictionaries of LGP and LSP. One explanation for this might be that technical language concerns the target user to a greater extent nowadays. When more and more people get highly dependent on technology, the need for technical vocabulary increases (cf. Urbinc and Urbinc 2013).

During the 18th century, however, neither of the disciplines of lexicography and terminography was clearly recognized as belonging to different theoretical onsets. Although Linnæus organized the naming of botanical items in a binominal conceptual system during the 18th century, theoretical discussions of terminology were not common until the end of the 19th century, following the rapid growth of electricity (Laurén et al. 1997: 16f); consequently, there is no reason to believe that 18th century lexicographers made a difference between *terms* and *words*. Therefore, I have chosen to use the designation *lexical item* in this study, since it might cover single lexical items as well as multi-word units regardless of their terminological status. Swedish and Latin lexical examples are written in italics, *domare*, *judex*, and the English translation in brackets and a single apostrophe ['judge']. The semantic meaning of a lexical item is used with straight apostrophes: 'judge'. Lexicographical examples are written in smaller typography.

4. Material

In this section the material of the study is presented. In subsection 4.1 the lexical sample is described and in subsection 4.2 the dictionaries are presented in chronological order.

4.1 The lexical items

This study seeks to establish which legal vocabulary was included in two important dictionaries, by Serenius and Sahlstedt respectively, from the 18th century when bilingual lexicography became important for linguistic standardization in Sweden (Hannedóttir 2000). In order to find an appropriate sample of such lexical items, I have taken central legal lexical items from the first two legal handbooks in Swedish, by Rålamb (1674) and Kloot (1676). The lexical items were originally extracted by Rogström and Landqvist (2015, 2018a, b) and used in their studies on Rålamb's and Kloot's use of Latin and Swedish. All the lexical items in their sample originate from the authors' chapters on the Civil Procedural Law. Rålamb and Kloot treat the same content, i.e. legal processes, and both books were written for educational purposes. The handbooks most probably were used in the lower courts, thus being tools for the growing professionalization of Swedish judges and lawyers. There is no evidence of the two authors being aware of each other's books (Björne 1995: 36).

For the study at hand, a total of 40 lexical items were used, and the ones that were omitted from the original sample were mostly long phrases, probably originating from Latin structure. One such example is the phrase: "*processens begynnelse igenom saksens inför rättens föreställande*" ['The opening of the process by presenting the case to the court'] [my translation]. The Latin expression is: *Litis Contestatio*.

In the handbooks, many of the lexical items are used synonymously in Swedish and Latin, i.e. *appellera/vädja* ['to appeal']; *domare/judex* ['judge']; *kärande/actor* ['plaintiff']. A few of the Latin lexical items could be called *true Latin* (i.e. *judex*) since they are used in their Latin spelling and form, whereas other Latin lexical items have been adjusted to Swedish spelling and morphology (i.e. *citera*), hence called *adjusted Latin*. By studying which legal lexical items from the handbooks are included in Serenius's and Sahlstedt's dictionaries and the way they were described, conclusions can be drawn about how lexicographers understood the lexical items according to their general and/or technical status, and to what extent the Latin lexical items were included in relation to the Swedish ones (cf. Teleman 2002: 59).

In order to make sure that the lexical items from the handbooks were used also in other contemporary legal contexts, the protocols of the King's Law Committee of 1686, (edited by Sjögren 1900–1909), have been used for comparison. In the protocols, both general vocabulary as well as legal lexical items are discussed in detail including several of the lexical items chosen for this study. Many of the same lexical items were included in the draft of the code of law (Rogström 2010). This enhances their chances of being included in a dictionary, since it could be assumed that they would be of general interest. The sample is made up of the following 40 lexical items (written in modern spelling):

Table 1: The sample of lexical items used in the study

Lexical items	Translation
actor	plaintiff
appellera	to appeal
bilägga	to settle (a dispute)
citation	writ of summons
citans	plaintiff
citatus	defendant
citera	to summons
contumacia	insubordination
dilation	postponement

dom	verdict
domare	judge
domsaga	judicial district
döma	to judge
execution (effektuera)	to execute (effectuate)
frikänna	to acquit
fälla	to pass (a sentence)
förlikning	reconciliation
judex	judge
jäva	take exception to
kärande	plaintiff
lagsöka	to summons
ogilla	to disallow, reject
olydnad	insubordination
part	party
process	lawsuit
rannsaka	to cross-examine
rekvisit	requisites
resolution	resolution
rätt	court
rättegång	trial
sak	case
sakförare	lawyer
stämna	to summons
stämning	writ of summons
svara för rätta	to answer in court
svarande	defendant
tredska	insubordination
uppskov	postponement
vädja	to appeal
åklagare	prosecutor

4.2 The dictionaries

There are three dictionaries used in this study, two of which were compiled in the 18th century and represent the central part of the study: Serenius (1741) and Sahlstedt (1773). A third dictionary, Dalin (1850–55), represents the period when Swedish lexicography became fully developed. Two modern dictionaries are also used for comparisons, *SO (Svensk ordbok)* (2009) and *NSEOP (Norstedts svensk-engelska ordbok (professionell))* (2018) (See section 4.2.1. for a detailed description of the dictionaries.). The dictionaries were chosen according to three parameters: good lexicographical quality, an expectation that they would reflect the contemporary vocabulary of their time, and, finally, they would represent two different lexicographical typologies, monolingual and bilingual dictionaries. Thus, the dictionaries represent both the pioneer era of Swedish lexicography and the fully developed stage of the 1850's and onwards. This also means that the dictionaries could be subject to different lexicographical techniques in their descriptions of technical language.

4.2.1 The 18th century dictionaries: Serenius (1741) and Sahlstedt (1773)

Serenius's *Dictionarium Suethico-Anglo-Latinum* (1741) is the first dictionary of Swedish and English. In 1734, Serenius published an English–Swedish dictionary that he later turned into a Swedish–English one by selecting a smaller range of lemmas adjusted to the new function of the dictionary (see Rogström 1998 for a detailed description). The dictionary explains the Swedish lemmas in both English and (often, but not always) Latin, but the main languages are Swedish and English and the structure is clearly bilingual. It comprises 8,133 lemmas (Rogström 1998). The dictionary is the first Swedish dictionary to give information of the parts of speech for every headword and for most derivatives.

The preface explains that the dictionary was compiled with the purpose of facilitating trade between England and Sweden. Subsequently, the vocabulary encompasses many lexical items that are central for tradesmen (Rogström 1998). The intended target users are native speakers (hence L1-speakers) of Swedish who need to communicate in English. Since the dictionary includes a section with rules of English pronunciation (Olofsson 1991), the dictionary was most probably intended as a resource also for spoken English. The dictionary is thereby a bilingual dictionary for active use for L1-speakers of Swedish (Svensén 2009: 15). The dictionary has no information of how the articles are structured. Since the target users were expected to know Swedish, explanations of different senses and information concerning stylistic register of the Swedish headwords need not be given in detail. On the other hand, more extensive information is needed for the target language, English, in order for the user to be able to choose the right equivalent (Svensén 2009: 15). A certain kind of symbol is used as sense indicator, a small hand with a pointing index finger (☞), also

known as *manicule* or *printer's fist*.

There is reason to believe that Serenius was interested in linguistic standardization. He dedicates the dictionary to the Swedish parliament, hoping that one day the members of the parliament will use their power to bring some order into the standardization of the Swedish language. Rogström (1994) has also shown that Serenius made use of the orthographical rules that were promoted by the Royal Swedish Academy of Science, who also sponsored the edition of Sahlstedt's dictionary, and in their early days had a pronounced interest in Swedish language planning (Teleman 2002: 69ff.)

Sahlstedt's *Svensk ordbok* ['*Swedish Dictionary*'] (1773) comprises approximately 21,500 lemmas (Hannesdóttir 1998: 106f.). Sahlstedt's dictionary was very influential in its day. His main purpose was to improve the principles for orthography and morphology for Swedish, and he was the first lexicographer to present a system of declinations for the nouns, based on gender and plural endings. Sahlstedt's dictionary is a bilingual Swedish–Latin dictionary, but Sahlstedt never had the intention of writing a dictionary that could be used for production in Latin. Instead, his dictionary aimed at explaining Swedish vocabulary using Latin. Consequently, the form of the target language differs somewhat between equivalents and more explanatory phrases. Sahlstedt put some effort into describing on what grounds he had compiled his dictionary. He explains his choice of headwords, and he discusses the importance of an orthographic standard; this is a very relevant factor for both lexicographical compilation as well as standardization on a larger scale. He does not explicitly mention which target users he has in mind, but he states that dictionaries are important both for foreigners who strive to learn a foreign language, as well as for natives to learn the correct usage of words (Sahlstedt, preface p. 2).

Sahlstedt obviously aims at a broad target-user group consisting of learners as well as L1-speakers of Swedish, and with that approach, his dictionary could be considered as bi-functional rather than bi-directional (cf. Hannesdóttir 2014). The problem of a foreign learner understanding the Swedish source language is solved with the use of a Latin index where the Latin equivalents to the Swedish lemmas are used as the source language. Hence, the index functions as an inverted entry list to the main part of the dictionary. However, the learners would most probably have to know some Swedish in order to use the index successfully, since many declinations and compounds are placed in the article of the headword — important information that is lacking in the index.

In his foreword (p. 6), Sahlstedt discusses LSP in conjunction with the lexical material he has included in his dictionary: "*Books on Theology and Law use special words, and a certain way of writing. I have only included those words that are used in general language or in books that treat such subjects. They are marked with a special sign.*" [My translation.] At the end of the preface (p. 16), Sahlstedt lists the abbreviations that he uses in the dictionary: "*V. jur. term. jur. Vox juridica. Terminus juridicus. Words or sayings that are used in the court.*" [My translation.]

Sahlstedt compiled his dictionary on commission from the Royal Swedish Academy of Sciences that also had supported his grammar, published in 1769, to accompany his dictionary. The two books have played an important role in the codification of Swedish grammar and orthography. Sahlstedt could not have been aware of the impact his books would have on linguistic standardization, but he was well aware of the advantage of combining a grammar and a dictionary to exemplify linguistic use of certain lexical items and constructions (Sahlstedt 1773, preface, p. 9).

4.2.2 The 19th century dictionary: Dalin (1850–55)

Dalin's *Ordbok öfver svenska språket* ['Dictionary of Swedish'] (1850–55), could be said to be the representative of the fully established Swedish lexicographical development, since it was the first complete monolingual defining dictionary of Swedish. It comprises 60,000 lemmas (Malmgren 1988: 201) and it held the position of being the standard Swedish dictionary for nearly a century. Dalin's dictionary was first described in detail in Norén (1991) and later in Hannesdóttir (1998). Even compared to modern lexicography, Dalin's dictionary shows a high standard, especially considering the semantic description with senses and subsenses (Malmgren 1988: 205). The articles are structured with separate numeric sense indicators, and the preface gives comprehensive information concerning both the structure and intended usage of the dictionary. Dalin makes use of approximately 60 diasystematic labels, listed in the preface. One of the labels is *Lagt.*, short for *Lagterm* ['Law term']. Dalin has no information, however, of how the subject fields should be understood or what the difference is between a *word* and a *term*.

4.2.3 The modern dictionaries: SO (2009) and NSEOP (2018)

Two modern dictionaries are used as comparative material, in order to see which legal lexical items in the study are still in use. The monolingual dictionary is *Svensk ordbok utgiven av Svenska Akademien* (SO) ['The Contemporary Dictionary of the Swedish Academy'] (2009). This is the contemporary, standard, defining dictionary in Swedish, available both in print and on the Internet (www.svenska.se). It comprises 65,000 headwords and the intended target users are speakers of Swedish and learners at an advanced level. The dictionary is suitable for both passive and active use, i.e. for both understanding and producing Swedish.

The modern bilingual dictionary is *Norstedts svensk-engelska ordbok, Professionell* (NSEOP) ['Norstedt's Swedish-English Dictionary, Professional']. The printed edition was published in 2010 and the electronic, somewhat enlarged edition, was released in 2018. The electronic edition comprises approx. 87,000 headwords and 53,000 phrases. NSEOP (2018) can be categorized as an active dic-

tionary for L1-speakers of Swedish (cf. Svensén 2009: 15). The preface in the printed dictionary (2010) gives a brief description of how the dictionary is structured in order to give the user the best opportunities to choose the right equivalent, but the digital version used for the present study has no such information.

5. Results

In this section, the results of the study are presented. In section 5.1, the representation of the lexical items in the dictionaries is presented and discussed, followed by section 5.2. in which the results from the lexicographic study are presented. Finally, a summary of the results is given in section 5.3.

5.1 Lexical representation in dictionaries

The first of the research questions set out for this study concerns the choice between Latin and Swedish legal lexical items in the dictionaries. In relation to this question, the 18th century dictionaries are of special interest. One assumption is that an overlap between the vocabulary in the legal handbooks and the older dictionaries reveals something of the stance of the dictionaries in the process of lexication (Haugen 1987), where one strategy practised was to replace Latin legal lexical items with Swedish. By tracking the sample of 40 lexical items over time, it is possible to get a notion of how the lexical items were represented in early Swedish lexicography and how the legal character of the lexical items was expressed with regard to the typology of each dictionary.

The results are presented in alphabetical order in Table 2. ("X" in the table means that the actual lexical item is included in the dictionary with a legal sense. "—" means that the lexical item is not included in the dictionary. "Not legal" means that the lexical item is included but lacking a clear legal sense, and "index" means that the lexical item is placed in the Latin index.) Some lexical items function as sublemmas instead of headwords, and in those cases the headword of the article is given. A few of the lexical items have been considered equals to the lexical items in the sample, although they have a different form. All lexical items are written using modern spelling.

Table 2: Lexical representation in the dictionaries by Serenius (1741), Sahlstedt (1773), and Dalin (1850–55)

Lemma	Serenius (1741)	Sahlstedt (1773)	Dalin (1850–55)
actor	—	X (<i>index: actor</i>)	X (<i>aktor</i>)
appellera	—	—	X
bilägga	X	X	X

citans	—	—	—
citation	—	—	X (<i>not judicial</i>)
citatus	—	—	—
citera	X	X	X (<i>not judicial</i>)
contumacia	X	X (<i>index: contumax</i>)	X
dilation	—	X (<i>index: dilatio</i>)	X
dom	X	X	X
domare	X	X	X
domsaga	—	X	X
döma	X	X	X
execution	X (<i>in EXECUTORIAL</i>)	—	X
frikänna	—	X	X
fälla	X	X	X
förlikning	X (<i>in FÖRLIKA</i>)	X	X
judex	—	X (<i>index: judex</i>)	—
jäva	X	X	X
kärande	X (<i>in KÄRA til en</i>)	X (<i>in KÄRA</i>)	X
lagsöka	X (<i>lagsökning, in LAG</i>)	X	X
ogilla	X	X	X
olydnad	X (<i>in OLYDIG</i>)	X	X (<i>not judicial</i>)
part	X	X	X
process	X	X	X
rannsaka	X	X	X
rekvisit	—	—	X
resolution	—	—	X
rätt	X	X	X
rättegång	X	X	X
sak	X	X	X
sakförare	—	—	X
stämma (vb)	X	X	X
stämning	X	X	X
svara för rätta	X (<i>svaranden för rätta, in SVARA</i>)	X (<i>in SVARA</i>)	X (<i>in SVARA</i>)
svarande	X (<i>in SVARA</i>)	X (<i>in SVARA</i>)	X (<i>in SVARA</i>)
tredska	X (<i>in TRESK</i>)	X	X (<i>not judicial</i>)
uppskov	X	X	X
vädja	X	X	X
åklagare	X	X	X

Jacob Serenius, *Dictionarium Suethico–Anglo–Latinum* (1741)

Serenius's dictionary (1741) lacks twelve lexical items out of 40 (*actor*, *appellera*, *citation*, *citans*, *citatus*, *dilation*, *domsaga*, *frikänna*, *judex*, *rekvisit*, *resolution*, and *sakförare*). Nine of these are of Latin origin and three could be defined as Swedish, i.e. *domsaga* ('judicial district'), *frikänna* ('acquit') and *sakförare* ('lawyer').

Serenius's omission of the Latin lexical items is interesting and might indicate that, by omitting them, he made a deliberate decision (cf. Telemann 2002: 59). *Actor*, *appellera*, *citation*, *dilation*, *rekvisit* and *resolution* were common in legal domains, and their frequent use can be seen in their formal adjustment to Swedish morphology and, in some cases, orthography. Still, Serenius chooses not to include them. *Citans*, *citatus* and *judex*, on the other hand, were probably conceived as true Latin, and no obvious candidates for the Swedish lemma list. However, the Latin lexical items lacking in Serenius's dictionary are represented by corresponding Swedish equivalents, most of them were also included in the sample. The only exceptions are *rekvisit* and *resolution* (cf. Table 3). There is no obvious explanation, though, why Serenius has not included the verb *frikänna* ['acquit'] when he includes its antonym *fälla* ['pass a sentence']. *Domsaga* is mostly used in a professional administrative discourse, and Serenius might have considered it less useful in his dictionary. Instead of *sakförare* ['lawyer'] Serenius uses the (Latin) lexical item *advocat*, which is still the natural lexical choice in Swedish (with the spelling *advokat*). (*Advocat* was also actually used by Rålamb and Kloot, but not in the chapters chosen for this study. The earliest usage is noted in the King Gustaf Vasa's Bible (1541) according to SAOB (the Swedish historical dictionary), and this was, of course, a strong incentive of the usage at the time.)

The comparison between the sample of lexical items extracted from the legal handbooks and their representation in Serenius (1741) shows that Serenius preferred not to include the Latin legal headwords if there were corresponding Swedish equivalents at hand. The result indicates that Serenius could have promoted the strategy of replacing Latin with Swedish lexical items during the 18th century.

Abraham Sahlstedt, *Svensk ordbok* ['Swedish Dictionary'] (1773)

Sahlstedt (1773) omits eight lexical items of the sample from the main part of his dictionary; one Swedish (*sakförare*), and seven Latin (*appellera*, *citation*, *citans*, *citatus*, *execution*, *rekvisit*, *resolution*). Apart from *execution*, Sahlstedt omits the same Latin lexical items as Serenius, and just like Serenius, Sahlstedt also prefers *advocat* to *sakförare*. One of the Latin lexical items included by Sahlstedt, (*citera*), shows no clear judicial meaning, since it is only used in the sense of 'to quote'. Sahlstedt uses the Latin index for some of the Latin lexical items: *actor*, *contumacia* (in the form *contumax*), *dilatatio*, and *judex*. It seems as if Sahlstedt prefers to use Swedish headwords in the main part of the dictionary and Latin lexical items (with Swedish equivalents) in the index. *Citation*, *citans*, and *citatus* are not included at all; one reason for this might be that they are regular word

formations related to *citera*, thereby easy to understand anyway by someone who is familiar with Latin.

The two lexicographers' treatment of the Latin lexical items points in the same direction — to promote Swedish headwords if possible — except when lexical items of Latin origin are the most established and (probably) frequently used. *Part* and *process* are two more examples of such lexical items. According to SAOB both lexical items have been in use since at least 1540.

A.F. Dalin, *Ordbok öfver svenska språket* ['Dictionary of Swedish'] (1850–55)

Just like Sahlstedt, Dalin (1850–55) does not include the lexical items *citans*, *citatus* or *judex*. He is, on the other hand, the first one to include *appellera*, *rekvisit*, *resolution* and *sakförare*. In the period when Dalin's dictionary was published, several of the lexical items had undergone a semantic change during which the judicial meaning had faded. Besides *citation* and *citera*, the lexical items *olydnad* and *tredska* (both meaning 'obedience', 'insubordination') had lost their judicial meaning.

SO (2009); NSEOP (2018)

It does not come as a surprise that the modern dictionaries also lack the same Latin lexical items as the older dictionaries. SO (2009) lacks six lexical items from the sample (*actor*, *citans*, *citatus*, *contumacia*, *dilation*, and *judex*). The lexical item *svara för rätta* is represented by a synonymous verb, i.e. *stå inför rätta* in an example in the article of RÄTT (not SVARA, as in Serenius 1741). There are also some lexical items from the sample that are included in the dictionary but without their earlier, judicial meaning (*citation*, *citera*, *olydnad*, *tredska*, *resolution*, and *uppskov*).

NSEOP (2018) lacks only four of the 40 lexical items in the sample, all of them true Latin: *citans*, *citatus*, *contumacia*, and *judex*. Nine lexical items have lost their judicial meaning and are used in a more general sense: *actor*, *bilägga*, *citation*, *citera*, *dilation*, *execution*, *olydnad*, *resolution*, and *uppskov*.

5.1.1 Summary of results

Two results can be discerned in this study. The first is that the Swedish equivalents from the sample were preferred to their Latin counterparts. The Latin lexical items included in the main parts of the dictionaries are all adjusted Latin. This means that the lexical items were most likely integrated into general Swedish vocabulary well before the 18th century. Some of them are still used in Swedish legal context (*appellera*, *kontumacie-dom*, *rekvisit*, *part*, and *process*). True Latin is avoided altogether, except for four examples, placed in Sahlstedt's Latin index. Although *citatus* is used in the 17th century Swedish legal handbooks by Rålamb and Kloot respectively in that form, it was probably perceived as true Latin and could not be expected in a dictionary for general purposes. Probably the same could be said about *judex*, which is listed only in

Sahlstedt's Latin index. This shows that Sahlstedt did not consider all the lexical items in the sample to be suitable for the general part of the dictionary and that he probably recognized the true Latin forms as part of a more professional, legal context, not relevant for his dictionary. Table 3 gives an overview of the Latin lexical items, their Swedish equivalents and which lexical item is used in each dictionary.

Table 3: Latin lexical items and their representation in all of the dictionaries (i = index; nj = not judicial)

Lexical items	Serenius (1741)	Sahlstedt (1773)	Dalin (1850-55)	SO (2009)	NSEOP (2018)
actor; citans; kårande (plaintiff)	— kårande	actor (i) kårande	aktor kårande	— kårande	aktor kårande
appellera; vådja (to appeal)	— vådja	— vådja	appellera vådja	appellera vådja	appellera vådja
citation; stämning (writ of summons)	— stämning	— stämning	citation (nj) stämning	citation (nj) stämning	citation (nj) stämning
citatus; svarande (defendant)	— svarande	— svarande	— svarande	— svarande	— svarande
citera; lagsöka; stämna (to summons)	citera lagsöka stämna	citera lagsöka stämna	citera (nj) lagsöka stämna	citera (nj) lagsöka stämna	citera (nj) lagsöka stämna
contumacia; olydnad; tredska (insubordination)	contumax olydnad tredska	contumax (i) olydnad tredska	contumacia olydnad (nj) tredska (nj)	— olydnad (nj) tredska (nj)	— olydnad (nj) tredska (nj)
dilation; uppskov (postponement)	— uppskov	dilation (i) uppskov	dilation uppskov	— uppskov (nj)	— uppskov
execution (effectuation)	execution	—	execution	exekution	exekution (nj)
judex; domare (judge)	— domare	judex (i) domare	— domare	— domare	— domare
part (party)	part	part	part	part	part
process; rättegång (lawsuit, trial)	process rättegång	process rättegång	process rättegång	process rättegång	process rättegång
rekvisit (requisites)	—	—	rekvisit	rekvisit	rekvisit
resolution (resolution)	—	—	resolution	resolution	resolution

The second result is that most of the Swedish lexical items are still in use in legal vocabulary although a few of them have been subject to semantic change due to a process of determinologization (Svensén 2009: 71). These are used in other contexts than the judicial, i.e. *actor* (now *aktör*), *olydnad*, *rannsaka*, *uppskov* and *tredska*. The process of determinologization is starting to show in Dalin's

dictionary but is most explicit in the contemporary dictionaries (SO 2009 and NSEOP 2018) that still include many of the lexical items from the sample, but without their judicial meaning.

One conclusion that can be drawn from this study is that the coverage of the sample is very good in all the dictionaries and that the lexicographers prefer Swedish headwords to Latin. This strengthens the impression that the 18th century dictionaries are reliable as sources of contemporary, 18th century vocabulary.

5.2 Lexicographical treatment of the lexical items

In 18th century Sweden, a vast number of dictionaries were produced, all of them bi- or multilingual, covering all the major European languages. During the 18th century, Swedish lexicography slowly stabilised and the structure of the dictionaries became more and more refined. In the prefaces, Serenius as well as Sahlstedt write about their strenuous labour with the compilation of their dictionaries, but there is no explicit information on the lexicographical structure given by the authors themselves.

The study in this section is based on analyses of the articles that treat the lexical items of the sample. The analyses are illustrated by a detailed description of the article FÄLLA [*to pass a sentence*] from the three oldest dictionaries, thus giving a more detailed glimpse of the establishment of lexicographical structure up until 1850. The choice of FÄLLA is based upon the fact that the word form is homonymous in Swedish and could be either a noun or a verb. The verb is polysemous with four different senses (according to SO 2009). The combination of the homonymous as well as the polysemous aspects of FÄLLA is interesting from a lexicographical point of view and especially the way the judicial meaning is described in the different dictionaries. In addition, the articles of FÄLLA (v) in the older dictionaries all include examples of constructions, phrases and word formation — the treatment of which is interesting to examine from a historical point of view. And — not the least important — the length of the article in each dictionary is manageable for these circumstances.

Jacob Serenius, *Dictionarium Suethico–Anglo–Latinum* (1741)

The oldest dictionary, Serenius (1741), is a bilingual dictionary aimed at L1-speakers of Swedish (see section 3). Out of the 40 lexical items in the sample, Serenius includes 28, and 20 are used as headwords introducing an article. Six lexical items are found as a sublemma in semantically related articles (see Table 2), each one supplied with information concerning its part of speech, in this case a noun: *execution; förlikning; kärande; olydnad; svarande* and, *tredska*.

The verb *lagsöka* is represented as a nominalization, *lagsökning*, that could be found in the article of LAG. The phrase *svara för rätta* [*to answer (in court)*] is represented by its nominal counterpart, *svaranden för rätta* [*the defendant*], and it is placed in the article of SVARA.

Serenius usually treats homonyms in separate articles. The senses of the monosemous headwords are explained by English and Latin equivalents and sometimes examples in Swedish and English. Polysemous headwords are explained in the same technique, but also have their different senses separated by the sense indicator (☞). Serenius does not make use of subject field labels; this is something that Vrbinč and Vrbinč (2013), for example, think should be used in all instances where technical language is treated in bilingual dictionaries.

Since the dictionary is a bilingual one aimed at L1-speakers, the target user is expected to have a good understanding of the semantics of the source language but needs guidance regarding the choice of what corresponding equivalent to choose in the target language. That is probably the reason why Serenius puts very little effort into explaining different senses and stylistic register in the Swedish source language.

The different equivalents are introduced in the beginning of the article in both English and Latin (see picture 1): To fell, *sternere*; *Fälla* (*perdere*) To drop. ☞ *Fälla* (*condemnare*) To cast, To condemn. ☞ *Fälla ned eller kull* (*subvertere*) To give one a fall. *Fälla ankar*, To let go the anchor. *Fälla från*, To cut or fell trees. *Han fälde bort sin klocka*, He dropt his watch. ☞ *Han fälde et ord*, He dorpt a word, he let fall a word. *Fälla ladden*, To cast his coat. *Fälla horn eller fjedr*, To mew. *Fälla fjedrarna* (*de avibus*) To moult or moulter. *Balkarna fälla fjedrarna*, The hawks mew. ☞ *Rets ten lät fälla Ert*, The Court will cast you. *Fäll*, *adj.* Felled. *Fälla*, *f. i en Rog*, Fall of wood. *Fäll*, *bör*, Folding door. *Fällport*, *fäll-bom*, *Portcullis*. *Fällstol*, Folding chair.

Picture 1: Article FÄLLA in Serenius (1741)

FÄLLA, v. a. To fell; *sternere*. ☞ *Fälla* (*perdere*) To drop. ☞ *Fälla* (*condemnare*) To cast, to condemn. ☞ *Fälla ned eller kull* (*subvertere*) to give one a fall. *Fälla ankar*, To let go the anchor. *Fälla från*, To cut or fell trees. *Han fälde bort sin klocka*, He dropt his watch. ☞ *Han fälde et ord*, He dorpt a word, he let fall a word. *Fälla ladden*, To cast his coat. *Fälla horn eller fjedr*, To mew. *Fälla fjedrarna* (*de avibus*) To moult or moulter. *Balkarna fälla fjedrarna*, The hawks mew. ☞ *Rets ten lät fälla Ert*, The Court will cast you. *Fäll*, *adj.* Felled. *Fälla*, *f. i en Rog*, Fall of wood. *Fäll*, *bör*, Folding door. *Fällport*, *fäll-bom*, *Portcullis*. *Fällstol*, Folding chair.

Serenius then gives a selection of examples, sometimes separated by the manicule. One would expect the examples to exemplify the usage of the equivalents, but this is not always the case, since Serenius also illustrates equivalents that are not used earlier in the article. The first example, *Fälla ankar*, uses an equivalent that is not explained earlier [to let go], and the sentence *Han fälde et ord* He dorpt [sic] a word, he let fall a word is placed together with examples of how to use the verb To mew or To moult, two other verbs that are not used earlier as equivalents.

It is not altogether clear how the different senses are to be understood in the article. Two Latin equivalents can be identified, *condemnere* and *subvertere*,

each one representing a sense of FÄLLA. The judicial meaning is signalled by a sense indicator but has no equivalent and is only given by the means of a linguistic example in the form of a phrase: *Retten lät fälla Ehr*, The Court will cast you (written in another tense). The correspondence between the Swedish verb *fälla* and the English equivalent To cast is the important information in the example, and the user is supposed to understand that this is the judicial sense, since the example treats a legal situation signalled by the noun *Retten* The court. The English and Latin equivalents (cast and *condemnere*) of the verb are given already in the beginning of the article. After this linguistic example, Serenius finishes the article by giving the adjective (i.e. past participle) of the verb (*Fälld*, adj.) and the homonymous noun, placed in the context of a linguistic example (*Fälla. s. i en skog* / Fall of wood). The article ends with the composites *Fäll-dör*, *Fäll-port*, *fäll-bom*, and *Fällstol*, which all are compounds to the sense 'to fold'. However, this different sense is not marked with any sense indicator. The noun FÄLLA ['trap'] is also headword in an article of its own, preceding the verb. Of course, the example *Fälla. s. i en skog* / Fall of wood had been more suitable there.

Although Serenius makes an effort to structure the article logically, he does not succeed completely. A major failing of the structure is that the division into senses is not arranged and described in a systematic way. One explanation for this might be that Serenius had difficulties in handling the material when he turned his English–Swedish dictionary into a Swedish–English one, mixing up some of the senses and categories. It is clear though that Serenius is aware of the importance of distinguishing the different senses, and he tries to do so by using three different techniques: Latin equivalents, manicules and examples. He identifies the judicial meaning of the lexical item, and the examples are relevant.

Abraham Sahlstedt, *Svensk ordbok* ['Swedish Dictionary'] (1773)

One group of target users of Sahlstedt's dictionary is Swedes who want to learn more about their vernacular, especially orthography and morphology. Another group is foreigners who want to learn Swedish. This is also the reason why Sahlstedt uses Latin as the defining language (preface p. 11, see also Hannesdóttir 1998: 238, 277 and Hannesdóttir 2014: 193ff.). In doing so, Swedish readers with little or no knowledge of Latin could also learn what the Swedish lexical items are in Latin; Sahlstedt especially stresses the technical lexical items (see section 4.2.1) and he explains what subject field labels he uses: *V. jur. term. jur. Vox juridica. Terminus juridicus. Words or sayings that are used in the court* [my translation] (Sahlstedt 1773: 16). The labels are not used in a systematic way, however, and cannot be of much help to the user.

The multiple function of the dictionary would demand more complex information about the Swedish headwords than that which Sahlstedt offers, since a foreign user could not be expected to know which of the Swedish lexical items correspond to the judicial meaning of the Latin equivalent. However, the

Latin index helps the user to solve that problem, since the Latin index and the headwords of the main part of the dictionary interact. The Swedish and Latin equivalents often make up the kind of synonyms that are used in the legal handbooks, thus showing that these Latin and Swedish legal lexical items must have been considered to be normal equivalents (or synonyms in a monolingual perspective) (see Table 3).

Out of the 40 lexical items in the sample, 32 are included in Sahlstedt's dictionary, and four of these are found in the index in their Latin form (*actor, contumax, dilatio, judex*). *Kärande* is found in the article of KÄRA, and the phrase *svara för rätta* is found in the article of SVARA, as is the nominalization *svarande*.

Homonyms are treated in separate articles, and the structure of the articles on a macro level is systematic. The microstructure is sometimes unclear, and it is not always obvious on what grounds Sahlstedt makes his subdivision of the senses. The monosemous articles have no sense indicators but all of the polysemous articles make use of the labels *Item.* and *Aliter.* as sense indicators. The semantic relations between the senses are difficult to understand (see Picture 2 with the example of *Fälla*).

Sahlstedt also uses Latin equivalents, Latin definitions, examples in Swedish and Latin, and word formation, both compounds and derivatives.

Picture 2: Section of article FÄLLA in Sahlstedt (1773)

Item: Fälla, ex. gr. Fälla någon, Aliquem
condemnare. Fälla dom, Sententiam ferre.
Fälla en på bötor, Multam alicui irroga-
re. Fälla djur i skogen, Animalia in silvis
necare. Fälla skog, Silvam cadere. Fälla
förbön, Precibus intercedere. Fälla priset
på en vara, Mercis pretium diminuere.
Fälla några ord, Verba sibi quædam elabi
permittere. Fälla tårar, Lacrymare. Fälla
träd, Arbores cadere.

In the example, the special judicial meaning is marked with the Latin label *Item.* The three first phrases in the section have a judicial sense: *Fälla någon* ['to convict someone'], *Fälla dom* ['to pass a sentence'], *Fälla en på bötor* ['to fine someone']. In *Fälla dom* the object *dom* ['verdict', 'sentence'] tells us that the verb should be interpreted as judicial. The same goes for *Fälla en på bötor* where the object *bötor* ['fines'] puts the verb in a legal context. For the example *Fälla någon*, you must know Latin to know that the sense is judicial for this verb.

The rest of the article has examples illustrating other senses of the polysemous verb, but there are no sense indicators that separate the senses. The article is made up from a mixture of several other senses of the verb *fälla*: *Fälla djur i skogen* ['to shoot animals in the woods']; *Fälla skog* ['fell a tree']; *Fälla förbön* ['offer up prayers']; *Fälla priset på en vara* ['to cut the price on a piece of

merchandise']; *Fälla några ord* ['say some words']; *Fälla tårar* ['shed tears']; *Fälla träd* ['fell a tree']. The phrases *Fälla skog* and *Fälla träd* illustrate the same sense of the verb: ['to cut down one or several trees']. The Latin equivalent to *fälla* (*cædere*) is also the same in both cases, so the only way in which the phrases differ is the way the Latin equivalents to the Swedish lexical items *skog* ['forest'] and *träd* ['tree'] differ from each other; this is information that might not be altogether relevant in this specific article.

Sahlstedt's relevant two subject field labels (*v. jur* and *term.jur*) are used in four of the articles in the sample: *åklagare* ('prosecutor') (monosemous) and the polysemous lexical items *kära* ('to accuse'), *part* ('part'), and *svara* ('to answer in court'). In the article of PART Sahlstedt also uses another subject field label that is not mentioned in the preface: *term. forens*. It is not at all clear how he uses the different notations or how he has chosen the lexical items that are labelled. It is important, however, to stress that Sahlstedt's main purpose with his dictionary was to bring more order into Swedish orthography and morphology; a commitment that was successful and resulted in a far more codified norm (Hannesdóttir 2000). Sahlstedt does not discuss semantics at any length in his preface and he probably did not intend to explain the meaning of the lexical items in detail. Still, Sahlstedt's mentioning of two different subject field labels denoting legal lexical items indicates that he was aware of the differences between general and technical lexical items and probably had the ambition of letting it show in his dictionary. However, the article of the verb FÄLLA does not give us any clues as to how Sahlstedt was going to carry out that idea.

A.F. Dalin, *Ordbok öfver svenska språket* ['Dictionary of Swedish'] (1850–55)

Dalin's dictionary is aimed at users with Swedish as their first language. Part of the long article FÄLLA is presented below.

Picture 3: Section of article FÄLLA in Dalin (1850–55)

marken. — 4) (fig.) a) Störta, bringa i förderf. Söka f. någon. — b) Göra lägre, nedsätta, afslå. F. priset på en vara. F. sina anspråk. — c) Döma. F. till ed, genom laga dom ålägga edgång. F. någon till plikt. — d) Döma skyldig. F. en anklagad. Det är bättre att fria än f., när den anklagade ej är fullt öfverbevist. — e) Förebringa eller innefatta öfverbevisning emot en anklagad, så att han kan dömas skyldig. Detta vittne, vittnesmål f-ller den anklagade. — f) Yttra, utsäga, afsäga, afkunna. F. ett ord, ett uttalande, ett yttrande. Han fällde några ord, som blefvo illa upplagna. F. hotelser. F. förböner för någon. F. dom, utslag. Han har fällt sin egen dom, dömt sig sjelf. F. omdöme, yttra sitt omdöme, säga sina tankar. F. sitt omdöme om någon, något, öfver något. — 5) (geom.)

Dalín treats homonyms in different articles. Polysemous headwords have subdivisions. The article of FÄLLA is divided into a total of six senses, each sense marked with a numeral, 1–6. The subsenses are marked with a long hyphen, and identified by small, Latin letters "a–f". Thereby, the subsenses are clearly separated from the main senses. In sense 4, six different subsenses are listed: (a) = ['bring sb. down']; (b) = ['lower, reduce']; (c) = ['judge']; (d) = ['convict']; (e) = ['Put forward or comprise conviction against the accused, in order to find him guilty']; (f) = ['pass (a word, opinion, verdict)']. The first two subsenses (a–b) are of general nature. The following three senses have a clear judicial meaning (c–e) and the last one, (f), can be used in a general as well as a judicial sense.

All senses are defined with (sometimes very) comprehensive definitions. The senses are exemplified with many phrases, some of which are fixed phrases or collocations: *F. omdöme* ['to give an opinion']; *F. dom, utslag* ['to pass a verdict'].

In the beginning of the article, Dalín makes use of the subject field label (fig.), ['figurative, metaphorical'], but he uses no labels to mark the judicial meaning of certain senses in the article. Probably, these are supposed to be explained by the formulation of the definitions. Dalín usually treats nominalizations in separate articles and there are no examples of word formations described in the article of FÄLLA.

Dalín presents a long list of approximately 60 diasystematic labels in his preface (pp. 19–20) but it is quite unclear on what grounds he uses them in the dictionary. Out of Dalín's 30 lexical items from the sample, only three have subject field labels, one of which is *contumacia*, and the other is *jäva*. *Contumacia* is marked with the label (lat.) for 'Latin', while *jäva* is marked with (Lagt.) for 'law term'. Dalín's use of subject field labels might be unclear but the semantic description has a systematic and consistent structure, although the definitions and examples are very verbose.

***Svensk ordbok* (2009); NSEOP (2018)**

Both SO (2009) and NSEOP (2018) meet all the requirements of modern dictionaries. In SO (2009), the senses are clearly marked with numerical sense indicators and supported by examples, illustrating constructions, collocations, fixed phrases etc. SO (2009), being a monolingual dictionary, has no reason to include subject field labels since the judicial sense can be understood by the definition. Still, subject field labels are sometimes used, and among the lexical items in this study *ogilla* is marked with <jur.>, maybe because the definition is somewhat insufficient: "*inte godkänna*" ['not approve of'].

The structure of the articles in the printed edition of NSEOP (2010) is clearly explained in the preface. The lemmas are provided with numeric sense indicators as well as Swedish explanations written in small type. The use of sense indicators is consistent, in accordance with the idea of the dictionary. Subject field labels are used for the "most important words" of different fields that are included, one of which is law, labelled as *jur.* (short for *juridik* ['law']). The labels are in Swedish and they are placed directly before the English

equivalent. There is no information about how to interpret "most important" or where the boundary between general words and terms is drawn.

It appears that most lexical items with judicial meaning are marked with a field label, except when the Swedish headword and English equivalent are both monosemous and have the same meaning, i.e. judicial. One example of this is *åklagare* ['prosecutor'].

5.3 Summary of the results

The lexicographical study shows, as expected, a clear development over time. The oldest dictionary, Serenius (1741), shows an awareness of the target users' needs in terms of isolated senses, usually marked with sense indicators, but the accomplishment of the lexicographical technique does not fully meet the ambitions of the strategy. The number of different senses is hard to discern and the examples and equivalents do not match altogether, which might be a result of a troublesome method when compiling the dictionary. The judicial sense is clearly expressed however and emphasized with several examples. The lexicographical design is focused on the target users' needs, which meets the demands of the typology.

The second oldest Swedish dictionary, Sahlstedt (1773), has definitions in Latin, but a structure that is more similar to a monolingual dictionary than a bilingual. Since the dictionary aims at two quite different groups of target users representing both L1 and L2-speakers, the mission of fulfilling the needs of both groups seems difficult to achieve without a more refined structure. Sahlstedt's main purpose is to try to codify Swedish morphology and orthography, and his dictionary is best used together with his grammar, published in 1769. The structuring of the semantics in the article studied shows many inconsistencies, and it is obvious that the reader must know Latin to fully understand the structure of the article and the senses of the lexical items. This is something of a paradox considering that the description and codification of Swedish were the main purpose of the dictionary. The judicial sense is easy to discern but it is not marked in any special way compared to the other senses.

Dalin's dictionary (1850–55) was published less than one hundred years after Sahlstedt but it reveals a consistent lexicographical structure with detailed semantic descriptions and elaborate definitions. Although Dalin lists judicial sense indicators in his preface, he does not make use of them in the article FÄLLA. Nevertheless, the judicial senses of FÄLLA studied are clearly identified with several examples to support them.

The modern dictionaries (SO 2009 and NSEOP 2018) both show a lexicographical structure in accordance to modern, lexicographical standards.

6. Final conclusions

The subject of this article has been to study in what way three influential 18th

and 19th century dictionaries treat a sample of 40 legal lexical items of Swedish and Latin origin in their dictionaries. In the 18th century, the usage of Latin as opposed to Swedish was intensely discussed, especially in science and the fine arts. At the same time, the period also saw a rapidly growing development of lexicography and the role of lexicography in relation to questions of language planning; this issue is interesting to examine in more detail. A material extracted from the two first legal handbooks in Swedish has been used: the handbooks were published by Rålamb (1674) and Kloot (1676), independently of each other.

The study is based on two specific research questions. The first question concerns the inclusion of Latin and Swedish legal lexical items from the handbooks in the lemma lists of the dictionaries, the second question regards the lexicographical treatment of the lexical items of the sample, especially the division into senses and whether the judicial meaning is clearly recognized or not.

Both Serenius and Sahlstedt, the two 18th century lexicographers, prefer the Swedish equivalents to the Latin, except when the Latin lexical items seem to be already well integrated into Swedish. Their infrequent usage of Latin legal lexical items in the dictionaries seems to be the result of an active decision, since the Latin legal lexical items were probably in use during the whole period examined. One proof of this is their presence in the later dictionaries included in the study, where several of the Latin lexical items are still in use but have been subject to a process of determinologization. Serenius's and Sahlstedt's treatment of the Latin judicial lexical items seems to indicate a strive to reduce the usage of Latin in favour of Swedish, thus trying to follow the direction of Swedish language planning of the time (cf. Telemann 2002: 59, Rogström 2017). They both had an outspoken interest in linguistic standardization, and there is reason to believe that they had an interest in taking active part in the language planning process.

The second research question focuses on the lexicographical treatment of the lexical items in the sample. The study shows that both Serenius and Sahlstedt are aware of the need to structure the article in accordance with the different senses of the headwords, but their lexicographical skills are far from perfect and leave much to be desired. The microstructure of the articles does not follow a set order, and the user cannot rely on the structure of the articles in order to discern the different senses. The most certain way to find a specific piece of lexical information is probably to read the whole article from the beginning to the end. However, in both Serenius's and Sahlstedt's dictionaries, the judicial meaning of the lexical items from the sample are all identified and described through equivalents and illustrated with examples; this indicates that both lexicographers were aware of the importance of separating different senses from each other. Neither of the lexicographers masters the special requirements for bilingual dictionaries, especially not Sahlstedt who is aiming at a bidirectional dictionary for both L1-speakers of Swedish as well as learners.

In Dalin's dictionary (1850–55), the lexicographic development is fully

established, and his articles are systematically written with clear distinctions between homonyms and related senses of polysemous lexical items within the main sections of the article. The two contemporary dictionaries SO (2009) and NSEOP (2018) are also systematically compiled, and the division into senses clearly identifies the judicial senses that are easy to find and make use of. The only thing still lacking, is the treatment of subject field labels (cf. Vrbinc and Vrbinc 2013). NSEOP (2018) seems to use them in a consistent way, but neither Dalin (1850–55) nor SO (2009) uses their labels in a clear way.

To conclude — the study confirms the assumption that Swedish 18th century dictionaries could be expected to support certain lexical language planning strategies at the time. Both Serenius and Sahlstedt seem to be loyal to the effort of promoting Swedish in favour of Latin during the 18th century. Their selection of headwords indicates a preference for the Swedish vocabulary, and their lexicographical treatment of the chosen headwords also shows an identification of legal lexical items and an awareness of the special requirements needed for the presentation of the judicial meaning of the lexical items. But most importantly, the results show that more detailed studies of the dependence between old dictionaries and early lexical language planning strategies probably could shed more light on a part of Swedish language history that is still waiting to be elucidated.

References

Dictionaries

Dalin, A.F. 1850–55. *Ordbok öfver svenska språket* I–II. Stockholm: Joh. Beckman.

NSEOP = *Norstedts svensk-engelska ordbok Professionell*. 2010. Stockholm: Norstedts Förlag AB.

NSEOP = *Norstedts svensk-engelska ordbok Professionell*. 2018. Stockholm: Norstedts Förlag AB.
<<https://www-ne-se.ezproxy.ub.gu.se/ordböcker/#/search/norstedts-pro-sv-en?q=dunkel>>

Sahlstedt, A. 1773. *Svensk ordbok*. Stockholm: Carl Stolpe.

SAOB = *Ordbok öfver svenska språket, utgiven av Svenska Akademien*. 1893–. Lund: Gleerups.

Serenius, J. 1734. *Dictionarium Anglo-Suethicum-Latinum*. Hamburg: Rudolphum Beneken.

Serenius, J. 1741. *Dictionarium Suethico-Anglo-Latinum*. Stockholm: Pet. Momma.

SO = *Svensk ordbok utgiven av Svenska Akademien* 2009. Stockholm: Svenska Akademien & Norstedts.
<www.svenska.se>

Other literature

Becker, H. 2015. Scientific and Technical Dictionaries; Coverage of Scientific and Technical Terms in General Dictionaries. Durkin, Ph. 2015. *The Oxford Handbook of Lexicography*: 393–408. Oxford: Oxford University Press.

Bergenholtz, H. and R.H. Gouws. 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13–45.

- Björne, L.** 1995. *Patrioter och institutionalister. Den nordiska rättsvetenskapens historia. Del I. Tiden före 1815.* Rättshistoriskt bibliotek 52. Skrifter utgivna av Institutet för rättshistorisk forskning. Grundat av Gustav och Carin Olin. Serien I. Stockholm/Lund: Institutet för rättshistorisk forskning.
- Bowker, L.** 2018. Lexicography and Terminology. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 138-151. London/New York: Routledge.
- Faber, P.** 2012. *A Cognitive Linguistics View of Terminology and Specialized Language.* Berlin/Boston: Mouton de Gruyter.
- Fontenelle, T.** 2014. From Lexicography to Terminology: A Cline, not a Dichotomy. Abel, A., C. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, Bolzano/Bozen, Italy, 15–19 July 2014*: 25-45. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Geeraerts, D.** 2015. Lexicography and Theories of Lexical Semantics. Durkin, Ph. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 425-438. Oxford: Oxford University Press.
- Gunnarsson, B.-L. (Ed.).** 2011. *Languages of Science in the Eighteenth Century.* Berlin/Boston: Walter de Gruyter.
- Hannesdóttir, A.H.** 1998. *Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. Meijerbergs arkiv för svensk ordforskning* 23. Gothenburg: University of Gothenburg.
- Hannesdóttir, A.H.** 2000. Tvåspråkig lexikografi och språkplanering i 1700-talets Sverige. Edlund, L.-E. (Ed.). 2000. *Studier i svensk språkhistoria* 5. *Förhandlingar vid Femte sammankomsten för svenska språkets historia Umeå 20–22 november 1997*: 224-235. Nordsvenska 11. Umeå: University of Umeå.
- Hannesdóttir, A.H.** 2011. From Vernacular to National Language: Language Planning and the Discourse of Science in Eighteenth-century Sweden. Gunnarsson, B.-L. (Ed.). 2011. *Languages of Science in the Eighteenth Century*: 107-122. Berlin/Boston: Walter de Gruyter.
- Hannesdóttir, A.H.** 2014. Lemman och ekvivalenter i nya roller — en reviderad ordbokstypologi. Vatvedt Fjeld, R. and M. Hovednak. (Eds.). 2014. *Nordiske studier i leksikografi* 12. *Rapport fra Konferanse om leksikografi i Norden. Oslo 13.–16 August 2013*: 193-211. Oslo: Skrifter utgitt av Nordisk forening for leksikografi.
- Hannesdóttir, A. and B. Ralph.** 1988. Early Dictionaries in Sweden: Tradition and Influences. Hyldgaard-Jensen, K and A. Zettersten (Eds.). 1988. *Symposium on Lexicography IV. Proceedings of the Fourth International Symposium on Lexicography April 20–22, 1988 at the University of Copenhagen.* Lexicographica Series Maior 26: 265-279. Tübingen: Niemeyer.
- Haugen, E.** 1984. *A Bibliography of Scandinavian Dictionaries.* White Plains NY: Kraus International Publications.
- Haugen, E.** 1985. Lexicography and Language Planning. Jankowsky, K.R. (Ed.). 1985. *Scientific and Humanistic Dimensions of Language. Festschrift for Robert Lado on the Occasion of his 70th Birthday on May 31, 1985*: 571-580. Amsterdam: John Benjamins.
- Haugen, E.** 1987. *Blessings of Babel. Bilingualism and Language Planning. Problems and Pleasures.* Contributions to the Sociology of Language 46. Berlin/New York/Amsterdam: Mouton de Gruyter.
- Hoare, M.R.** 2009. Scientific and Technical Dictionaries. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography. Vol. II. Specialized Dictionaries*: 47-93. Oxford: Clarendon Press.

- Holm, L. and H. Jonsson.** 1991. Swedish Lexicography. Hausmann, F.J, O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1991. *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie.* Handbücher zur Sprach- und Kommunikationswissenschaft 5(2): 1933-1943. Berlin/New York: De Gruyter.
- Holmer, L.** 2016. *Grammatik i SAOL. En undersökning av grammatisk information i Svenska Akademiens ordlista över 130 år.* Meddelanden från institutionen för svenska språket 65. Institutionen för svenska språket. Gothenburg: University of Gothenburg.
- Jacoby, M.** 1990. *Historische Lexicologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin.* Wiesbaden: Otto Harrassowitz.
- Johansson, M.** 1997. *Lexicon Lincopense. En studie i lexikografisk tradition och svenskt språk vid 1600-talets mitt.* Meijerbergs arkiv för svensk ordforskning 21. Gothenburg: University of Gothenburg.
- Josephson, O.** 2018. *Språkpolitik. Språkrådets skrifter* 25. Falun: Morfem.
- Kloot, C.** 1676. *Then Swenska Lagfarenheetz Spegel [...].* Gothenburg: Amund Grefwe.
- Korpiola, M.** 2014. General Background: From Judicial Crisis to Legal Revolution through the Svea Court of Appeal? Korpiola, M. (Ed.). 2014. *The Svea Court of Appeal in the Early Modern Period: Historical Reinterpretations and New Perspectives.* Rättshistoriska studier 26: 23-54. Stockholm: Institutet för rättshistorisk forskning/Jure.
- Larsson, L.** 2003. *Varifrån kom svenskan? Om den svenska vokabulären i en fyrspråkig ordbok utgiven i Riga 1705.* Acta Academia Regiæ Gustavi Adolphi 81. Uppsala: Kungl. Gustav Adolfs Akademien för svensk folkkultur.
- Laurén, C., J. Myking and H. Picht.** 1997. *Terminologi som vetenskapsgren.* Lund: Studentlitteratur.
- Malmgren, S.-G.** 1988. Almqvist, Dalin och den svenska definitionsordbokens födelse. Pettersson, G. (Ed.). 1988. *Studier i svensk språkhistoria.* Lundastudier i nordisk språkvetenskap A 41: 195-213. Lund: Lund University Press.
- Malmgren, S.-G.** 2000. Projektet Det svenska ordförrådets utveckling 1800–2000: utgångspunkter. <<https://svenska.gu.se/publikationer/ordat>>
- Malmgren, S.-G. and E. Sköldbberg.** 2013. The Lexicography of Swedish and Other Scandinavian Languages. *International Journal of Lexicography* 26(2): 117-134.
- Mattila, H.E.S.** 2013 [2007]. *Comparative Legal Linguistics. Language of Law, Latin and Modern Lingua Francas.* Second edition. Farnham, Surrey, UK/Burlington, VT: Ashgate.
- Modéer, K.Å.** 2014. Mixed Legal Systems in Early Modern Sweden: Judicial Traditions and Reforms in an Expanding European Great Power — Some Concluding Remarks. Korpiola, M. (Ed.). 2014. *The Svea Court of Appeal in the Early Modern Period: Historical Reinterpretations and New Perspectives.* Rättshistoriska studier 26: 395-422. Stockholm: Institutet för rättshistorisk forskning/Jure.
- Nielsen, S.** 2015. Legal Terms in General Dictionaries of English: The Civil Procedure Mystery. *Lexikos* 25: 246-261.
- Nkomo, D.** 2018. Dictionaries and Language Policy. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography:* 152-165. London/New York: Routledge.
- Norén, K.** 1991. Utformningen av artiklarna Dalins Ordbok öfver svenska språket (1850–55). *Språk och Stil. Tidskrift för svensk språkforskning.* NF 1: 109-138.

- Olofsson, A.** 1991. Bilingual Lexicography Swedish–English, English–Swedish. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1991. *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie.* Handbücher zur Sprach- und Kommunikationswissenschaft 5(2): 3047-3051. Berlin/New York: De Gruyter.
- Ralph, B.** 2012. Biskopar, ordböcker och verkligheten — vad är lexikografihistoria bra för? Eaker, B., L. Larsson and A. Mattisson (Eds.). 2012. *Nordiska studier i lexikografi 11. Rapport från Konferensen om lexikografi i Norden. Lund 24–27 maj 2011:* 30-54. Lund: Skrifter utgivna av Nordiska föreningen för lexikografi.
- Rogström, L.** 1994. Nya tider, nya seder? Om ortografiska principer i Jacob Serenius' ordböcker. *Språk och Stil. Tidskrift för svensk språkforskning.* NF 4: 65-97.
- Rogström, L.** 1998. *Jacob Serenius lexikografiska insats. Meijerbergs arkiv för svensk ordforskning* 22. Gothenburg: University of Gothenburg.
- Rogström, L.** 2009. Språkplanering i praktiken? En lexikalisk studie i förarbetena till 1734 års lag. *Språk och Stil Tidskrift för svensk språkforskning.* NF 19: 5-34.
- Rogström, L.** 2010. I akt och mening. En jämförelse av ordförrådet i 1734 års lag med samtida ordböcker. Reinhammar, M. (Ed.). 2010. *Studier i svenska språkets historia 11. Förhandlingar vid Elfte sammankomsten för svenska språkets historia i Uppsala 23–24 april 2010:* 189-199. Acta Academiae Regiae Gustavi Adolphi 113. Uppsala: Kungl. Gustav Adolfs Akademien för svensk folkkultur.
- Rogström, L.** 2017 "Om våra swänske ord komma i bruk, kunna de såwäl som fremmande förstås". En studie av lexikalisk förändring i juridiskt språk på 1600-talet. *Språk och Stil. Tidskrift för svensk språkforskning.* NF 27: 67-95.
- Rogström, L. and H. Landqvist.** 2015. Ett tidigt försök att etablera en svensk juridisk vokabulär? Lexikaliska och fackspråkliga perspektiv på en juridisk handbok från 1674. Rellstab, D. and N. Siponkoski (Eds.). 2015. *Rajojen dynamiikkaa, Gränsernas dynamik, Borders under Negotiation, Grenzen und ihre Dynamik. VAKKI-symposiumi XXXV. Vaasa 12–13.2.2015:* 199-208. VAKKI Publications 4. Vaasa: Vaasan Yliopisto — University of Vaasa.
- Rogström, L. and H. Landqvist.** 2018a. *A Terminologia Latina till Svenska Ord.* En studie av latinska och svenska lexikala enheter i två juridiska handböcker från 1670-talet. Lönnroth, H., B. Haagenzen, M. Kvist and K. Sandvad West (Eds.). 2018. *Studier i svensk språkhistoria* 14: 227-243. Vasa: Vaasan Yliopisto — University of Vaasa.
- Rogström, L. and H. Landqvist.** 2018b. Introducing a Local Legal Vocabulary in a Latin Context. A Study of Two Swedish 17th Century Approaches. *Fachsprache. Journal of Professional and Scientific Communication* 40(3–4): 161-178.
- Rålamb, C.** 1674. *OBSERVATIONES JURIS PRACTICÆ. Thet är / Åthskillige Påminnelser othi Rättegångs Saker [...].* Stockholm: Niclas Wankijf.
- Sahlstedt, A.** 1769. *Svensk Grammatika efter det nu för tiden i språket brukliga sättet författad.* Uppsala: Johan Edman.
- Sjögren, W.** 1900–1909. *Föarbetena till Sveriges rikets lag 1686–1736. Efter offentligt uppdrag utgifna af Wilhelm Sjögren.* Uppsala: Almqvist & Wiksells boktryckeri-aktiebolag.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.
- Teleman, U.** 2002. *Ära, rikedom och reda. Svensk språkvård och språkpolitik under äldre nyare tid.* Skrifter utgivna av Svenska språknämnden 85. Stockholm: Norstedts Ordbok.

- Teleman, U.** 2003. *Tradis och Funkis. Svensk språkvård och språkpolitik efter 1800*. Skrifter utgivna av Svenska språknämnden 87. Stockholm: Norstedts Ordbok.
- Teleman, U.** 2005. The Role of Language Cultivators and Grammarians for the Nordic Linguistic Development in the 16th, 17th and 18th Centuries. Bandle, O. et al. (Eds.). 2005. *The Nordic Languages. An International Handbook of the History of the North Germanic Languages. Volume 2*. Handbücher zur Sprach- und Kommunikationswissenschaft 22(2): 1379-1396. Berlin/New York: Mouton de Gruyter.
- Tiisala, S.** 2018. Subscribers and Patrons: Jacob Serenius and his Dictionarium Anglo-Svethico-Latinum 1734. McConchie, R. and J. Tyrkkö (Eds.). 2018. *Historical Dictionaries in their Paratextual Context*. Lexicographica. Series Major 153: 237-260. Berlin/Boston: De Gruyter.
- Westman, K.G.** 1912. *De svenska rättskällornas historia: föreläsningar*. Uppsala: Appelberg.
- Vrbinc, M. and A. Vrbinc.** 2013. Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8. *Lexikos* 23: 440-455.

A Critical Evaluation of Three Sesotho Dictionaries¹

Mmasibidi Setaka, *South African Centre for Digital Language Resources, South Africa and University of Pretoria, Pretoria, South Africa*
(mmasibidi.setaka@nwu.ac.za)

and

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, South Africa* (danie.prinsloo@up.ac.za)

Abstract: This article gives a perspective on Sesotho lexicography and a critical analysis of the macrostructures and microstructures of three selected Sesotho dictionaries. The monolingual paper dictionary *Sethantšo sa Sesotho*, the bilingual paper dictionary *Southern Sotho–English Dictionary* and the *Sesotho online Bukantswe v.3* are evaluated. Their virtues and shortcomings as reference works will be viewed against dictionaries of high lexicographic achievement in order to establish to what extent they fulfil the most basic requirements of macrostructures and microstructures. The inconsistencies addressed in this article reflect the need for Sesotho lexicographers to use corpora in dictionary compilation in order to enhance the quality of entries on both microstructural and macrostructural levels. It will be argued that much more research and description of lexicographic issues is required to bring Sesotho lexicography on a par with its sister languages, Sepedi and Setswana and with good dictionaries for major languages of the world. After decades in existence, currently available Sesotho dictionaries are in dire need for revision and new dictionaries aimed at specific target users should be compiled.

Keywords: LEXICOGRAPHY, DICTIONARIES, SESOTHO, CORPORA, LEMMATISATION, LEMMA TREATMENT, CONCORDANCES, MICROSTRUCTURE, MACROSTRUCTURE

Opsomming: 'n Kritiese evaluering van drie Sesotho woordeboeke. Hierdie artikel gee 'n perspektief op Sesotho-leksikografie en 'n kritiese ontleding van die makrostrukture en mikrostrukture van drie geselekteerde Sesotho woordeboeke. Die eentalige papierwoordeboek *Sethantšo sa Sesotho*, die tweetalige papierwoordeboek *Southern Sotho–English Dictionary* en die *Sesotho online Bukantswe v.3* word geëvalueer. Hulle deugde en tekortkominge as naslaanwerke sal beskou word teenoor woordeboeke van hoë leksikografiese gehalte om vas te stel in watter mate hulle aan die mees basiese vereistes van makrostrukture en mikrostrukture voldoen. Die teenstrydighede wat in hierdie artikel aangespreek word, weerspieël die noodsaaklikheid dat Sesotho leksikograwe korpora in woordeboeksamestelling gebruik om die gehalte van inskrywings op mikrostrukturele sowel as makrostrukturele vlak te verhoog. Daar sal geargumenteer word dat baie meer navorsing en beskrywing van leksikografiese kwessies nodig is om die leksikografie van Sesotho op gelyke voet te bring met die sustertale Sepedi en Setswana asook met goeie woordeboeke van wêreldtale. Na dekades van gebruik, moet die Sesotho woordeboeke wat tans beskikbaar is dringend hersien word en nuwe woordeboeke saamgestel word wat op spesifieke teikengebruikers gerig is.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEKE, SESOTHO, KORPORA, LEMMATISERING, LEMMABEWERKING, KONKORDANSIES, MIKROSTRUKTUUR, MAKROSTRUKTUUR

1. Introduction

Sesotho lexicography receives very little attention in the literature, compared to e.g. its sister languages Sepedi and Setswana. For Sepedi, numerous studies have been done on problematic macrostructural and microstructural aspects such as lemma selection, treatment of lemmas and the utilisation of electronic corpora to enhance lexicographic quality. On the macrostructural level, most modern Sepedi and Setswana dictionaries utilise frequency counts from corpora as an aid to decide on inclusion or omission of lemmas for newly compiled or revised dictionaries. On the microstructural level, concordance lines culled from corpora contribute to quality enhancement in the writing of definitions, selection of translation equivalents, selection of examples, etc. De Schryver and Prinsloo (2000a) give a detailed discussion of the shortcomings in African language dictionaries on the macrostructural level due to inadequate lemma offerings, mostly as a result of including lemmas in the dictionary "as they cross the compiler's mind" rather than by means of a specific selection strategy such as frequency lists from corpora. African language dictionaries also do not perform well at the microstructural level due to inadequate treatment of the lemmas, as will be shown in more detail below. See also De Schryver and Prinsloo (2000b), Gouws and Prinsloo (2005) and Otlogetswe (2009a and b, 2012, 2013), for detailed discussions for Sepedi and Setswana. With the exception of Moleleki (1999), Prinsloo (2013), and Motjope-Mokhali (2016), no in depth lexicographic research has been recorded for Sesotho. A work of merit is Motjope-Mokhali's (2016) critical comparison of the *Sesuto-English dictionary* and *Sethantšo sa Sesotho* with reference to lexical entries and dictionary design.

In this article we will focus on Sesotho lexicography and will explore and critically analyse one monolingual dictionary, *Sethantšo sa Sesotho* (henceforth referred to as *Sethantšo*), one bilingual, *Southern Sotho-English Dictionary*, (*SSED*) and an electronic dictionary, *Sesotho Online Bukantswe v.3* (*Bukantswe*) in terms of their macrostructural and microstructural characteristics. Benchmarking the quality of these Sesotho dictionaries will be done against dictionaries of high lexicographic achievement on the basis of a number of basic requirements of macrostructures and microstructures.

2. Criteria for the evaluation of Sesotho dictionaries

Prinsloo and Taljard (2017: 428-430) give a detailed discussion of the problematic aspects of evaluation of macrostructures and microstructures of dictionaries for African languages. They state that any comparison between African language dictionaries with dictionaries deeply rooted in a long and rich lexicographical tradition is somewhat unfair. Ideally, specific types of dictionaries with narrowly defined target users should be compared with each other in terms of the specific quality criteria applicable to that specific type of dictionary. So, for example, a Sesotho monolingual dictionary for advanced learners should be compared with English monolingual dictionaries for advanced learners. None of the three Sesotho dictionaries evaluated in this article, however, has narrowly defined target users. They are all aimed at undefined target users and have to serve all possible user profiles. The researcher has no other option than to revert to an evaluation based upon the most basic criteria for the judgment of lexicographic quality of Sesotho dictionaries. Gouws (1990: 52) emphasizes that good dictionaries, as containers of knowledge, are characterised by their offering of a variety of information types. *Macmillan Dictionary (MD)* gives a concise summary of such basic requirements of a dictionary:

A dictionary is a description of the vocabulary of a language. It explains what words mean, and shows how they work together to form sentences.
<http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/>

Gouws and Prinsloo (2005: 144) say that "the word must be defined in such a way that the users will get all the answers to the questions that made him or her consult the dictionary".

Finding the word he/she is looking for in the dictionary and finding sufficient information about the meaning and use of the word relates to the compilation of the lemma list (as part of the macrostructure) and treatment of the lemma (data types in the microstructure).

3. A survey of Sesotho dictionaries

Sesotho dictionaries known to the authors are listed with brief descriptions of author, type and size, where available, in table 1. This survey of dictionaries is focused on general dictionaries which according to Nkomo (2010: 372) "... have a very important role to play in the development, acquisition and use of indigenous African languages". The sizes and scope of these dictionaries are different, as some are mere word lists while others provide more comprehensive treatment of the lemma.

Table 1: Sesotho dictionaries

Name	Date of publication First published/ latest revision	Author	Type and size
Paper dictionaries			
English–Southern Sesotho dictionary	2015	Motsapi J.	Mono-directional bilingual 408 pages
English–Southern Sesotho dictionary	1965	Hamel, L. (OMI)	Bilingual. 6 volumes
English–Sesotho vocabulary	1905	A. Casalis	Bilingual: English lemma list with basic Sesotho translations
English–Sotho, Sotho–English pocket dictionary	1960	S. Christeller	Bilingual Pocket Dictionary 144 pages
English–Se-Suto vocabulary	(Unknown)	Ellenburger, D.F.	(Unknown)
Sethantšo sa Sesotho	2005	Hlalele, B.	Monolingual 325 pages
Sesuto–English dictionary	1937	Mabille, A. and Dieterlen, H.	Bilingual 445 pages
Bukantswe ya maleme-pedi, Sesotho–Seafrikanse	1974	J.A. du Plessis, J.G. Gildenhuis and J.J. Moiloa	269 pages
English–Sotho, Sotho–English pocket dictionary.	1960	(Unknown)	Bilingual
Sehlahosi: Sesotho cultural dictionary	1994	Matšela, F.Z.A.	74 pages
Southern Sotho–English dictionary	1988	Paroz, R.A.	598 pages
Khetsi ea Sesotho	1997	Pitso, T.T.E.	Monolingual 323 pages
Se-Sotho–English vocabulary	(Unknown)	Verdier, T.	(Unknown)
New South Sotho dictionary	1997	Chaphole, S.R.	Bilingual 103 pages
Oxford First Bilingual dictionary	2007	Paizee, D.	Bilingual 64 pages
English Sesotho Official Foundation Phase CAPS Picture Dictionary	2017	Sesotho National Lexicography unit	Bilingual dictionary 148 pages
Online Sesotho dictionaries			
Bukantswe http://bukantswe.sesotho.org/		Olivier J.A.K.	Bilingual dictionary 10,075 entries
Sesotho dictionary-Bilingo https://www.bilingo.co.za/sesotho-dictionary-2/			(Not available at time of consultation (9 Jan 2020))
Free English–Sesotho dictionary and translator-FREELANG https://www.freelang.net/dictionary/sesotho.php			Bilingual Sesotho > English: 9,980 words, English > Sesotho: 6,638 words
English–Southern Sotho Dictionary, Glosbe https://glosbe.com/en/st			Bilingual
Sesotho–English Dictionary — Apps on Google Play https://play.google.com/store/apps/details?id=alldictdict.alldict.sten&hl=en_US			Bilingual

4. Lemma selection and treatment

In this section the aim is to evaluate the macrostructures and microstructures of *Sethantšo*, *SSED* and *Bukantswe*. The lemma lists and the treatment of lemmas in these dictionaries were evaluated based on introspection as well as through comparison with Sesotho corpora. The focus will be on the merits and contributions made to the knowledge of Sesotho, but a number of shortcomings on macrostructural and microstructural level will also be highlighted and briefly discussed. These presumed shortcomings revolve around (a) insufficient basic information in respect of meaning and translation equivalents, (b) alphabetical ordering, (c) morphological information, (d) pronunciation guidance, (e) examples of usage, (f) inconsistencies in the presentation and treatment of lemmas and (g) inadequate search functions.

4.1 *Sethantšo*

This is a monolingual dictionary written in Lesotho Sesotho orthography and first published in 2005. It contains approximately 10,000 lemmas and the typical information given in the articles of each lemma includes paraphrase of meaning (a definition), part of speech, noun class indication and prefix of the plural form in the case of nouns, past tense derivation in the case of verbs and etymology as in figure 1. Basic orthographic differences between South African and Lesotho orthographies are SA *di* > Lesotho *li*, *kg* > *kh*, *tjh* > *ch*, *ff* > *fsh*, etc. See https://en.wikipedia.org/wiki/Sesotho_orthography for a detailed discussion and typical examples.

kalatis.a(.itse) /*kutu-ketso*/ ho pepesa makhabane kapa seo motho a se fihletseng 'me a ithorisa ka tsona; *ho ikalatisa*: ho ipolelisa; ho ipatla ka ho ipolela le ho ipolelisa makhabane le liketso tsa bohaka; ho se ikopanye le ba bang ka hore ba tlase, ka ho iphahamisa.
kale /*kutu-tlhakiso*/ e boholo bo bokae? e kholo bo bokana.
kaliana(li.) /*lereho 9*/ lijo tse jeoang mohla bale ba tsoang ba tebukang; lijo tsa bohlokoa tsa mekete e khethehileng.
kalim.a (.ile & .me) /*kutu-ketso*/ ho lumella tšebeliso ea seng sa hao se tla khutlisoa; ho nehana ka seng sa hao se sebelisoa ke e mong nakoana e itseng se tla khutlela ho uena ha tšebeliso eo e felile.

kamo /*sere*/ le supa ho robeha ha ntho e khaohang ho sa lebelloa, haholo e entsoeng ka letsopa kapa ka ho hong ha e le motopo.
kamoh.a(.ile) /*kutu-ketso*/ ho robeha ka boeona ho sa lebelloa; ho pomeha ha ntho ka boeona. (<kamo)
kamol.a(.tse) /*kutu-ketso*/ ho poma ntho e motopo; ho khaola e entsoeng ka letsopa kapa ho roba sejo kotoana se itseng. (<kamo)
kamore(li.) /*lereho 9*/ phapusi ea ntlo; karolo e khaotsoeng ka lerako kahar'a ntlo. (<A)
kampel.a(.tse) /*kutu-ketso*/ ho koalla ka motero; ho thibeletsa ka motero k.h.r. terata. (<kampo: A & E)
kampo /*sehokelo*/ kapa; afa; empa.

Figure 1: An extract from *Sethantšo*

In the treatment of *kamore* 'room' the plural form is indicated in brackets as the final part of the lemma as (*li.*), i.e. the plural form is *likamore* 'rooms'. This is followed by part of speech indication given in italics between forward slashes indicating that it is a noun from class 9 followed by a definition and that it is borrowed from Afrikaans. The articles in this dictionary are relatively short — on average 32 articles per page in double columns. No examples of usage, collocations, pronunciation, etc. are given. In particular, examples of usage could be valuable in illustrating the different types of rooms and related terms, as has been successfully done in the *English–Sesotho Dictionary (ESD)* for the lemma *room* in figure 2.

room *n.* (~ *in a building*) kamore, phaposi;
(sitting ~ in a house) foreisi, kamore ya
 boikgathollo, phaposi ya boikgathollo;
(space) sebaka; *(waiting ~)* kamohelong,
 kotoise

Figure 2: 'Room' in the *English–Sesotho Dictionary*

The *English–Sesotho Dictionary* did well in giving a clear scope of different types of rooms by means of labelling them in brackets but still being economical, using only five lines of dictionary column space. Adding examples of use would assist users in text production and should be considered for future revisions of this dictionary. Consider the corpus lines for *kamore* in table 2.

Table 2: Corpus lines for *kamore* 'room' in Sesotho

<i>Tjhopho o ile a re 59</i> 'Tjhopho said. 59'	<i>kamore</i> [... room]	<i>ya ho robala. Karabo: Tjho</i> 'the bedroom. Karabo: Tjho'
<i>sona, o ne o ka tiisa hore</i> 'it, you would be certain that'	<i>kamore</i> [... room]	<i>ya habo ya ho phomola ke s</i> 'the sittingroom is s'
<i>Ka ho ya kaBarbara,</i> 'According to Barbara,'	<i>kamore</i> [... room]	<i>ya bonaya ho robala, ke mo</i> 'their 'the bedroom, is here'
<i>e tsohle tsa bona di na le</i> 'all of them have'	<i>kamore</i> [... room]	<i>e ikgethang ya ho hlapela</i> 'a special bathroom'

The concordance lines in table 2 show the different types of rooms which ought to be put in the article to, help the user to understand the meaning of the word, its range of application and the opportunity to extract authentic examples of usage from the corpus as emphasized by De Schryver and Prinsloo (2000b). Such examples bring the meaning of different types of rooms to the fore and can be regarded as a natural extension of the definition. Future compilers of

Sesotho dictionaries are advised to consult Sesotho corpora in the compilation of dictionaries to enhance the quality of their micro- and macrostructural compilations. Consider the value added by an example of usage of *room* in the *Oxford Bilingual School Dictionary: Northern Sotho and English (ONSD)* in figure 3.

room *** *noun* (pl. **rooms**) ■ phapoši ♦ Paledi's house has four **rooms**. *Ntlo ya Paledi e na le diphapoši tše nne.*

Figure 3: Phapoši 'room' in ONSD

In figure 3 the three-star markup (***) is a valuable indication to the user that *room* is a highly frequent word and such markup is often wisely perceived by learners as an implicit recommendation to not only find its meaning but also to learn such a word to extend their vocabulary of the language. The structural markers ■ and ♦ help to demarcate the different information categories, translation equivalents and examples, respectively. These markers also contribute to a user-friendly layout and are appealing to the eye.

Collocations could economically be indicated as part of examples to give an indication of words which more often than chance predicts co-occur with *kamore* 'room', thus giving users a clearer picture of its meaning.

One has to keep in mind that space is limited in single-volume paper dictionaries. Compilers often have to strike a balance between the number of entries that can be accommodated versus the exhaustiveness of the treatment. So, for example, Prinsloo (2009: 162) says the compiler is caught up in a triangulation of number of lemmas versus exhaustiveness of treatment versus price.

In principle, these limitations leave the compiler with two basic options: the inclusion of a large number (e.g., 20,000–30,000) of lemmas with limited (e.g., 1–2 lines double column) treatment, or a limited number (e.g., 10,000) of lemmas with more exhaustive (e.g., 5–7 line) treatment.

This is also true for Sesotho dictionaries.

Sesotho monolinguals should nevertheless strive towards giving a more detailed treatment of lemmas in order to meet the basic requirements as stated above in terms of MD and Gouws and Prinsloo (2005).

Consider example (1) as an attempt at a model entry for *mala* 'entrails'.

(1)

mala¹ (ma-la) 5/6 [*mala*]

Setho sa mmele se fumanwang ka mpeng ya motho kapa ka mpeng ya phoofolo se jarang dijo, ho di tsamaisa le ho di kenya maro a itseng a tsoang mabopong a ona. *Dikgoho di na le mala a masesane. Batho bohle ba na le mala. Ke ja mala le mohodu. Mala a kgomo a maholo.*

mala²

Lefu le tshwarang motho ha a dubehile ka mpeng, a jele dijo tse senyehileng kapa dijo di sa dula hantle ka mpeng. *Ntate o tshwerwe ke mala. Mala a mmangwane a bohloko, ebile o a tsholla. Nthabiseng o sebedisitse mala. Ke tshwerwe ke mala. Mme o mathiswa ke mala.*

mala³

Lela la pene. *Pene ya Tshepo e na lela le le lelele. Dipene di na le mala a fapaneng. Enke e ka hara lela lena e omme.*

In example (1) the three homonym distinctions are separated and indicated by superscript numbers. Syllable division, noun class singular/plural is given as well as a specific class indication, class 6 in this case, and a phonetic transcription. The class and class pair indication is done in a very economical way showing the relation 5/6 with the relevant one in this entry given in boldface.

Consider also the treatment of *mala* in figure 4, in the paper version of Sepedi, a sister language of Sesotho in *Pukuntšutlhaloši ya Sesotho sa Leboa (PTLH)* which is a reflection of a well-compiled article.

mala¹ (*leina ka bontši*) (5/6) kanale
ye teleletelele ye boleta yeo e
phuthaganego ka teng, e hlotla le
go sepetša dijo tšeo di šitšwego ke
mogodu go di ntšhetša ntle : **Ge go**
hlabilwe kgogo, bana ba fiwa mala,
hlogwana le menatlana

Figure 4: Treatment of the first sense of *mala* in *PTLH*

The treatment of *mala* in example (1) and figure 4 is more appropriate: it distinguishes between the different senses and it gives usage examples. It also indicates the different data categories — someone who does not understand Sesotho will perhaps not be able to distinguish the different categories. The absence of frequency indication in (1) indicates relatively low frequency. Examples of usage are clearly illustrated in bold and paraphrased in such a way that the user is able to understand the true meaning of *mala*.

On the level of the macrostructure the lemma offering of *Sethantšho* for the alphabetical stretch "L" was compared with words occurring more than 200 times in a Sesotho corpus of approximately 1.5 million words. Only non-derived words were taken into consideration as it cannot necessarily be expected from a dictionary to lemmatise nominal and verbal derivations. For example, a dictionary cannot be criticized if it lemmatizes only the non-derived forms of frequently used words, (frequencies given in brackets), such as the

verb stem *lahla* (729) 'lose, throw away' and the noun *lebenkele* (799) 'shop' and not any of their derivations. Typical derivations are verb stems containing verbal suffixes e.g. the perfect, applicative, passive and relative forms, or nouns occurring with locative, diminutive or augmentative suffixes. Consider, for example, the derived verb stems *lahlile* (215) (perfect) 'lost' and *latelwang* (251) (applicative + passive + relative) 'which is followed' as well as the locative derivation *lebenkeleng* (906) 'at the shop'. It would be user friendly if a dictionary did lemmatise frequently used derivations as has been done in *ONSD*. Lemmatisation of derivations, however, cannot be put as a requirement in this evaluation because the editorial policy of the dictionary could simply be *not* to lemmatise certain regularly derived forms of verbs and nouns. This is typically the approach for passive, perfect and locative forms irrespective of the frequency of the specific derivation, as in *Pukuntšu* (*PUKU*). *Sethantšo* did, however, lemmatise a number of frequently derived words, e.g. *lapile* (281) 'hungry'. Consider table 3 for an edited list of these words compared to the lemma list of *Sethantšo*.

Table 3: Top frequencies in the Sesotho corpus for the alphabetical stretch "L" compared to the lemma list of *Sethantšo*

Corpus word	Freq	Sethant.	LEBOYA	1748	NO	LAHLA	729	YES
LE	496927	NO	LEHAE	1682	YES	LAELA	673	NO
LA	79427	YES	LAOLA	1668	YES	LEBATOWA	610	NO
LEHA	12217	YES	LELOKO	1571	NO	LEHE	607	YES
LEBITSO	6783	YES	LEBESE	1543	NO	LEEME	605	NO
LEBAKA	5725	NO	LAKATSA	1272	YES	LEMA	567	NO
LELAPA	5443	NO	LEKALA	1266	YES	LEHLAKORE	500	YES
LEFU	5287	YES	LEBA	1257	NO	LEBE	490	YES
LEKA	4833	YES	LEKANA	1160	YES	LAESENSE	480	NO
LEFATSHE	4799	NO	LEIHLO	1079	YES	LABORARO	474	NO
LATELA	4245	NO	LEKE	1026	YES	LEBO	429	NO
LEFAPHA	4173	NO	LEBALE	1024	NO	LATA	425	YES
LEANO	4149	NO	LEBELLA	939	NO	LEBOHELA	416	NO
LEFA	3914	NO	LEHOLO	878	YES	LEFIFI	410	YES
LEHLOHONOLO	3884	YES	LEKOLA	874	NO	LEFEELA	408	NO
LEBALA	2009	YES	LEETO	814	NO	LEKANYA	407	YES
LEFELLA	1837	NO	LEBENKELE	799	NO	LEE	404	NO
LEBOHA	1762	YES	LEBELO	790	NO	LEFUBA	394	YES

LEHLABULA	387	YES	LABOBEDI	315	NO	LEHANO	228	YES
LELEME	380	NO	LEHLATSIPA	311	YES	LEHOLIMO	224	NO
LABOHLANO	379	NO	LEA	304	NO	LATOLA	221	YES
LELEKA	365	YES	LEBOLLO	297	YES	LEHARE	214	YES
LEBOKOSE	358	NO	LAPILE	281	YES	LELEIRI	214	NO
LABONE	343	NO	LABALABELA	243	YES	LEETSI	204	NO
LEKUNUTU	320	NO	LEHLABATHE	231	YES			
LAKALETSA	316	YES	LEFUTSO	229	NO			

The lemma list of *Sethantšo* does not compare well with the Sesotho words which occur more than 200 times in the Sesotho corpus. The dictionary lemma-tised and treated only 35 of the 75 i.e. 47% of these top frequencies that can be assumed to be words likely to be looked up especially by learners of the language. Common words such as *lebaka* (5,725) 'reason', *Laboraro* (474) 'Wednesday' and *leano* (4,149) 'a plan', are obviously missing. De Schryver and Prinsloo (2000c) suggest that compilers should do much better in the compilation of lemma lists on intuition. Compilers should at least be able to capture the most frequently used words in a language even without the help of frequency lists culled from corpora.

4.2 SSED

The *Sesuto-English Dictionary* was published in 1876 in Lesotho. Several editions followed. The 4th edition, enlarged by Dieterlin was published in 1911. In 1959 a revised and enlarged edition, the *Southern Sotho-English Dictionary* by Paroz was published. This edition was presented in the South African orthography (Moleleki 1999: 243). The edition under discussion in this article is the 8th edition of 1961.

SSED is a classic example of dictionaries that were not compiled through the use of corpora but on introspection over time by Mabile, Dieterlin and Paroz. Moleleki (1999: 243) regards *SSED* as "the most useful and consulted work for Sesotho". He, however, bluntly states that "the work is not user-friendly. It is structured by the presupposition that the user is very conversant with the structure of Sesotho". He is of the opinion that the dictionary is not meant for learners but for those who already have a sound competence in the language. Narrowing the target users down to users who have a sound basic knowledge of the grammar of Sesotho is fine in itself, and ideally dictionaries should be aimed at clearly defined target users (Gouws and Prinsloo 2005: 3). However, if a dictionary is the only significant available reference work for a specific language it has to serve by necessity the needs of the broader Sesotho-speaking community. This includes learners of the language who are mother-tongue

speakers as well as non-mother-tongue speakers. It will be briefly argued in terms of Van Wyk (1995) below that this dictionary unnecessarily excludes users who do not have the required grammatical knowledge of Sesotho simply because it opted for bad choices in lemmatisation strategy and alphabetical ordering.

Consider the following typical examples of articles from *SSED* in figure 5.

<p><i>se·baka</i> (<i>di·</i>) n., place; distance, space; opportunity, time, occasion, chance; <i>ho hloka -</i>, to have no time; <i>ho</i></p>	<p><i>mo·hla</i> (<i>me·</i>) n., day, date; season; <i>le ka - o le mong</i>, never; <i>mehla e a ja, e a roba</i>, time leaves its trail of disabilities; loc. <i>mohla(ng)</i>, <i>me-</i></p>
<p><i>molêthema</i>, n.cl.2, length; adj., long. <i>bolêthô</i>, n., <i>hlonepho</i> substitute for <i>bosiu</i>, night. [cp. <i>bošlôthô</i>].</p>	<p><i>mo·phata</i> (<i>me·</i>) n., cracked pot; big clay pot. <i>maphata</i>, n.plur., tapeworm.</p>
<p><i>le·kgabunyane</i> (<i>ma·</i>) n., homeless person taken as a servant; female servant;</p>	<p><i>lēbèle</i> (<i>ma-</i>) n., grain or plant of corn;</p>

Figure 5: Selected entries in *SSED*

The entry *sebaka* in this extract is indicated as a noun with the translation equivalent paradigm 'place, distance, space, opportunity, time, occasion, chance'. The plural form is indicated in brackets as (*di*). This is followed by several examples of usage e.g. *ho hloka sebaka* 'to have no time'. Nouns are lemmatised according to the first letter of the noun stem, thus the word *lebele* is lemmatised in the alphabetical stretch B on its stem form *-bele* and *mohla* under *-hla*. Likewise *lekgabunyane* is lemmatised under *-kga*, *molethema* and *boletho* under *-le*, and *mophata* and *maphate* under *-pha*. If the user wants to look up *mohla* 'day' in the *SSED* he/she won't find it under M because the user is supposed to know that, since a stem lemmatisation strategy is followed, the prefix *mo-* has to be removed and that it should be looked up on the stem under H, i.e. *-hla*. However, the lemma is not found under H in a normal alphabetical order. The alphabetical stretch H runs from stems beginning with *ha-*, *hi-*, *ho-*, ... *hwi-* but words starting with *hl-* are not listed — they are given under a following main stretch HL as *mo.hla*.

Van Wyk (1995), supported by Prinsloo and De Schryver (1999) and Prinsloo and Theletsane (2018), strongly argue against the use of stem lemmatisation for disjunctively written languages such as Sesotho. They argue that stem lemmatisation is unnecessary and unwanted for disjunctively written languages and that future compilers of paper dictionaries should stick to word lemmatisation. It simply means that instead of expecting the user to identify the stems *-bele*, *-lethema*, *-kgabunyane*, etc. as a prerequisite for look-up, the lexicographer could simply have lemmatised *lebele* and *lekgabunyane* under L, *boletho* under B and *mohla*, *molethema*, *mophata* and *maphata* under M. Van Wyk (1995) also dismisses all claims that stem lemmatisation is superior, more scientific and more economical than word lemmatisation.

Thus, we believe that for *SSED* changing from a word to a stem tradition in the 7th edition was a step in the wrong direction. In the front matter the motivation for changing to stem lemmatisation was the belief that the noun prefix is a mobile and exchangeable element (*SSED*: xii). Even this belief was refuted by Van Wyk (1995: 86) saying that "this assumption is, however, wrong; the morphology of the noun differs in crucial ways from that of the verb. The noun prefix is not mobile or freely exchangeable as Paroz claims".

SSED motivates its viewpoint with the example that *motho* 'a person' should be lemmatised on its stem form *-tho* in order to bring together words "which are similar in origin and related in meaning and of showing better the relative place of a given word in the language". This would hold true for *motho* 'person', *batho* 'people', *setho* 'culture' and *botho* 'mankind'. The resulting entry is given in figure 6 as two columns.

<p>mo·tho (<i>bā</i>) n., human being; person; – <i>ò ka hana</i>, one could refuse; – <i>mo-tonana o kaa ka tlou</i>, old description of the Great God; <i>mmuso wa batho ka batho</i>, democracy; <i>wa batho</i>, chief; <i>ngwana wa batho</i>, poor, unlucky child; <i>m. wa batho</i>, poor, weak person; harmless person; <i>re batho ka ba bang</i>, prov. advocating solidarity; <i>re batho ba mmele wa kgomo</i>, description of the link constituted by the possession of cattle; <i>motho-(ka-)lebitso</i>, one unworthy of the name of human; dim. <i>mothwa(ka)na</i>, small, despicable person; <i>mothohadi</i>, elderly person; <i>batho</i>, interj., shame! [<i>mòthò</i>]. <i>mmabatho</i>, n., good, helpful person; good chief.</p>	<p><i>mmamothwanyane</i>, n., a variety of sorghum bearing large white corn. <i>setho</i>, n., human manners or customs; culture; conduct, morals; adv., as a human being; in a dignified manner; kindly. <i>botho</i>, n., condition of a person, a human being; mankind; humanity; human ways; kindness. <i>bothohadi</i>, n., condition of a great person, of a superman. <i>mothwanahēlēlē</i>, n., bread made with whey. > <i>nthofala</i>, <i>tho</i>.</p>
---	--

Figure 6: The entry for *motho* in *SSED*

Since *setho*, *botho*, *bothohadi*, etc. are treated separately no significant space saving is achieved. The only real advantage is that no separate entry is required for the plural form *batho*, thus saving less than a single line in the dictionary article. Providing plural forms and their treatment as separate articles does require duplication of the treatment of the singular form. However, plural forms are handled in a very economic and effective way in *ONSD* by sacrificing one line for the lemma and a skeleton treatment thereof, with a cross-reference to the singular as in example (2).

(2) **batho** *** *pl. noun* 1/2 See sg. MOTHO

So, in *ONSD* the lemmas *motho*, *batho*, *setho* and *botho* can all be found alphabetically under their first letters by even inexperienced users who do not have any grammatical knowledge of the language.

A second drawback is the phonetic ordering of lemmas in *SSED* instead of an ordinary alphabetical ordering. Digraphs and trigraphs are treated as single letters in their own right instead of as two or three individual letters for alphabetical ordering, i.e. as *a, b, c, (ch), d, e, f, ff, g, h, hl, i, j, k, kg, kh, l, m, n, nc, ng, nq, nx, ny, o, p, ph, pj, pjh, q, qh, r, s, sh, t, th, tj, tjh, tl, tlh, ts, tsh, u, v, w, x, (xh), y, z*. This means that ordinary alphabetical categories are divided into different subsections. For example, T is split up into no less than eight categories e.g. *t, th, tj, tjh, tl, tlh, ts* and *tsh*. For example, *th* in a word such as *thaba* 'mountain' comes after *tetetsa* 'bruise, beat' separated by 24 dictionary pages where in the *New South Sotho dictionary (NSSD)* *tetetsa* and *thaba* are separated by a single entry. To the ordinary user who does not have in-depth phonetic knowledge, this arrangement is user-unfriendly, difficult to comprehend and simply does not make sense. This affects the alphabetical ordering of bigrams and trigrams inside words as well.

It can be concluded in terms of Van Wyk (1995), Prinsloo and De Schryver (1999) and Prinsloo and Theletsane (2018) that stem lemmatisation brings no gain but imposes an unnecessary burden on the user, — it makes it very difficult for especially the inexperienced user to find words. Exactly the same goes for a phonetic instead of an ordinary alphabetical ordering. When stem lemmatisation and phonetic ordering are combined it is even worse and even experienced users struggle to look up words in such dictionaries. They often have to revert to a guidance page, if provided, or can even incorrectly conclude that the word is not in the dictionary. Compilers of future Sesotho dictionaries should seriously consider abandoning both stem lemmatisation and a phonetic alphabetical ordering.

On the level of the macrostructure, the lemma list of *SSED* compares well with top frequencies in the Sesotho corpus as indicated in table 4.

Table 4: Top frequencies in the Sesotho corpus compared to the lemma list of *SSED*

Corpus word	Freq	SSED	LEFATSHE	4799	YES	LEBOHA	1762	YES
LE	496927	YES	LAPENG	4632	YES	LEBOYA	1748	YES
LA	79427	YES	LATELA	4245	YES	LEHAE	1682	YES
LEHA	12217	YES	LEFAPHA	4173	YES	LAOLA	1668	YES
LEBITSO	6783	YES	LEANO	4149	YES	LELOKO	1571	NO
LEKGOTLA	5749	YES	LEFA	3914	YES	LEBESE	1543	YES
LEBAKA	5725	YES	LEHLOHONOLO	3884	YES	LAKATSA	1272	YES
LELAPA	5443	YES	LEBALA	2009	YES	LEKALA	1266	YES
LEFU	5287	YES	LEFELLA	1837	NO	LEBA	1257	YES
LEKA	4833	YES	LEKGETHO	1787	YES	LEKANA	1160	YES

LEHLO	1079	YES	LABORARO	474	YES	LAKALETSA	316	YES
LEKE	1026	YES	LEKGOLO	435	YES	LABOBEDI	315	YES
LEBELLA	939	NO	LEBO	429	YES	LEBETSE	312	YES
LEHOLO	878	YES	LATA	425	YES	LEHLATSIPA	311	YES
LEKOLA	874	YES	LEBOHELA	416	YES	LEA	304	YES
LEHODIMO	871	NO	LEFIFI	410	YES	LEBOLLO	297	YES
LEETO	814	YES	LEFEELA	408	YES	LEKGOWA	295	NO
LEBENKELE	799	YES	LEKANYA	407	YES	LAPILE	281	YES
LEBELO	790	YES	LEE	404	YES	LABALABELA	243	YES
LAHLA	729	YES	LEFUBA	394	YES	LEHWATATA	240	YES
LEKGETLO	703	YES	LEHLABULA	387	YES	LEHLABATHE	231	YES
LAELA	673	YES	LELEME	380	YES	LEFUTSO	229	YES
LEBATOWA	610	NO	LABOHLANO	379	YES	LEHANO	228	YES
LEHE	607	YES	LELEKA	365	YES	LATOLA	221	YES
LEEME	605	NO	LEBOKOSE	358	YES	LEHARE	214	YES
LEMA	567	YES	LABONE	343	YES	LELEIRI	214	NO
LEHLAKORE	500	YES	LEKGOBA	343	YES	LEETSI	204	YES
LEBE	490	YES	LEHLOKWA	323	YES			
LAESENSE	480	NO	LEKUNUTU	320	YES			

SSED lemmatised and treated 75 of the 84 words considered, i.e. 88%. In this case the compilers did well in the selection of top frequencies.

4.3 *Bukantswe*

It is generally believed that electronic dictionaries made a slow start, but will eventually supersede paper dictionaries in many ways.

Though 'electronic lexicography' — the use of digital media for delivering dictionary data — dates back at least as far as 1990, the pace of change has picked up dramatically in the last five years, after a leisurely start. (Rundell 2012: 72)

Good electronic dictionaries are characterised by the utilisation of electronic features enabled by computer technology and utilisation of virtually unlimited space on the internet. The interested reader is referred to De Schryver (2003), and Prinsloo (2019a) for a more detailed discussion of such features and to Bothma, Prinsloo and Heid (2018), Prinsloo, Prinsloo and Prinsloo (2018), Prinsloo (2019a), Prinsloo and Bothma (2020) and Prinsloo and Taljard (2019) for

detailed discussions on user support tools in electronic dictionaries.

Bukantswe has more than 10,000 Sesotho entries with their English equivalents available from <http://bukantswe.sesotho.org/>. Searches can be done in English and Sesotho. In its self-description it is stated that *Bukantswe* is a "Bilingual English–Sesotho dictionary, [the] dataset represents a basic Sesotho dictionary compiled in the creation of a Sesotho language resource". The dictionary was developed by Jako Olivier and is "based on an on line word list published and revised since 1996" (<https://repo.sadilar.org/handle/20.500.12185/419>).

At first glance when a user opens *Bukantswe* a search box is presented and users can start right away by typing in the word they are looking for. Consider figure 7 for the search word *lapeng*.



Figure 7: *Bukantswe*'s opening screen and results for *lapeng*

The user is informed about the size of the dictionary — it contains 10,075 entries. The current screen layout has been changed from the previous one which offered a clickable A–Z alphabetical option as in figure 8.



Figure 8: A–Z sequence of alphabetical stretches from *Bukantswe*

Bukantswe gives users translation equivalents of words and not any pronunciation guidance or examples of usage.

It is a virtue of the dictionary that *noun classes* and persons are indicated as in (3a–3d). It indicates singular forms of nouns with an "s.", followed by the class number in brackets and the same for plural forms with "pl." as in (3a). Class numbers and indication of first and second person singular and plural are also given for pronouns as in (3b–3d), but are missing in cases such as (3e) and (3f) where class indication as (s.9) and (pl.02) respectively should be given.

- (3)
- | | |
|--|-----------------------------------|
| a. agente (s.9) diagente (pl.10) | agent |
| b. Ana demonstrative pronoun (06) | these |
| c. lohle | all (05) (quantitative pronoun) |
| d. wena | you (singular) (absolute pronoun) |
| e. kakaretso [1] | abstract |
| f. badimo [1] | ancestors |

Part of speech (POS) is indicated for nouns (n.), verbs (v.), adjectives (adj.), pronouns (e.g. quantitative pronoun), etc. as in example (4)

- (4)
- | | |
|--|---------------------------------|
| karabo (s.9) dikarabo (pl.10) [1] | answer (n.) |
| baleha | flee (v.) |
| a matonana | huge (06) (adj.) |
| ohle | all (06) (quantitative pronoun) |
| bjja | slap (ideophone) |

However, POS indication is not done consistently — consider the missing POS indication for nouns in (5a) and (5b) and missing indications for verbs in (5c) and (5d).

- (5)
- | | |
|---|-----------|
| a. avenyu (s.9) diavenyu (pl.10) | avenue |
| b. ketso (s.9) diketso (pl.10) [1] | action |
| c. bipetsana | suffocate |
| d. ntjhafatsa | modify |

The dictionary indicates etymology by means of a convention "<="">" as in (6a). In this case it indicates that *saena* is derived from the English word *sign*. Nothing is inserted between the two double quotation marks. If, however, the origin is from another language, e.g. Afrikaans the original word or language is indicated between these two double quotation marks followed by a closed bracket as in (6b–6f). In (6b), for example, it is indicated that *kalaka* is borrowed

from 'kalk'. The placement of the closed bracket directly after the Afrikaans word is unclear.

- (6)
- | | |
|---|------------------------------------|
| a. saena | sign (v.) (<=""> |
| b. kalaka | lime (<="" kalk)=""> |
| c. bora | drill (v.) (<="" boor)=""> |
| d. amen | amen (<="" afr)=""> |
| e. ankere (s.9) diankere (pl.10) | anchor (<="" anker)=""> |
| f. borashe ba terata | wire brush (<="" draad)="" afr=""> |

Another virtue of *Bukantswe* is that *homonyms* are distinguished. *Amohetse* has three unrelated translations, 'accepted', 'accommodated' and 'adopted'. Homonym distinction is made by homonym numbers in square brackets following the lemma as in (7), indicating three unrelated meanings for *amohetse*. Homonym numbers are even supplied for translation equivalents as in (7b) but in such cases it merely looks like synonym paradigms, i.e. *bjara* and *bjaratsa* as translation equivalents of *crush*.

- (7)
- | |
|---------------------------------------|
| a. amohetse [1] accepted (v.) |
| amohetse [2] accommodated (v.) |
| amohetse [3] adopted (v.) |
| b. bjara crush (v.) [1] |
| bjaratsa crush (v.) [2] |

Scientific and domain labels are used throughout the dictionary. Consider (8a) and (8b) for natural elements and (8d) for domain indication.

- (8)
- | | |
|----------------------|--------------------------|
| a. Aluminiamo | Aluminium (Al) [Element] |
| b. Argone | Argon (Ar) [Element] |
| c. bela | boil (liquids) (v.) |
| d. Amose | Amos (Biblical Name) |

In (8) *Aluminiamo* and *Argone* are labelled as [Element] i.e. belonging to the periodic table of natural elements and *Amose* as a Biblical name in the religion domain. Such labels are valuable to the user to distinguish between words belonging to the general language versus words occurring in specific domains. So, for example, *solution* refers to the solving of a problem in the general language but is domain specific if referring to a chemical solution.

A number of shortcomings were, however, noticed.

Users who want to look up frequently used words, e.g. *mosadi* (5,140) 'woman' and *monna* (7,191) 'man' find no results. In the case of *monna* the full

string "monna (s.1) banna (pl.2) [2]" as search node is required; even searching for "monna (s.1) banna (pl.2)", i.e. without "[2]" renders no results. Exactly the same holds true for *mosadi*, i.e. searching for the full string "*mosadi (s.1) basadi (pl.2) [3]*" is required. This is a serious problem which existed at the time of consultation (December 2019–January 2020) and needs to be corrected urgently. As it stands the user would simply conclude that these top frequency items, *monna* and *mosadi* are not in the dictionary.

Another major problem is that the results obtained from the look-up often reflect partial-matches, i.e. mere blind data base hits instead of reflecting a dictionary article. For example, the result for the search of the second-most frequently used word in the Sesotho corpus, *le* (426,927) returns no less than 270 results of which (9) is an extract. It is clear that all words containing '*le*' (bold-faced in (9)), either in Sesotho or English have simply been blindly extracted from the database.

(9)	
tenehile	irritated
teotsa ya pensele	pencil sharpener
themperetjhara ya mmele	body temperature
thomello dikerafike ho tswa ka ntle	importing graphics
Thuto (le bitso) [1]	L esson (female name)
tsamaisa [2]	l ead (v.)
tsamaisa [3]	l et go (v.)
tsamaisa [4]	l et someone/something walk (v.)

The user who wants to know the meaning of the word *le* has to read through 270 unwanted entries and ironically, the most basic meanings of *le* 'and, with, also' are not given, with *hammoho le* 'together with' as the closest match. This probably represents the worst case of information overload — something that is frequently cautioned against in the literature, cf. Gouws and Tarp (2017).

Consider also the results for *motho* (23,052) 'person' in (10):

(10)	
ha ho motho	nobody
tidima ya tse jang motho	pathologist

Not indicating the basic meaning 'human being, person' but giving 'pathologist' is completely illogical and misinforms the user.

From all of these examples it is clear that even for the correctly treated lemmas there are not much microstructural information or data types offered, as discussed in terms of Gouws and Prinsloo (2005) above. The information types are limited to a translation equivalent and hardly fulfil the basic requirements of a dictionary. The dictionary does not give any synonyms where applicable, no related words, no examples of usage, no pronunciation guidance, etc. Pronuncia-

tion guidance can be very effectively given in electronic dictionaries by means of clickable icons; see figure 10. These entries in *Bukantswe* are examples of what Prinsloo and Taljard (2017: 431) call "ontoereikende bewerking" (insufficient treatment) of the lemmas. Consider the information given for *ja* 'eat' in (11) compared to corpus extracts in table 5, *SSED* and *ONSD* in figure 9 and *MD*, figure 10 respectively. Insufficient treatment is unacceptable in electronic dictionaries because virtually unlimited space and true electronic features enabled by the computer are available, cf. Prinsloo (2019b) for a detailed discussion.

(11)
ja [1] eat (v.)

Prinsloo (2015) indicates the value of even a very limited corpus in the compilation of dictionaries. Corpus lines suggest a number of senses that can be distinguished for *-ja*. The boldfaced words in the final column in table 5 describe the sense in which the lemma *-ja* 'eat' has been used in the different lines.

Table 5: Concordance lines and senses of *ja*

0 etsetswang Jehova; le tla	ja	mahobe a se nang ditomoso ka	to eat
anyetsa, a hopola hare 0 tla	ja	mofufutso wa bohlae ba hae	to succeed
ka ba Modimo wa rona. Le tla	ja	menono ya ditjhaba, le ithor	to enjoy
ka, o ne a nahana hore o tla	ja	mohlapeng? O ne a nahana ho	to use what is near
etsang. 35 Ba bohlae ba tla	ja	lefa la hlomphe; ho e le dit	to respect
la di sebetsa." "Monna a tla	ja	batho." "Ke ile ka tshoha ha	to deceive
di tla be di 0 dumella 0 tla	ja	monate. Morwetsana 0 habile	to live
e tla nyopa, mme mollo 0 tla	ja	moAho wa ya kgopo. 35 O emol	to burn down
hapanyetsa taba ena, empa ya	ja	setsi sekolong mona. Eitse	to happen frequently
a, ba re ho yena: O ke ke wa	ja	lefa la ntlo ya ntata rona	to take the inheritance
ba 0 romileng? 22 0 se ke wa	ja	mofutsana ka hoba e le mofut	to rob a poor person
a bolela, feel a 0 se ke wa	ja	tlwaye ere ha ke 0 roma, 0 b	get used to
ba molemo haholo ha 0 ka wa	ja	masapo a hlooho hore ka moso	to think
marao a tshabile disale. Tsa	ja	mothinya wa ho qetela di nts	to run away
leleme; ba le ratANG ba tla	ja	tholwana tsa lona. 22 Ya fum	to reap rewards of

Such corpus lines are invaluable to the lexicographer to distinguish the different senses of *eat* for consideration for inclusion in the dictionary. It often happens that lexicographers are alerted to senses that they might have missed if they had to rely on intuition only. It has to be realised, however, that not every single concordance line represents a different sense — it is the task of the lexicographer to decide on the number of senses to be distinguished.

The entry for *ja* in *SSED*, although not compiled using a corpus, as well as in *ONSD* is richer because it captures a number of senses through translation equivalents such as 'to eat' 'to despoil', 'to cost', 'to cause pain', 'to ache', etc.

SSED	ONSD
<p>ja^a (<i>jelē</i>, imper. <i>eja</i>, <i>jaa</i>, pres., past <i>eja</i>, pass. <i>jewa</i>, <i>jowa</i>, perf. <i>jelwē</i>, <i>jewē</i>, <i>jowē</i>, obs. perf. <i>lle</i>, obs. pass. perf. <i>hlewē</i>) v.t., to eat; fig., to despoil; to cost; to cause pain, to ache; to affect (a disease); to put on; to bump; <i>ho ja</i>, food; act of eating;</p>	<p>ja *** verb ■ eat • Ke ja nama ya kgomo. I eat beef. ○ ja bogale ■ be angry ○ ja fase ■ fail; run away ○ ja hlogo ■ think; be deep in thought ○ ja leonyane ■ stalk ○ tša go ja ■ something to eat</p>

Figure 9: *-ja* in SSED and ONSD

The entry for *eat* in MD also indicates lexicographic richness of treatment.

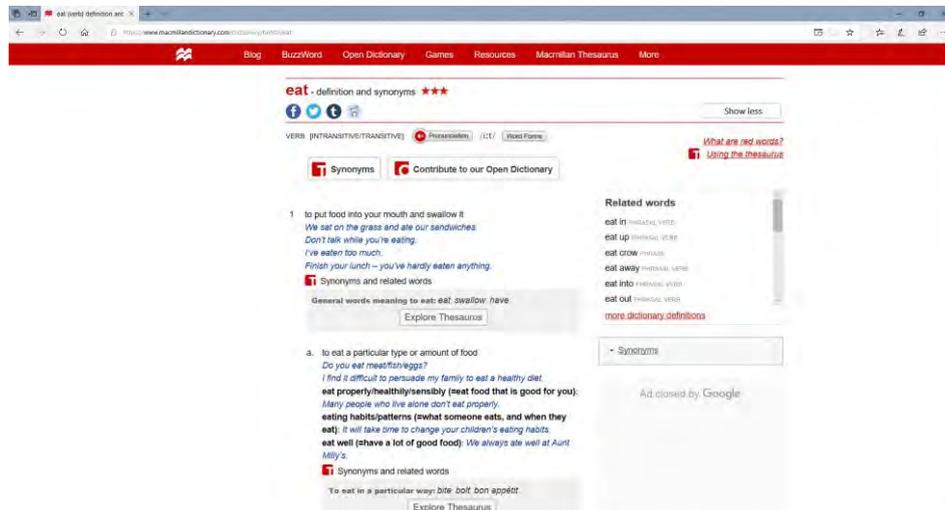


Figure 10: *eat* in MD
(https://www.macmillandictionary.com/dictionary/british/eat)

In figure 10 a wealth of information types are given such as different senses, translation equivalents, examples of usage, frequency indication, pronunciation, word forms, definitions, etc. This is a good example of what future Sesotho electronic dictionaries should look like.

A further shortcoming in *Bukantswe* is the lack of even a very basic user's guide to the dictionary.

On the level of the macrostructure, as indicated in table 6, the lemma list of *Bukantswe* does not compare well with top frequencies in the Sesotho corpus. This is aggravated by the presumed technical detection problem described above, i.e. cases where the dictionary gives look-up results only if the full string

is entered. It simply means searches for hundreds, if not thousands of especially nouns will not render any results.

Table 6: Top frequencies in the Sesotho corpus compared to the lemma list of *Bukantswe*

Corpus word	Freq	Bukantswe						
LE	496927	NO	LEKANA	1160	NO	LEE	404	YES
LA	79427	NO	LEIHLO	1079	NO	LEFUBA	394	NO
LEHA	12217	NO	LEKE	1026	NO	LEHLABULA	387	NO
LEBITSO	6783	NO	LEBALE	1024	NO	LELEME	380	NO
LEKGOTLA	5749	NO	LEBELLA	939	YES	LABOHLANO	379	YES
LEBAKA	5725	NO	LEHOLO	878	NO	LELEKA	365	NO
LELAPA	5443	NO	LEKOLA	874	NO	LEBOKOSE	358	NO
LEFU	5287	NO	LEHODIMO	871	NO	LABONE	343	YES
LEKA	4833	NO	LEETO	814	YES	LEKGOBA	343	NO
LEFATSHE	4799	YES	LEBENKELE	799	NO	LEHLOKWA	323	NO
LAPENG	4632	YES	LEBELO	790	YES	LEKUNUTU	320	YES
LATELA	4245	NO	LAHLA	729	NO	LAKALETSA	316	YES
LEFAPHA	4173	NO	LEKGETLO	703	NO	LABOBEDI	315	YES
LEANO	4149	NO	LAELA	673	NO	LEBETSE	312	YES
LEFA	3914	YES	LEBATOWA	610	NO	LEHLATSIPA	311	NO
LEHLOHONOLO	3884	NO	LEHE	607	NO	LEA	304	NO
LEBALA	2009	YES	LEEME	605	NO	LEBOLLO	297	YES
LEFELLA	1837	NO	LEMA	567	NO	LEKGOWA	295	NO
LEKGETHO	1787	NO	LEHLAKORE	500	NO	LABALABELA	243	NO
LEBOHA	1762	YES	LEBE	490	NO	LEHWATATA	240	NO
LEBOYA	1748	YES	LAESENSE	480	NO	LEHLABATHE	231	NO
LEHAE	1682	NO	LABORARO	474	YES	LEFUTSO	229	NO
LAOLA	1668	NO	LEKGOLO	435	NO	LEHANO	228	NO
LELOKO	1571	NO	LEBO	429	NO	LATOLA	221	NO
LEBESE	1543	NO	LATA	425	NO	LEHARE	214	NO
LAKATSA	1272	NO	LEBOHELA	416	YES	LELEIRI	214	NO
LEKALA	1266	NO	LEFIFI	410	NO	LEETSI	204	NO
LEBA	1257	NO	LEFEELA	408	NO			
			LEKANYA	407	NO			

Bukantswe lemmatised and treated 19 of the 84 words considered i.e. 23% which reflects insufficient coverage of the selected top frequencies. Even without a corpus, the lexicographer is expected to capture a greater percentage of the most frequently used verbs on intuition as mentioned above in terms of De Schryver and Prinsloo (2000c).

5. Conclusion and future work

Sesotho lexicography is in a developing phase and much more research and description of lexicographic issues is required to bring the body of knowledge for this language on a par with its sister languages Sepedi and Setswana and lexicographic achievement of dictionaries for major languages of the world such as English, French, German, etc. The virtues and shortcomings raised in this article in respect of *Sethantšo*, *SSED* and *Bukantswe* are a true reflection of most African language dictionaries. Gouws and Prinsloo (2005: 9) say that "the publication of any dictionary should not only be the result of the preceding compilation activities but it has to be regarded as the culmination of a much more comprehensive set of activities, the so-called lexicographic process". Furthermore, the inconsistencies addressed in this article reflect the need for Sesotho lexicographers to use corpora in dictionary compilation in order to enhance the quality of entries on both microstructural and macrostructural levels. Corpus utilisation will also enable compilers to indicate frequencies of words in the dictionary as has been done in *ONSD* by means of a 3-star rating system, as in figure 9 above. After decades in existence, currently available Sesotho dictionaries are in dire need for revision and new dictionaries aimed at specific target users should be compiled. A language cannot be served by only a few dictionaries compiled as a one-size-fits-all for user needs. Gouws and Prinsloo (2010: 505) state that no single dictionary can be everything for everyone. There is also a strong need for community involvement in the compilation of Sesotho dictionaries in a true Afro-centric approach where more mother-tongue speakers of Sesotho take the initiative to compile good Sesotho dictionaries, cf. Prinsloo (2017) and Prinsloo (2019b).

Acknowledgements

This research is supported in part by (a) the South African Centre for Digital Language Resources (SADiLaR) and (b) the National Research Foundation of South Africa (Grant specific unique reference number 85763). The Grant holder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the authors, and that the sponsors accept no liability whatsoever in this regard.

Endnote

1. The term 'Bantu' got stigmatised during the Apartheid Era in South Africa. Therefore the term 'African' will be used in this article even in reference to what is internationally referred to as 'Bantu languages'.

Bibliography

Dictionaries

- (*Bukantswe*) *Sesotho Online, Bukantswe v.3*. Available at <http://bukantswe.sesotho.org/>. [accessed: January 7, 2020].
- (*ESD*) Motsapi, M. (Ed.). 2015. *English–Sesotho Dictionary*. Cape Town: South African Heritage Publishers.
- (*MD*) *Macmillan Dictionary*. Available at <http://www.macmillandictionaries.com/> [accessed: May 5, 2019].
- (*NSSD*) Chaphole, S.R. 1997. *New South Sotho Dictionary. English–South Sotho, South Sotho–English*. Pietermaritzburg: Shuter & Shooter.
- (*ONSD*) De Schryver, G.-M. (Ed.). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*. Cape Town: OUP Southern Africa.
- (*PTLH*) Mojela, M.V. (Ed.). 2007. *Pukuntšutlhaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
- (*PUKU*) Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšu*. Pretoria: J.L. van Schaik.
- (*Sethantšo*) Hlalele, B. 2005. *Longman Sethantšo sa Sesotho*. Maseru: Longman Lesotho.
- (*SSED*) Mabile, A. and H. Dieterlen. 1988. *Southern Sotho–English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.

Other

- Bothma, T.J.D., D.J. Prinsloo and U. Heid.** 2018. A Taxonomy of User Guidance Devices for e-Lexicography. *Lexicographica. International Annual for Lexicography* 33: 391-422.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000a. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries. Part 1: The Macrostructure. *South African Journal of African Languages* 20(4): 291-309.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000b. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries. Part 2: The Microstructure. *South African Journal of African Languages* 20(4): 310-330.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000c. (In)consistencies and the Miraculous Consistency Ratio of '(x 1.25)⁴ = x 2.44', A Perspective on Corpus-based versus Non-corpus-based Lemma-sign Lists. Paper presented at the Fifth International Conference of the African Association for Lexicography (AFRILEX), University of Stellenbosch, Stellenbosch, 3–5 July 2000.

- Gouws, R.H.** 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–25 March 1989*: 52-65. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: AFRICAN SUN MeDIA.
- Gouws, R.H. and D.J. Prinsloo.** 2010. Surrogaatekwivalensie in tweetalige woordeboeke met spesifieke verwysing na zero-ekwivalensie in Afrikataalwoordeboeke. [Surrogate Equivalence in Bilingual Dictionaries with Special Reference to Zero-equivalence in Dictionaries for African Languages]. *Tydskrif vir Geesteswetenskappe* 50(4): 502-519.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Moleleki, M.A.** 1999. The State of Lexicography in Sesotho. *Lexikos*. 9: 241-247.
- Motjope-Mokhali, Tankiso Lucia.** 2016. *A Comparative Analysis of Sesuto–English Dictionary and Sethantso sa SeSotho with Reference to Lexical Entries and Dictionary Design*. Unpublished doctoral thesis. Pretoria: UNISA. Available at URI: <http://hdl.handle.net/10500/22205>.
- Nkomo, D.** 2010. Affirming a Role for Specialised Dictionaries in Indigenous African Languages. *Lexikos*. 20: 371-389.
- Otlogetswe, T.J.** 2009a. *English–Setswana Dictionary*. Second edition. Pentagon Publishers: Gaborone.
- Otlogetswe, T.J. (Ed.).** 2009b. *MLA Kgasa: A Pioneer Setswana Lexicographer*. CASAS Book series 64. Cape Town. CASAS.
- Otlogetswe, T.J.** 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.
- Otlogetswe, T.J.** 2013. Oxford English–Setswana, Setswana–English School Dictionary. Oxford: OUP.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Prinsloo, D.J.** 2013. Lexicography of the Sotho Languages. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries: An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 929-946. Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science. HSK Vol. 5.4. Berlin/Boston: Mouton de Gruyter.
- Prinsloo, D.J.** 2015. Corpus-based Lexicography for Under-resourced Languages — Maximizing the Limited Corpus. Paper read at AELINCO 2015, 7th Conference on Corpus Linguistics, Valladolid, Spain, 5–7 March 2015.
- Prinsloo, D.J.** 2017. Analyzing Words as a Social Enterprise: Lexicography in Africa with Specific Reference to South Africa. Miller, J. (Ed.). 2017. *Analysing Words as a Social Enterprise: Celebrating 40 Years of the 1975 Helsinki Declaration on Lexicography. Collected Papers from AustralLex 2015*: 42-59. <https://www.adelaide.edu.au/australlex/publications/>.
- Prinsloo, D.J.** 2019a Detection and Lexicographic Treatment of Salient Features in e-Dictionaries for African Languages. *International Journal of lexicography*. ecz031, <https://doi.org/10.1093/ijl/ecz031>. Published: 27 November 2019.
- Prinsloo, D.J.** 2019b A Perspective on the Past, Present and Future of Lexicography with Specific Reference to Africa. Gürlek, Mehmet, Ahmet Naim Çiçekler and Yasin Taşdemir. 2019. *Proceedings of the 13th International Conference of the Asian Association for Lexicography, ASIA-*

LEX 2019, 19–21 June 2019, Istanbul University Congress and Culture Center, Istanbul Turkey: 148-160. Istanbul: Asos.

- Prinsloo, D.J. and T.J.D. Bothma.** 2020. A Copulative Decision Tree as a Writing Tool for Sepedi. *South African Journal of African Languages* 40(1): 85-97.
- Prinsloo, D.J. and G.-M. de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, D.J., J.V. Prinsloo and Daniel Prinsloo.** 2018. African Lexicography in the Internet Era. Pedro A. Fuertes-Olivera (Ed.). 2018. *The Routledge Handbook of Lexicography*: 487-502. London: Routledge.
- Prinsloo, D.J. and E. Taljard.** 2017. Afrikataalleksikografie: Gister, vandag en môre [African Language Lexicography: Yesterday, Today and Tomorrow]. *Lexikos* 27: 427-456.
- Prinsloo, D.J. and E. Taljard.** 2019. The Sepedi Helper Writing Assistant: A User Study. *Language Matters* 50(2): 73-99.
- Prinsloo, D.J. and T. Theletsane.** 2018. Stem Lemmatization and Phonetic Ordering of Lemmas in Sotho Dictionaries from a User Perspective. *Conference Booklet of the 23rd International Conference of Afrilex, University of the Western Cape, Cape Town, South Africa. 27–29 June 2018*: 40-42. Available online at <https://afrilex.africanlanguages.com/AFRILEX%202018%20Booklet.pdf?timestamp=1578114060000>.
- Rundell, M.** 2012. 'It Works in Practice but Will it Work in Theory?' The Uneasy Relationship between Lexicography and Matters Theoretical. Vatvedt Fjeld, Ruth and Julie Matilde Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012, Oslo*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture

Sven Tarp, *Department of Afrikaans and Dutch, Stellenbosch University, South Africa; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; International Centre for Lexicography, Universidad de Valladolid, Spain; and Centre for Lexicography, University of Aarhus, Denmark (st@cc.au.dk)*

and

Rufus Gouws, *Department of Afrikaans and Dutch, Stellenbosch University, South Africa (rhg@sun.ac.za)*

Abstract: This paper deals with the design of digital lexicographical products. It introduces the philosophy of human-centered design, as explained in the work of Don Norman, and discusses central design concepts like affordances, signifiers, feedback, and other forms of good communication from lexicographer to user. The successful use of traditional dictionaries often relied on the presumed reference skills of the envisaged target user group. Especially in printed dictionaries with their space restrictions, lexicographers used condensed entries, abbreviations, and different types of structural indicators to save space. This often was to the detriment of the user who struggled to retrieve the required information from the data on offer. The digital environment has created new opportunities for lexicographers to assist their users in a far better way. Using principles of human-centered design, this paper shows the emergence of the application of some of these approaches in existing online dictionaries. Going beyond the scope of traditional online dictionaries, examples are taken from integrated e-reading dictionaries and lexicography-assisted writing assistants where even stronger user-centered design features come to the fore with regard to respectively text reception and text production assistance. Where the original concept of a dictionary culture was primarily based on the reference skills of the user, a new lexicographical culture is suggested that adheres to human-centered design principles. It places the responsibility on the shoulders of the lexicographer to design lexicographic products that enable intuitive use and ensure improved lexicographic success.

Keywords: REFERENCE SKILLS, DICTIONARY CULTURE, LEXICOGRAPHERS' RESPONSIBILITY, HUMAN-CENTERED DESIGN, INTUITIVE USE, AFFORDANCES, SIGNIFIERS, FEEDBACK, INTERDISCIPLINARY COLLABORATION, NEW LEXICOGRAPHICAL CULTURE

Opsomming: Naslaanvaardighede of mensgerigte ontwerp: Op pad na 'n nuwe leksikografiese kultuur. Hierdie artikel handel oor die ontwerp van digitale leksikografiese produkte. Dit stel die filosofie van mensgerigte ontwerp bekend, soos uiteengesit in die

werk van Don Norman, en bespreek sentrale ontwerpsbegrippe soos beskikbaarhede, aanduiders, terugvoering en ander vorme van goeie kommunikasie tussen leksikograaf en gebruiker. Die suksesvolle gebruik van tradisionele woordeboeke het dikwels berus op die vermeende naslaanvaardighede van die veronderstelde teikengebruikersgroep. Veral in gedrukte woordeboeke met hulle ruimtebeperkings het leksikograwe verdigte inskrywings, afkortings en verskillende tipes struktuurmerkers gebruik om ruimte te spaar. Dit was dikwels tot die nadeel van die gebruiker wat moes sukkel om die nodige inligting aan die aangebode data te onttrek. Die aanlyn omgewing het nuwe moontlikhede vir leksikograwe geskep om hulle gebruikers op 'n baie beter manier by te staan. Met behulp van beginsels van mensgerigte ontwerp, soos uiteengesit in die werk van Don Norman, bespreek hierdie artikel die inwerkingstelling van die toepassing van sommige van hierdie benaderings in bestaande aanlyn woordeboeke. Deur verder as tradisionele aanlyn woordeboeke te kyk, word voorbeelde uit geïntegreerde e-leeswoordeboeke en leksikografiese ondersteunde skryfhupe geneem waar nog sterker mensgerigte ontwerpskenmerke m.b.t onderskeidelik teksresepsie- en teksproduksiehupe na vore tree. Waar die oorspronklike begrip van 'n woordeboek-kultuur primêr berus het op die naslaanvaardighede van die gebruiker word 'n nuwe leksikografiese kultuur voorgestel wat op mensgerigte ontwerpsbeginsels steun. Dit plaas die verantwoordelikheid op die skouers van die leksikograaf om leksikografiese produkte te ontwerp wat intuïtiewe gebruik moontlik maak en verbeterde leksikografiese sukses verseker.

Sleutelwoorde: NASLAANVAARDIGHEDE, WOORDEBOEKKULTUUR, LEKSIKOGRAWSE VERANTWOORDELIKHEID, MENSGERIGTE ONTWERP, INTUÏTIEWE GEBRUIK, BESKIKBAARHEDE, AANDUIDERS, TERUGVOERING, INTERDISCIPLINÊRE SAMEWERKING, NUWE LEKSIKOGRAFIESE KULTUUR

Great designers produce pleasurable experiences.
Norman (2013: 10)

0. Introduction

Wiegand (1998) rightly defined dictionaries as utility tools produced with the genuine purpose of satisfying human needs, or more precisely, information needs of specific types. This fundamental approach to the discipline, which can be considered the cornerstone of his theoretical building, is shared by many other lexicographers although frequently disagreeing with Wiegand in other aspects, among them Bergenholtz and Tarp (2003).

During most of their existence, the possession and usage of dictionaries have been the privilege of the few due to social and technological constraints. For centuries, the artisanal compilation methods and the material used to produce dictionaries strongly limited the number of available copies and made them inaccessible and unaffordable for most people who, in addition, could frequently not even read and write. All this changed with the introduction and improvement of the printing technology and the general alphabetization required by the industrial revolution. Especially from the late nineteenth century on-

wards, the editions grew bigger and bigger allowing dictionaries to reach out to the majority of people. Like other classes of utility tools, dictionaries became increasingly *everyday things*, at least in the most developed countries.

Being products ideally conceived to meet human needs, dictionaries and other lexicographical tools would be expected to be human-centered in all aspects, including their design. However, technological progress, although laying the foundation for lexicographical products of still higher quality, has not been unequivocally favourable to lexicography. Hanks (2013: 512), for instance, reports how the new printing technology and typographic achievements enabled "lexicographers to cram vast quantities of information elegantly and legibly onto each page and to disseminate large numbers of identical copies of completed dictionaries quickly and efficiently". The downside of this positive development is that the large amount of "information" (read: data) gave rise to more and more complex dictionary articles and structures.

Wiegand (1990) and Wiegand et al. (2013) have detected and described numerous lexicographical structures representing different types of relationship between the various classes of data contained in dictionaries. Practical lexicographers often negotiate only a few structures (like macro- and micro-structures) and take little cognizance of the many and diverse structures prevailing in their dictionaries. But these structures exist objectively. Together with data condensation, they are probably a major reason why dictionary successful consultation often still turns out to be difficult for many users. To this should be added the still bigger editions that also contributed to a growing distance between lexicographers and users. Whereas lexicographers in the small-edition era usually had personal knowledge of a relevant segment of their target group, this segment grew smaller and less significant over time. The inevitable result was lexicographical alienation with the dictionary user ending up as a "well-known unknown" (Wiegand 1977).

The alienation, combined with a lack of personal feedback, impeded a comprehensive insight into the problems experienced by dictionary users. It may also have created a sort of academic arrogance among some lexicographers who shifted the responsibility from themselves to their users. The latter's inadequate "reference skills" were blamed for the consultation problems. The need for a special "dictionary culture" with school children and other potential users being trained in look-up techniques was promoted (see Yamada 2014, among many others). This situation may look comfortable from the lexicographer's point of view. But it strongly contradicts the very nature of dictionaries and other lexicographical products as utility products that are not only designed to meet human information needs but also expected to allow easy, fast, and successful consultation. That is, it contradicts the principle of human-centered design.

These imperfections may, to a certain extent, have been justified in the era of printed dictionaries where the technical room for manoeuvring was tight. But this is no longer the case in the present era where disruptive technologies

make allowance for completely new ways of consulting lexicographical products. We therefore strongly advocate that the responsibility for a good consultation experience should shift from the user to the producer. With this purpose, we will introduce the modern concept of human-centered design as developed by Norman (2013). In the subsequent sections, we will try to apply this concept to "traditional" dictionaries as well as other lexicographical products integrated into digital devices like e-readers and writing assistants. Based on the positive aspects from this discussion, we will finally outline a new lexicographical culture under the current technological possibilities.

1. Concept of human-centered design

There may be few lexicographers who know the term *Norman doors*. But among industrial designers of everyday things, the term is well-known. It refers to doors whose design makes it difficult, or even impossible, to see how they open and should be used. The doors are named after the American engineer and cognitive scientist Don Norman who has severely criticized their user-unfriendly design and used them to illustrate the need for a human-centered approach.

Norman has authored several books on industrial design, among them *Emotional Design: Why We Love (or Hate) Everyday Things* (Norman 2004), *Living with Complexity* (Norman 2010), and *The Design of Everyday Things* (Norman 2013). He was one of the experts that were called upon to analyze the nuclear accident at Three Mile Island back in 1979. Here, the committee he was on discovered that "the plant's control rooms were so poorly designed that error was inevitable: design was at fault, not the operators" who initially were blamed for the accident (Norman 2013: 7).

This conclusion goes like a red thread through his writings. Far too often users are blamed when they make mistakes or do not know how to use modern products. This holds true both for simple everyday objects where difficulties may lead to frustration, and for complex technological products where the consequences may be much more severe.

It is time, Norman tells us, to "reverse the situation". It is the duty of designers "to understand people", and not the other way around. A major problem is that design to a large extent is done by "engineers who are experts in technology but limited in their understanding of people". These experts mistakenly think "that logical explanation is sufficient", and that everything would work smoothly if "only people would read the instructions". However, "humans are amazingly complex". We therefore have to "accept human behavior the way it is, not the way we would wish it to be" (p. 6).

As a solution, Norman suggests *human-centered design* (HCD) which he defines as "an approach that puts human needs, capabilities, and behavior first", and "then designs to accommodate those needs, capabilities, and ways of behaving" (p. 8). The overall objective is to create products that are not only

"understandable and usable". In the best of cases, they should also be "delightful and enjoyable" which means that "attention must be paid to the entire experience", i.e. "the aesthetics of form and the quality of interaction" (p. 4). This focus on aesthetics is due to research on emotion and cognition showing that "attractive things really do work better" (Norman 2004: 17).

The human-centered approach regards the best design to be invisible in the sense that it does not draw attention to itself and allows its users to intuitively know both the purpose and the functioning of the product. Besides a solid understanding of both technology and human psychology, this kind of design requires "good communication" where the device indicates to its user "what actions are possible, what is happening, and what is about to happen" (Norman 2013: 8). In this respect, good communication is especially important when there are problems:

This is where good design is essential. Designers need to focus their attention on the cases where things go wrong, not just on when things work as planned. Actually, this is where the most satisfaction can arise: when something goes wrong but the machine highlights the problems, then the person understands the issue, takes the proper actions, and the problem is solved. When this happens smoothly, the collaboration of person and device feels wonderful.

(Norman 2013: 9)

As can be seen, the human-centered design philosophy is exciting. Its relevance to modern dictionaries and other lexicographical products is evident. But how can this philosophy actually be applied? In his book, Norman (2013) discusses a number of concepts, or principles, and illustrates them with a large number of practical examples. An in-depth discussion of all these interesting issues goes beyond the scope of this article. We have therefore selected a few questions that we consider particularly relevant to lexicography.

Norman initially draws our attention to two important characteristics of good design, namely *discoverability* and *understanding*:

Discoverability: Is it possible to even figure out what actions are possible and where and how to perform them? Understanding: What does it all mean? How is the product supposed to be used? What do all the different controls and settings mean? (Norman 2013: 3)

The answers to these questions are usually given in manuals or through personal instruction. Norman, by contrast, insists that manuals and instructions are only required, if "the device is indeed complex, but it should be unnecessary for simple things" (p. 3). This is certainly a discussion relevant to lexicography: Do we need user guides? Are our products simple or complex? The discussion of user guides has been going on for years and no final agreement has been reached, cf. Kirkpatrick (1989), Nielsen (2005), Vrbinc and Vrbinc (2020). Hopefully, the discussion can now be placed on more solid ground.

According to Norman (2013), discoverability results from the designers' appropriate application of six fundamental psychological concepts: *affordance*, *signifiers*, *feedback*, *constraints*, *mapping*, and *conceptual model*. We will briefly discuss the first three concepts which seem to be most relevant to the purpose of this article.

An *affordance* is not a property in itself, but a relationship between the properties of both a physical object and a person that is interacting with this object. An affordance helps the person figure out how the object can be used and "what actions are possible without the need for labels or instructions" (p. 13). For instance, "a chair affords ('is for') support and, therefore, affords sitting" (p. 11). Affordances may be visible or not. But to be effective, they should be perceivable and discoverable. If this is not the case, "some means of signaling its presence is required" (p. 12). This is the role of the signifier.

A *signifier* is "any perceivable indicator that communicates appropriate behavior to a person" (p. 14). An example could be the sign PULL on a door. Norman considers signifiers to be more important than affordances, because "they communicate how to use the design." (p. 19). As to the mutual relationship between the two concepts, he writes:

Some affordances are perceivable. Others are invisible. Signifiers are signals. Some signifiers are signs, labels, and drawings placed in the world, or arrows and diagrams indicating what is to be acted upon and in which direction to gesture, or other instructions. Some signifiers are simply the perceived affordance, such as the handle of a door or the physical structure of a switch. (Norman 2013: 18)

Feedback is a well-known concept in information science. Norman views it as an important form of communication that must be given after any action in order to confirm the action and communicate its results. Feedback must be immediate and informative. Norman (2013) holds that poor feedback can be worse than no feedback at all, because "it is distracting, uninformative, and in many cases irritating and anxiety-provoking" (p. 24). But the solution is not to resort to information overload as "too much feedback can be even more annoying than too little" (p. 24). From this perspective, feedback must be planned and prioritized. Unimportant information must be presented in an "unobtrusive fashion", whereas important signals should be "presented in a way that does capture attention" (p. 25).

Norman (2013) stresses several times that the human-centered design process starts with a good understanding of people and their needs, i.e. "the needs that the design is intended to meet" (p. 9). How should this understanding be obtained? Norman recommends that it is done primarily through *observation*. His argument in favor of this method is that "people themselves are often unaware of their true needs, even unaware of the difficulties they are encountering" (p. 9). To design based on the results from observation, he advocates the use of *iterative processes with test-driven development*:

[...] the HCD principle is to avoid specifying the problem as long as possible but instead to iterate upon repeated approximations. This is done through rapid tests of ideas, and after each test modifying the approach and the problem definition. The results can be products that truly meet the needs of people. (Norman 2013: 9)

The design of modern technological products requires interdisciplinary collaboration. Such collaboration, however, is far from trouble-free. Experts from different fields tend to have different goals and principles and consider their specific contribution to the final product the most important:

One discipline argues that it must be usable and understandable, another that it must be attractive, yet another that it has to be affordable. Moreover, the device has to be reliable, be able to be manufactured and serviced. It must be distinguishable from competing products and superior in critical dimensions such as price, reliability, appearance, and the functions it provides. Finally, people have to actually purchase it. It doesn't matter how good a product is if, in the end, nobody uses it. (Norman 2013: 35)

Norman believes that "everybody is right", and that the successful product has to meet all these requirements. This requires a strong sense of business, great management skills, and the ability to coordinate the many, separate disciplines. The different experts must be convinced to "understand the viewpoint" of the other experts and "to think of the design from the viewpoint of the person who buys the product and those who use it, often different people" (p. 35).

2. Design of dictionaries in the digital era

Dictionaries are utility tools. For many people, they have become everyday things. The principles of human-centered design are therefore relevant and applicable also to this group of objects. It is no secret that dictionaries vary considerably in terms of quality, and it is not difficult to find *Norman dictionaries*. Many printed dictionaries, for instance, are not sufficiently focused on the particular user needs they intend to meet. Instead, they apply a sort of scattergun technique based on the old paradigmatic idea that the compilation of a dictionary is only realistic if "it takes into account the likely needs of various users in various situations" (Hartmann 1989: 104). In addition, space constraints frequently compel lexicographers to negotiate quality in some aspects at the expense of shortcomings in other aspects (see, e.g., Rundell 2015a). Labels like *Norman dictionaries* are uncalled for in this context.

The digital technology changes the rules. It opens completely new horizons and poses new challenges as well. The capacity to store lexicographical data is now unlimited whereas the presentation of these data is subject to new types of constraints. The temptation of resorting to data overload, for instance, should be resisted (see Gouws and Tarp 2017). Lexicographers have to adapt to the new reality. They must develop observational skills to detect not only the

foreseen users' information needs but also their "capabilities and ways of behaving" in the digital environment. The determination and inclusion of lexicographical data to satisfy the mentioned needs are paramount. Yet, it is clearly not enough. Users should also be able to consult the dictionary intuitively, i.e. to find these data and retrieve the needed information as fast and easy as possible and without any instructions. Moreover, the entire consultation experience should be "delightful and enjoyable" in terms of aesthetics and interaction between user and device. To provide such an experience, interdisciplinary collaboration becomes a necessity as Norman rightly states.

In the following, we will briefly look at various online dictionaries. Our aim is not to analyze these dictionaries in all details or describe the current state-of-the-art but to highlight some positive aspects as well as challenges still to be solved. Today, users typically access online dictionaries using bookmarks, apps, or searches on the Internet. Basically, it is quite similar to the old days when they took the printed dictionaries down from the shelves or went to the library to get them. In both cases, the users are aware that they have a reference tool in front of them. In this respect, dictionaries do not need to call attention to themselves. But they do need to make their affordances visible to the users.

In "traditional" online dictionaries, the consultation process starts in the search bar, i.e. the box that affords the search. Netizens and other navigators on the Internet are familiar with search bars and do not need additional signifiers. Nonetheless, many dictionaries add such signifiers, for instance, the *Oxford Dictionary on Lexico* (see Figure 1). The text *Type word or phrase* is presented in subdued color. As such, it is unobtrusive and less annoying for frequent users.



Figure 1: Search bar in *Oxford Dictionary on Lexico*

The down-pointing arrow in the dark field to the left allows the user to choose between six mono and bilingual dictionaries with English and Spanish as well as two grammars of the respective languages. In the upper right corner of the default page (not shown in Figure 1), users also have the option to change the site language from English to Spanish. Other dictionaries have similar functionalities. *Cambridge Dictionary*, for instance, allows its users to choose between 16 different site languages and 26 dictionaries plus one grammar, in the last case by clicking on the signifier ":" (see Figure 2).

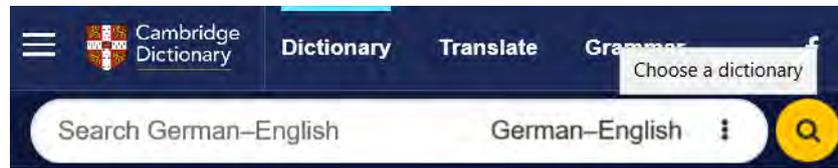


Figure 2: Search bar with the option to choose dictionary in *Cambridge Dictionary*

The search bar in *Lexico* has other affordances that users can discover by means of signifiers. A click on the keyboard icon visualizes a keyboard with special letters whereas a click on the magnifying glass initiates the search process. Once the user starts typing, the signifier *Type word or phrase* disappears and the typed letters appear in normal black color (see Figure 3). Besides, a box with lemmatized word terminations immediately visualizes together with a new button (X). The former allows the users to access one of the suggested words with a simple click whereas the latter allows them to reset the search query and start from scratch. These affordances and signifiers have become conventional and are used in many online dictionaries and search engines like Google. When applied correctly, they undoubtedly guarantee a pleasurable beginning of the consultation process.

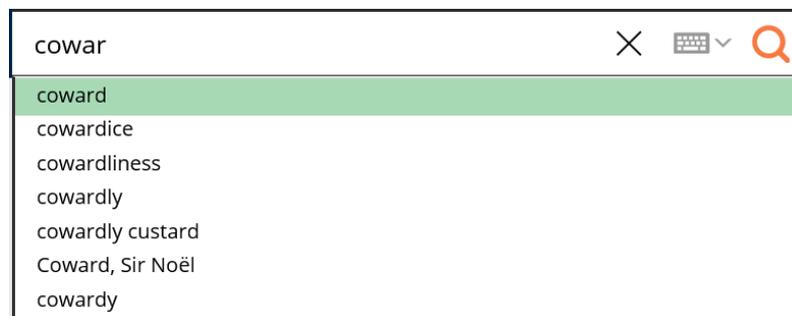


Figure 3: Search bar in *Lexico* with lemmatized word terminations and reset button

As Norman stated, good communication is important, especially when something goes wrong. Two main problems may occur: lemma lacunae or misspelling of the search query. In both cases, the users need a response. Let us take the archaic adjective *specifical* as an example. In various dictionaries like *Lexico*, *Cambridge*, *Longman*, and *Macmillan*, the users are informed that the search for *specifical* did not yield any result. Instead, several alternative spelling

candidates are suggested, among them the modern variant *specific* as can be seen in Figure 4.

Sorry, no search result for **specifical**.

DID YOU MEAN:

[specifically](#)

[specific](#)

[specifics](#)

[specificity](#)

[special](#)

[specified](#)

[specifies](#)

[specification](#)

[specifications](#)

[apolitical](#)

Figure 4: Response to the query "specifical" in *Macmillan*

This kind of feedback is not bad, but neither is it perfect. The underlying problem is the design and the challenges it poses to the users. The correct spelling variant has to be identified and clicked on in order to access the article *specific*. This takes some time and may create anxiety among some users. It would probably be the only possible solution if it were a case of misspelling. But here it clearly is a question of a lemma lacuna even though *specifical* is old fashioned. As can be seen in Figure 5, *Merriam-Webster* has opted for a different method. In this dictionary, *specifical* has its own article where the users are informed that the word is archaic (= should not be used) and means *specific*. Even so, if they do not know the meaning of *specific* or how to use it, they will have to click on the word to access the corresponding article. In both cases, a more human-centered solution with improved communication would have saved the extra click and prevented the possible anxiety.



Figure 5: The article "specifical" in *Merriam-Webster*

The next phase in a successful consultation process is the display of a dictionary article. Figure 6 shows such an article from the *Diccionario de la lengua española* edited by the Royal Spanish Academy. Although this dictionary has a number of relevant affordances and signifiers to assist the search process, the resulting article is disappointing. Like the monarchy, it seems like something from the past. The compact definitions and metatexts are clearly influenced by traditional conventions and space restrictions in printed dictionaries. Even for native speakers, it is difficult to grasp the meaning of the different senses. Grammatical abbreviations like *U. t. c. intr.* and *prnl. p. us.* are simply hogwash for most users. This *Norman article* represents a general problem and is typical for many online dictionaries that are based on lexicographical databases designed for the printed book format. The challenge is not to redress these dictionaries in digital bobbin laces but to rethink the whole concept from scratch starting with the databases. Apart from user needs and behavior, the design has to accommodate the foreseen user group's capabilities which seem to have been completely ignored in this case.

consentir Conjugar

Del lat. *consentīre*.
Conjug. c. *sentir*.

1. tr. Permitir algo o condescender en que se haga. *U. t. c. intr.*
2. tr. Mimar a los hijos, ser muy indulgente con los niños o con los inferiores.
3. tr. **creer** (|| tener algo por cierto).
4. tr. *Der.* Otorgar, obligarse.
5. tr. *Der. Arg.* Acatar una resolución judicial o administrativa sin interponer contra ella los recursos disponibles.
6. tr. *p. us.* Dicho de una cosa: Soportar, tolerar algo, resistirlo.
7. *prnl. p. us.* Dicho de una cosa: Resentirse, desencajarse, principiar a romperse. *El buque se consintió.*

Figure 6: Article from the Dictionary of the Royal Spanish Academy

Cambridge and Oxford are among the publishers of dictionaries that have advanced most towards a human-centered digital design. Figure 7 shows a screenshot of the article *consent* in Oxford's *Lexico*.

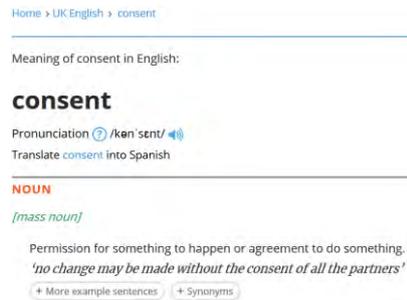


Figure 7: Extract of article from *Lexico*

The full article treats both noun and verb, two so-called phrases, as well as etymology. It also includes a large amount of hidden text (examples sentences and synonyms). Especially the former are very voluminous. If these data categories were displayed as default, it would be necessary to scroll down several pages to get an overview of the complete article. This would probably challenge many users and, in some cases, lead to anxiety and abortive consultation. From this perspective, the design is aesthetic and user-friendly. This also holds true for the two types of signifiers (+ *More example sentences* and + *Synonyms*). It is easy for the users to see how the hidden text can be visualized, and it is also easy to see how it can be hidden again. As can be seen in Figure 8, the signifier is now placed on a dark background color, and plus (+) has changed to minus (-). There is little room for misunderstandings.



Figure 8: Extract of article from *Lexico* with synonyms expanded

Synonyms are particularly useful in connection with text production. A major problem is here to find their exact meaning and choose the most appropriate candidate to be used in a given context. Users may, therefore, be tempted to click on *View synonyms* to get more information on meaning and usage. The disappointment will be big. They are referred directly to a classic thesaurus with more synonyms (and antonyms) but no definitions (see Figure 9). It is a blind alley. If they want definitions of the listed synonyms, they will have to start a new consultation and look them up one by one, a very time-consuming

operation that takes focus away from the writing process. The good news, however, is that they can opt to return to the previous page (Figure 8) by simply clicking on *consent* in *See definition of consent*. The possibility to close an expanded text or return to a previous page is another fundamental characteristic of good human-centered design.

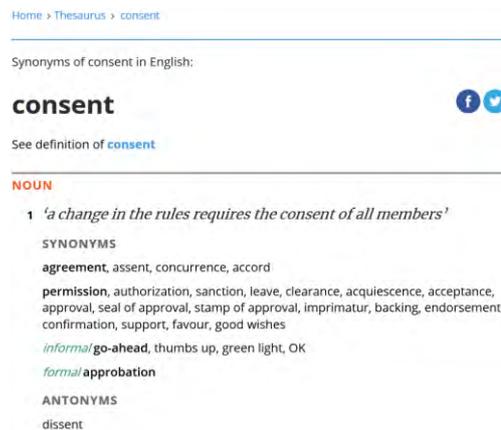


Figure 9: Extract of article from *Lexico* with synonyms

Cambridge, which has a design very similar to *Lexico*, misses this last option. When its users click on *See more results* in the thesaurus, they are directed to a page where there are no perceivable affordances that allow them to return to the previous page. By contrast, each of the synonyms and related words listed on this page affords an additional click that provides access to a definition, a very useful piece of information as mentioned above. The dark side of the coin, however, is that the users who follow this path once more end up in a remote area with no return ticket. Not even the best reference skills can bring them back. It is the design that fails. The database contains the required data but they are not presented and connected in a user-friendly way. If users want to know the meaning of various synonyms to pick up the most suitable one, the only option is to start from scratch and conduct a series of new consultations. The entire experience is anything but "delightful and enjoyable". In this respect, a marriage between *Cambridge* and *Oxford*, a lexicographical Oxbridge, would be an important step forward towards a more human-centered design of online dictionaries.

In the previous paragraphs, we have seen a few examples of how the most prestigious dictionaries are adapting to the online environment and introducing interesting solutions. But there are also serious challenges to be solved before they can claim to fully comply with the principles of human-centered design. The challenges seem to have two main sources: a hangover from the print era and problems with the business model.

Many databases sustaining online dictionaries still have their origin in the print era, where the "exclusion criteria" dominated due to space restrictions (Rundell 2015a: 312). The problem is not only ultra-short definitions and obscure abbreviations like the ones we saw in the dictionary from the Spanish Academy. Equally serious are all the possible lemmata, senses, and remaining data that have been discarded by the exclusion criteria. It takes time to remedy these lacunae. It entails a mental awakening and frequently also the design of completely new lexicographical databases. As long as the "various users in various situations" cannot find the lexicographical data they need in these situations, online dictionaries cannot be classified as fully human-centered.

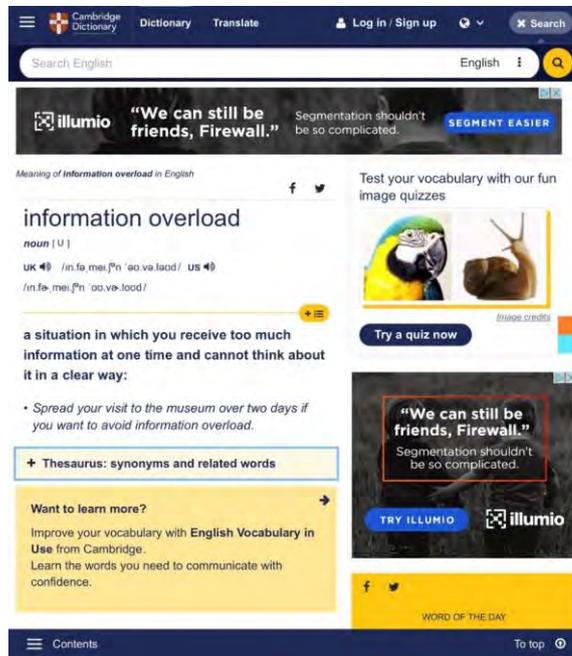


Figure 10: Dictionary article with advertisements from *Cambridge Dictionary*

The other problem is the business model. Due to financial constraints, all the online dictionaries consulted above, at least in their free-access editions, have "decorated" the pages with click-through banners and advertisements for themselves or other companies. An example is the screenshot from *Cambridge Dictionary* shown in Figure 10. The decoration may guarantee some revenue, but it may also divert the users' attention and give them a far from pleasurable experience. The Royal Spanish Academy has even honored its dictionary sponsor, one of the country's biggest banks, in the very search bar (see Figure 11). The obtrusive font sizes and colors are like vinegar in the morning coffee, a bad way to start the day.



Figure 11: Search bar in the Dictionary of the Royal Spanish Academy

A pleasurable and aesthetic dictionary experience requires, as a minimum, that all unnecessary "noise" is removed allowing users to exclusively focus on the lexicographically relevant items. One of the few providers of online dictionaries who have understood this is *Ordbogen.com*. Figure 12 shows the article *consent* from its monolingual English dictionary that has been licensed from Random House. If we abstract from the content and structure of the displayed article, we can see an uncomplicated design characterized by simplicity. The article is the central issue. In addition, we find the company logo, site languages, login icon, search bar, license reference, access button to all dictionaries, feedback icon (bulb), and language and technical support icon (speech bubbles). All of this presented in a balanced and unobtrusive way with the dictionary article as the main attraction. This is everything needed to make full use of this page. Only the content of the dictionary article calls for improvement in various aspects but the roots of this last problem have already been uncovered above.



Figure 12: Design of page with dictionary article in *Ordbogen.com*

3. Design of integrated e-reading dictionaries

By integrated e-reading dictionaries, we understand any dictionary that provides direct and immediate assistance to text reception in digital devices like e-readers, tablets, smartphones, and laptops. The benefit of such dictionaries is that readers can access them by simply touching or clicking on a word in the text they are reading. This saves time and is less disturbing for the reader who may be very focused on the content of the text. In the following, we will use the iPad to discuss some features and requirements for the design.

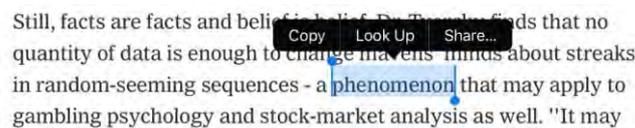


Figure 13: Touching a word on the screen

Figure 13 shows what happens when the reader touches a word in an article from the *New York Times* uploaded on an iPad. A small box with three options immediately pops up, one of them inviting the user to start a lexicographical consultation. In the figure, we see two lexicographically relevant affordances: The first is the individual words in the text that afford a touch in order to access the box, and the second is the central section of the box that also affords a touch, in this case to start the real consultation process. However, only this second affordance is perceivable. If the user does not know that a touch on the screen can provide access to further data, nothing will happen. In fact, one of the authors of this article had used his iPad several years before he discovered this functionality. The information can be found in the extensive user manual that can be consulted online, but how many users read such manuals? The designers should, therefore, have added a signifier making the affordance discoverable, for instance: "Touch a word to get a definition". Of course, it can be annoying for the users to see this message repeated every time they read a text in one of the applications that allow this functionality. A solution could therefore be to add "Don't show this message again", a standard text already used in similar cases.

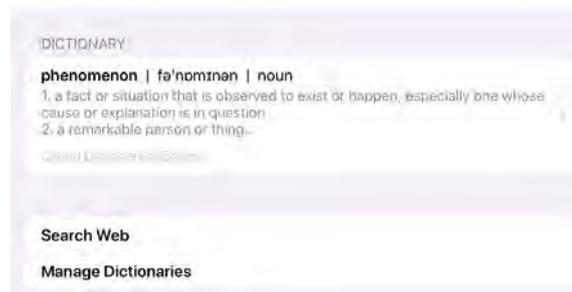


Figure 14: Dictionary overview: phenomenon

A click on "Look Up" will take the readers to a page with three options (see Figure 14). First, there is a link to a directory where the users can activate and deactivate the available dictionaries at their choice. Second, there is the possibility to search for more information about *phenomenon* on the Internet. And third, there is a list with extracts from the dictionaries that are activated and can be accessed (in this case only one dictionary has been selected). The use of perceivable affordances and signifiers makes it easy to figure out which actions can be performed. The access route is laid bare. Noticeable are the subdued colour of the dictionary extract and the two conventional signifiers attached to it: The sign "..." at the end of the extract is used to indicate that more text follows, whereas the small right-pointing arrow in the right margin signals how to access this text (by clicking or touching). This is an important piece of communication that allows the readers to continue the consultation.

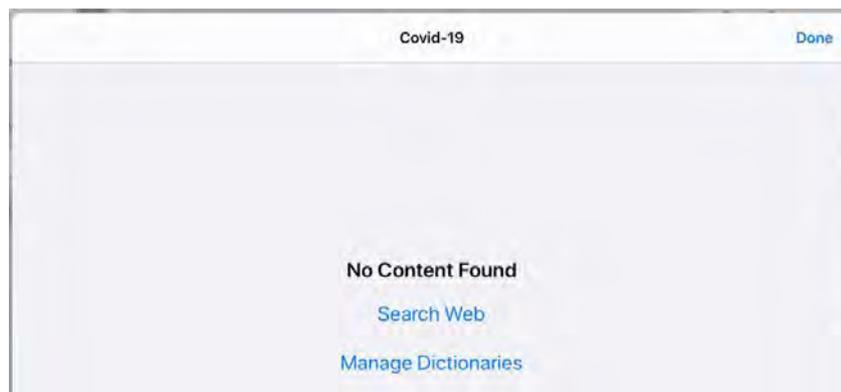


Figure 15: Feedback when no content is found

Before discussing the article we are directed to, we will briefly see what happens if a word is not included in the dictionary (see Figure 15). The term *Covid-19* taken from another article in the *New York Times* is a good illustration. After touching the screen, a page is visualized confirming the word consulted (at the top of the page), but informing us that no content has been found. Instead, it suggests two alternative actions: either to search on the Internet or to activate a different dictionary. The actions can be performed by simply touching the blue letters. This is an excellent example of good communication and the immediate feedback required by Norman (2013: 9) "where things go wrong", in this case due to a lemma lacuna.

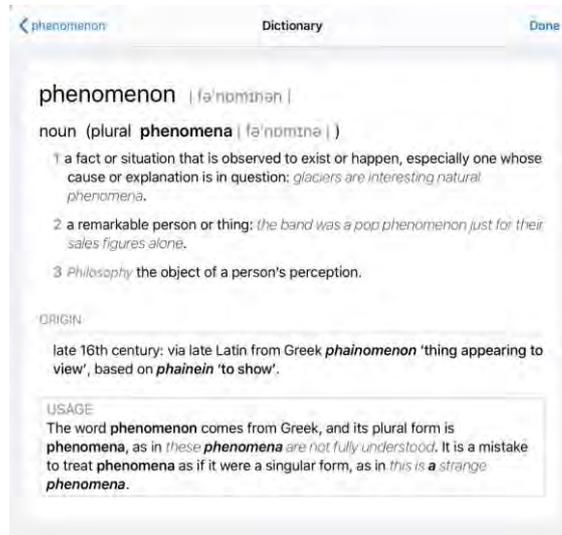


Figure 16: The article of *phenomenon* as presented in the iPad

Figure 16 shows the full article of *phenomenon* taken from the Oxford Dictionary of English. In terms of affordances and signifiers, the presentation appears perfect. If the users are happy with the information retrieved, they can easily end the consultation by clicking on "Done" in the upper right corner. And if they want to explore one of the other activated dictionaries and return to the previous page, they just need to click on the left-pointing arrow in the upper left corner. No special reference skills are needed so far.

However, in terms of "the needs that the design is intended to meet" (Norman 2013: 9), the article shown in Figure 16 is disappointing because the features do not match the requirements of the users. The designers seem to have forgotten that its genuine purpose is to assist the reading of a text. More than half of the lexicographical data included are superfluous in this respect. This applies to etymology, however interesting it is, as well as pronunciation (repeated twice) and usage as the consultation is not aimed at producing oral or written texts.

As Norman (2013: 24) comments, this kind of data overload can be "annoying", but it may also have more serious consequences. The treatment of *phenomenon* is relatively simple. But what about an orthographic word like *water* that belongs to two different word classes? If readers, for instance, touch the verb *watering* on the screen, they will be directed to an overview page where only the treatment of the noun is shown (Figure 17). This may create some confusion. Some users may try to activate other dictionaries (all of which are published by Oxford), but the result will be the same. Other users may immediately touch the arrow and go directly to the expanded article. However, here they will only find the various senses of the noun *water* on the default

page, and they will have to scroll down to find the corresponding verb and the specific meaning they are looking for. Apart from being unnecessarily time-consuming, this kind of structuring and relative data overload may, in the worst case, lead to an "obstructed" consultation that is "finally aborted with no result" (Gouws and Tarp 2017: 297)

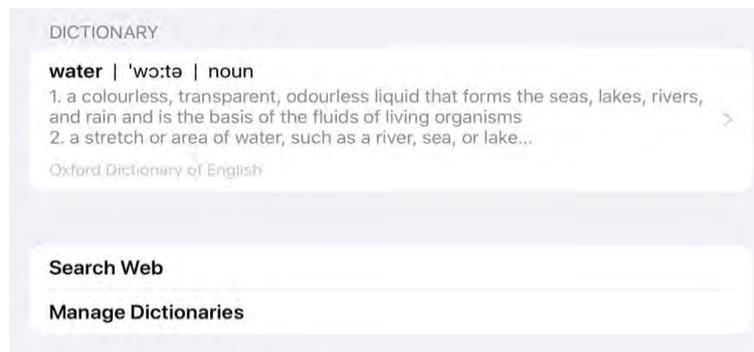


Figure 17: Dictionary overview: *water*

How can this sad ending to a happy journey be explained? We can only guess. It may result from a blurred definition of the users' real needs in a particular context (reading). The database sustaining the Oxford Dictionary of English may not, although it seems inconceivable, permit the extraction of the data categories required for the iPad. An explanation could also be that Apple, without having second thoughts, has just licensed this and other Oxford dictionaries due to their prestige and well-known quality. In any case, it is definitely a question of bad project management and lack of healthy interdisciplinary collaboration. It is a broad hint that lexicographers urgently need to dig in their own backyard, engage in interdisciplinary collaboration, and be better to "understand the viewpoint" of other experts and "think of the design from the viewpoint of the person" who is going to use it (Norman 2013: 35). The problem here is clearly not the users' reference skills but a design that is not sufficiently focused on their needs in a particular situation.

4. Design of lexicography-assisted writing assistants

Human-centered design must be understood in a broader perspective. Lexicography is no exception. Even if the target users' "needs, capabilities, and behavior" seem to have been properly accommodated in the design of a digital dictionary, there can be no absolute guarantee that they will pick up the right information in a concrete consultation. Some users may still erroneously feel that the information obtained meets their specific needs, and then be unaware of the difficulties they face when applying it in a concrete context. This is a well-

known phenomenon which most teachers of non-native languages may have observed hundreds, if not thousands, of times. Learners who consult traditional dictionaries may, for instance, unintentionally have chosen the wrong equivalent among several possible candidates. This is one of the arguments in favor of teaching reference skills. But is this really the best solution? And does it work at all in these complex cases?

In the last instance, the underlying problem is, obviously, that the properties of the dictionary consulted do not sufficiently cater for the users' capabilities and behavior. From a narrow lexicographical perspective, a solution would be to introduce additional data into the dictionary, for instance, more explicit definitions of the respective senses and equivalents followed by a larger number of context examples. However, this would probably have negative collateral consequences such as data overload, more complex access routes, longer consultation time, bigger workload for the lexicographers, and higher production costs. So, what could the alternative be?

One option is to take a broader perspective on dictionary usage and perform a so-called *root cause analysis*. This implies that the overall activity is separated into goals and subgoals by persistently "asking 'Why?' until the ultimate, fundamental cause of the activity is reached" (Norman 2013: 42). Norman provides an example: Somebody buys a quarter-inch drill, not because they want a drill, but because they need a quarter-inch hole. However, "perhaps they don't really want the hole, either; they want to install their bookshelves" (p. 44). Hence, the hole is "an intermediate goal", whereas the real goal is to install shelves that can be used to store books. In this connection, he philosophizes:

Most innovation is done as an incremental enhancement of existing products. What about radical ideas, ones that introduce new product categories to the marketplace? These come about by reconsidering the goals, and always asking what the real goal is. (Norman 2013: 43)

In the above example, a reconsideration of the ultimate goal could lead to "methods that don't require holes", or e-books "that don't require bookshelves". The trick is "to develop observational skills" to detect these goals (p. 43).

Now, what is the ultimate goal of a dictionary consultation? Nesi (2015: 584) observes that people usually "are doing something else" when they resort to dictionaries. The function theory (Tarp 2008) that subdivides this "something else" into reading, writing, translating, and learning, among others, may assist us to further develop this idea. In the previous section, we discussed problems related to reading. Let us now take writing in a second language as an example.

The information retrieved from dictionary consultation during the L2-writing process is not a goal in itself, it is a subgoal. To find the real goal, we first need to ask why the person writes the text, and then we should figure out what kind of text is required. Is it a job application, a love letter, a technical report, an academic article, or any other text genre? An application, for instance, usually serves the purpose of getting a job and earning some money

that can sustain the family. Hence, the "ultimate, fundamental cause of the activity" is to sustain the family. Instead of offering an "enhanced dictionary" to assist the job application, a "radical idea" could, thus, be to "reconsider the goal" and pay the applicant's family some money. This may be a wise step in specific cases, but it goes far beyond the scope of lexicography. The ultimate lexicographically relevant goal appears to be the writing of an L2 text belonging to a specific genre.

A "radical idea" would then be to focus on this goal and develop a tool that offers more comprehensive writing assistance than traditional dictionaries. The tool should also be context-aware so its users are placed in a much more favorable situation when they have to decide which word to use in a concrete context. Tarp et al. (2017) have described such a tool (*Write Assistant*) that works integrated into the software people typically use when they write on laptops, smartphones, and tablets. Like other digital writing assistants, *Write Assistant* is driven by a language model that has been trained on a corpus applying statistic programming and increasingly also deep learning. This technology makes it possible — based upon the previously written text — to suggest the words most likely to be the next in a concrete context. Existing writing assistants with these characteristics are mostly monolingual, but the one mentioned is bilingual. It does not only allow non-native writers to type L2 words but also L1 words if they have doubt about the words that can be used to express a particular idea in L2. As illustrated in Figure 18, the tool will then automatically look up in a lexicographical database, pick up the L2 equivalents, and present them in a prioritized order with the most likely candidates listed first.

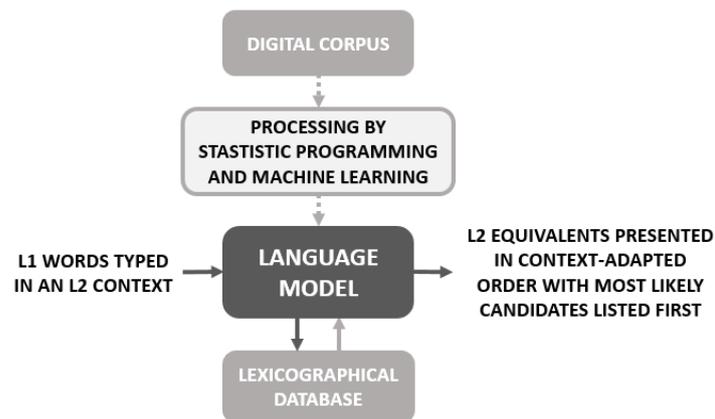


Figure 18: Schematic overview of an L2-writing assistant

Thus, a new principle has been added in the long-lasting lexicographical discussion of the most convenient ordering of senses and equivalents, namely the

principle of automatic, context-aware ordering that is unique to each consultation (see A and B in Figure 19).

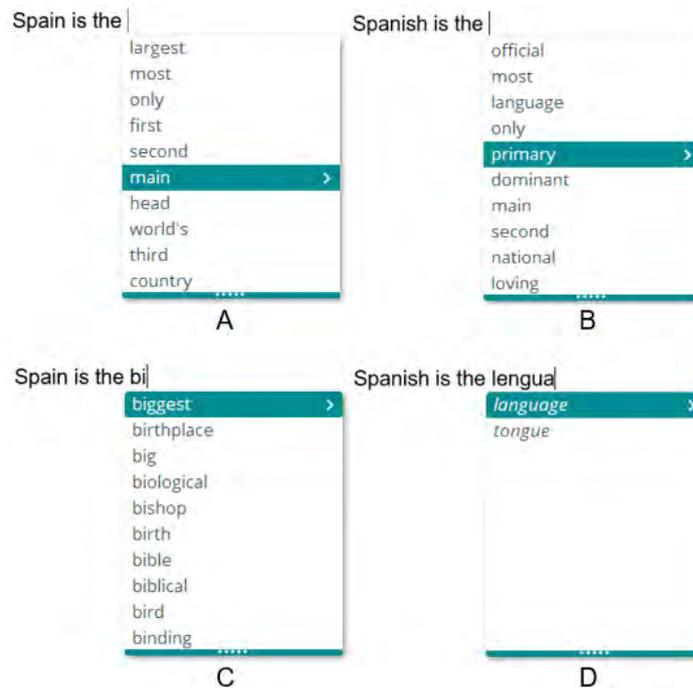


Figure 19: Some functionalities. A & B: context-adapted suggestions; C: likely word terminations; D: L2 equivalents (in italics)

As mentioned above, learners are often unaware of the mistakes they make when using equivalents in a concrete context. L2-writing assistants like the one described have the potential to considerably reduce this risk. If learners have doubts about the meaning of any of the suggested L2 words, they can activate the integrated dictionary by a simple click or touch on the screen, that is, without leaving the document they are working on. Although the technology can still be improved, the design with its affordances and signifiers is, up to this point, generally convincing from a human-centered perspective (see Figure 19). The suggestions for word terminations, next-words, and equivalents are automatically generated by the tool and presented to the users in an aesthetic design when the latter start typing. The green bar indicates how to navigate between the respective suggestions, and the small right-pointing arrow invites the users to start a lexicographical consultation. So far so good. Now the trouble starts. The users who continue to the integrated dictionary will experience something similar to what happened on the iPad.

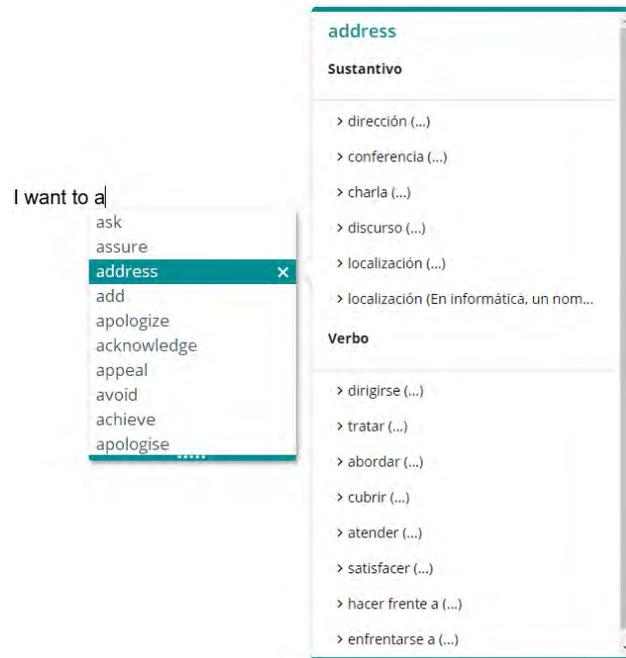


Figure 20: Consultation of *address*: default page

The overall design is still aesthetic and functional with the use of pertinent signifiers like the sign "... " and right-pointing arrows signaling how to proceed (see Figure 20). A slightly darkened background indicates when an item is open and a down-pointing, green arrow shows how to close it again (see Figure 21). Feedback is given when there is a lemma lacuna. Yet, the lack of relevant lexicographical data is obvious. The users' needs in terms of L2 writing are clearly much more complex than the ones related to L1-text understanding discussed in the previous section. Various types of lexicographical data are required to give maximum assistance to non-native learners: part of speech, inflection, L1 definitions of L2 words, equivalents, cultural and pragmatic notes, synonyms and antonyms, syntactic properties, collocations, and example sentences, among others. Most of these central data types are completely missing. The only data provided to the users are part of speech, equivalents, and example sentences. This is clearly not enough. Cumulative equivalents, for instance, are helpful to text understanding, but insufficient to L2 writing as many of them are polysemous and it is difficult for users to figure out which of the senses correspond to the meaning they want to express. Besides, they will not be less confused by the fact that all example sentences addressed to the eight verb senses in Figure 21 are identical, except for the first one (*dirigirse*). Regretfully, this is no exception to the rule.



Figure 21: Consultation of *address* with four expanded verb senses

According to Fuertes-Olivera and Tarp (2020), who have analyzed the lexicographical assistance offered by *Write Assistant*, the problems begin in the existing lexicographical databases:

These databases were originally compiled to sustain printed or digital dictionaries. The experience so far indicates that they are less suited to feed a tool like *Write Assistant*. They may not contain the required data types. They may not have them in the necessary quantity, or they have stored them in such a way that they cannot be used properly. In any case, it seems necessary to restart almost from scratch.

The two authors have come up with an alternative proposal based upon the bisopical English–Spanish *Valladolid-UVa Database* (under construction); see Fuertes-Olivera et al. (2018). The proposal aims at avoiding frustrating phenomena such as long access routes, data overload, scrolling down, and incomprehensible metatexts, all of which may lead to mistakes and abortive consultations. It is being gradually incorporated into *Write Assistant* *pari passu* with

the expansion of the database. The new articles do have definitions. But it takes time to complete such work.

We started this section stressing that "there can be no absolute guarantee" that users will retrieve "the right information in a concrete consultation". This is true in the sense that writers always have the sole responsibility for the final text. Irresponsible behavior may still lead to mistakes. But writing assistants can go a long way to reduce the risk. The technology has far from exhausted its potential. The enhanced application of artificial intelligence, for instance, will beyond any doubt make allowance for improved collocational, syntactic, stylistic, and genre-adapted assistance (see Tarp 2020). The service can be proactive and even furnished in the form of alerts. In this way, writing assistants with human-centered design can accommodate their users' "capabilities and behavior". The Achilles heel is the lexicographical component that, so far, takes insufficient account of the users' needs in connection with L2 writing. Hence, we can conclude that the fundamental problem is not the users' reference skills but the lexicographers' ability to determine and produce the data required to meet the latter's needs.

Cutting-edge technology does not make lexicography obsolete. But it implies interdisciplinary collaboration as well as a qualitative shift from the traditional stand-alone dictionary to an integrated lexicographical product that is context-aware and user-centered.

5. Towards a new lexicographical culture

The notion of *reference skills* must be seen from a historical perspective. Hausmann (1989: 13) observes that the history of lexicography shows a strained relationship between the discipline and society. According to him, the terms *dictionary culture* and *user-friendliness* are used to describe this friction. User-friendliness implies that lexicography adapts to society whereas dictionary culture means that society adapts to lexicography. User-friendliness prevails when dictionaries are made from which the intended target users can retrieve the kind of lexicographical information they require. By contrast, a dictionary culture prevails when lexicographers know the target users have acquired the necessary reference skills to successfully consult their dictionaries. The conflict described by Hausmann implies a complementary relation where users have to complement the efforts of the lexicographers to ensure successful dictionary usage. The lexicographers make the dictionaries, and the users have the responsibility to enable themselves to find and retrieve the required information from the condensed and often strongly codified items as the ones we saw in the dictionary from the Royal Spanish Academy (Figure 6).

The discussion has divided the waters between lexicographers in favor of one approach or the other. Gouws (2016), for instance, defends a so-called "comprehensive dictionary culture" that includes both the traditional concept of dictionary culture and that of user-friendliness. At the other extreme, we

have lexicographers like Yamada (2014) who still promotes the teaching of reference skills. In a review of Yamada's *Oxford Guide to the practical usage of English monolingual learners' dictionaries*, Rundell (2015b: 27) writes:

Yamada believes that, when a user's search for information is unsuccessful, "either the dictionary or the user is to blame". My default position is that if users can't readily find what they are looking for, the fault lies squarely with the dictionary. ... Few students will be fortunate enough to have a teacher who understands dictionaries as well as the author of the guide. In most cases, they must rely on their dictionary being well enough designed to make its use intuitive.

We fully agree with Rundell, especially when he stresses that users should be able to use dictionaries intuitively. The concept of *intuitive use* is even more advanced than that of traditional *user-friendliness*. It may just have been a dream a few decades ago, but it is now achievable due to new technologies. We refer to a practice with this user perspective as a new lexicographical culture as the discipline is no longer concerned only with dictionaries, but also with other types of information tools. In this culture, it is the lexicographers' sole responsibility that their products can be used successfully by the target group. No special reference skills are required. Of course, users should have some basic capabilities. They should be able to read and write. They should have a certain proficiency level in at least one relevant language. And for some products like specialized subject-field dictionaries, they should also have some basic knowledge of the discipline in question. But in terms of using the device, and navigating in it, no special skills should be required. The design should be human-centered, or user-centered, in full compliance with the principles outlined by Norman (2013) and discussed in the previous sections.

The new culture places more responsibility on the shoulders of lexicographers in terms of needs detection as well as data preparation and presentation. They should refine their observational skills to detect real user needs, as users themselves may not be aware of these needs. The initial observation should, therefore, not be performed when users are using lexicographical devices, but in the situations prior to the consultation process when the needs occur. The lexicographical data that can meet these needs should then be prepared meticulously and stored in well-designed databases that contain as much relevant data as possible. The lexicographical data are the central issue and should be high-quality. This implies, apart from giving response to the detected user needs, that definitions and metatexts are easily understandable and written in a plain, user-oriented language.

Finally, the data should be presented to the users in a way that guarantees intuitive use and smooth consultation by means of appropriate techniques. This requires test-driven development and elegant use of affordances and signifiers. Users should be able to find what they need in as few clicks as possible. Scrolling down should be reduced to a minimum and should not be employed as a means to limit the number of clicks. No relevant navigation routes should

be blocked. Blind alleys and no-return situations like the ones we saw in *Lexico* and *Cambridge* should be banned. Users should be allowed to close expanded texts and pop-up windows, return to the previous page, and jump to the top after scrolling down. Data overload with overcrowded pages that may divert the users' attention and create anxiety should be avoided. Any item, data, symbol, line, or button that cannot be justified from a user perspective should be eliminated.

To achieve all this, a strong engagement with experts from other fields is required, among them information engineers, programmers, industrial designers, and professional testers. Hopefully, this will increasingly become the new lexicographical normality.

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (UID) 85434). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard.

Bibliography

A. Online dictionaries

- Cambridge Dictionary*. <https://dictionary.cambridge.org/>
Diccionario de la Lengua Española. <https://dle.rae.es/>
Longman Dictionary of Contemporary English Online. <https://www.ldoceonline.com/>
Macmillan Dictionary. <http://www.macmillandictionaries.com/>
Merriam-Webster Dictionary. <https://www.merriam-webster.com/>
Ordbogen.com. <https://www.ordbogen.com/>
Oxford Dictionary on Lexico. <https://www.lexico.com/>

B. Other literature

- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Fuertes-Olivera, P.A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.
- Fuertes-Olivera, P.A., S. Tarp and P. Sepstrup.** 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Diccionarios Valladolid-UVa. *Lexikos* 28: 152-176.

- Gouws, R.H.** 2016. Increasing the Scope of the Treatment of Specialised Language Terms in General Dictionaries. Schierholz, S., R.H. Gouws, Z. Hollós and W. Wolski (Eds.). 2016. *Wörterbuchforschung und Lexikographie*: 101-118. Berlin: De Gruyter.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Hartmann, R.K.K.** 1989. Sociology of the Dictionary User: Hypothesis and Empirical Studies. Hausmann et al. (Eds.). 1989: 102-111.
- Hausmann, F.J.** 1989: Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann et al. (Eds.). 1989: 1-19.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989. *Wörterbücher, Dictionaries, Dictionnaires. An International Encyclopedia of Lexicography. Volume 1.* Berlin/New York: Walter de Gruyter.
- Kirkpatrick, B.** 1989. User's Guides in Dictionaries. Hausmann et al. (Eds.). 1989: 754-761.
- Nesi, H.** 2015. The Demands of Users and the Publishing World: Printed or Online, Free or Paid for? Durkin, P. (Ed.). 2015. *The Oxford Handbook of Lexicography*: 579-589. Oxford: Oxford University Press.
- Nielsen, S.** 2005. User's Guides. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 137-146. Frankfurt am Main: Peter Lang.
- Norman, D.** 2004. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.
- Norman, D.** 2010. *Living with Complexity*. Cambridge: MIT Press.
- Norman, D.** 2013. *The Design of Everyday Things*. New York: Basic Books.
- Rundell, M.** 2015a. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.
- Rundell, M.** 2015b. Review Article: Shigeru Yamada. Oxford Guide to the Practical Usage of English Monolingual Learners' Dictionaries: Effective Ways of Teaching Dictionary Use in the English Class. *Kernerman Dictionary News* 23: 26-27.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen. Max Niemeyer.
- Tarp, S.** 2020. Integrated Writing Assistants and their Possible Consequences for Foreign-Language Writing and Learning. Bocanegra-Valle, A. (Ed.). 2020. *Applied Linguistics and Knowledge Transfer: Employability, Internationalization and Social Challenges*: 53-76. Bern: Peter Lang.
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Vrbinc, M. and A. Vrbinc.** 2020. User Guides in Monolingual Slovenian Dictionaries: Users' Perspective. *Lexicography* 7(1-2): 123-134.
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zürich: Bibliographisches Institut.
- Wiegand, H.E.** 1990. Printed Dictionaries and their Parts as Texts. An Overview of More Recent Research as an Introduction. *Lexicographica* 6: 1-126.

- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Volume 1.* Berlin/New York: Walter de Gruyter.
- Wiegand, H.E., S. Beer and R.H. Gouws.** 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* Handbücher zur Sprach- und Kommunikationswissenschaft 5.4: 31-73. Berlin/Boston: Mouton de Gruyter.
- Yamada, S.** 2014. *Oxford Guide to the Practical Usage of English Monolingual Learners' Dictionaries: Effective Ways of Teaching Dictionary Use in the English Class.* Tokyo: Oxford University Press.

'n Model vir 'n aanlyn GIS-vakwoordeboek

Lize-Mari Theron, *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Suid-Afrika (theronlizemari@gmail.com)*

Opsomming: As deel van 'n MA-studie aan die Universiteit Stellenbosch, is 'n model vir 'n aanlyn GIS-vakwoordeboek ontwikkel om drie probleme aan te spreek. Die eerste probleem is die groot hoeveelheid vakwoordeboeke wat gemaak is sonder die leiding van die leksikografiëorie. Dit is min dat 'n woordeboek sy potensiaal van gebruikersvriendelikheid en toeganklikheid bereik sonder die leiding van die leksikografiëorie. 'n Vakgebied wat hierdie probleem het, is Geografiese Inligtingstelsels (GIS). As daar na aanlyn GIS-vakwoordeboeke gekyk word, word die tweede probleem geïdentifiseer: die bestaande GIS-vakwoordeboeke is nie genoegsame bronne vir GIS-pedagogie in Suid-Afrika nie. Die bestaande vakwoordeboeke is nie gemik op enige spesifieke gebruiker nie. Beginners, semi-deskundiges en deskundiges se onderskeie, spesifieke behoeftes word nie in ag geneem nie. Daar is ook geen vakwoordeboek wat 'n operasionele funksie bied nie, almal fokus net op teksrepsie. 'n Verdere blik op die kommunikatiewe funksie van hierdie vakwoordeboeke, wys die derde probleem: daar is geen GIS-vakwoordeboek wat vertaalekwivalente in Afrikaans (of enige van die ander amptelike Afrikaanse) bied nie. In hierdie artikel word daar 'n model voorgestel wat ontwikkel is met die leiding van die leksikografiëorie en wat die gaping in GIS-pedagogie in Suid-Afrika kan aanpreek, want dit is bedoel vir beginners en dit het beide 'n operasionele en kommunikatiewe funksie.

Sleutelwoorde: VAKWOORDEBOEKE, LEKSIKOGRAFIËORIE, OPERASIONELE FUNKSIE, KOMMUNIKATIEWE FUNKSIE, GEOGRAFIESE INLIGTINGSTELSELS, VAKLEKSIKOGRAFIE, WOORDEBOEKMODEL, GIS-PEDAGOGIE, AANLYN WOORDEBOEKE, WOORDEBOEK-KOËPTUALISERINGSPLAN, E-LEEROMGEWINGS, HIBRIDIESE WOORDEBOEKMODEL

Abstract: A Model for an Online Specialized Dictionary for GIS. As part of a MA study completed at Stellenbosch University, a model for an online specialized dictionary for GIS was developed to address three problems. The first problem is the large number of specialized dictionaries created without the guidance of the theory of lexicography. A dictionary seldom reaches its potential of user-friendliness and accessibility without the guidance of theory of lexicography. One field of study that has this problem is Geographic Information Systems (GIS). Looking at online GIS dictionaries, the second problem is identified: the existing specialized dictionaries for GIS are not sufficient sources for GIS pedagogy in South Africa. The existing dictionaries are not aimed at any specific user. Beginners, semi-experts and experts are all treated the same without taking into account their individualized needs. There is also no specialized dictionary that provides an operational function — all of the dictionaries focus only on text reception. A look at the communicative function of these specialized dictionaries shows the third problem: there is no GIS dictionary that offers translation equivalents in Afrikaans (or any of the other official African languages).

This article proposes a model developed with the theory of lexicography that can address the gap in GIS pedagogy in South Africa, as it is intended for beginners and has both an operational and a communicative function.

Keywords: SPECIALIZED DICTIONARIES, THEORY OF LEXICOGRAPHY, OPERATIONAL FUNCTION, COMMUNICATIVE FUNCTION, GEOGRAPHIC INFORMATION SYSTEMS, SPECIALIZED LEXICOGRAPHY, DICTIONARY MODEL, GIS PEDAGOGY, ONLINE DICTIONARIES, DICTIONARY CONCEPTUALIZATION PLAN, E-LEARNING ENVIRONMENTS, HYBRID DICTIONARY MODEL

1. Inleidende opmerkings

Daar is 'n probleem in die land van die leksikografie: daar is vakwoordeboeke wat saamgestel is sonder die leiding van die leksikografieteorie. Oor die afgelope twee dekades het die hoeveelheid gedrukte en aanlyn vakwoordeboeke vermenigvuldig, maar Tarp (2012: 118) bevraagteken die gehalte van hierdie vakwoordeboeke omdat baie van hulle sonder die insette van die leksikografieteorie gemaak is. 'n Aanlyn vakwoordeboek van gehalte moet, onder andere, gemaak word met die insette van die vakspecialis, die IT-tegnikus en veral die leksikograaf (Fuertes-Olivera en Tarp 2014: 197). As enige van hierdie drie komponente ontbreek, kan dit veroorsaak dat woordeboeke nie optimaal gebruikersvriendelik en toeganklik is nie (Fuertes-Olivera en Tarp 2014: 197), wat dit ook uiteraard moeilik sal maak om die gebruiker se behoeftes te bevredig.

'n Vakgebied waarvan die vakwoordeboeke problematies is, is Geografiese Inligtingstelsels (GIS). GIS is rekenaarapparatuur en -programmatuur wat gebruikers in staat stel om geografiese data te kan manipuleer, analiseer en visualiseer. 'n GIS kan gebruik word om 'n ingewikkelde, ruimtelike vraag soos: "Waar is die beste plek om 'n boerekoöperasie in die Tankwa-Karoo te bou?" te antwoord op grond van, byvoorbeeld, geografiese data oor landbougemeenskappe, vervoerroetes, eiendom te koop, ensovoorts. GIS is 'n belangrike en relevante hulpmiddel in vandag se inligtingsgedrewe wêreld. Elke persoon wat vandag geografiese data op 'n kaart wil tipeer, kan dit doen met GIS-programmatuur wat gratis op die internet beskikbaar is. Tog is dit nie so eenvoudig nie, want GIS vereis 'n mate van kennis oor die GIS-wetenskap en ook goeie rekenaarvaardighede.

Omdat die meeste GIS-vakwoordeboeke sonder die leiding van die leksikografieteorie saamgestel is, is die probleem nou dat die beskikbare aanlyn vakwoordeboeke nie effektiewe bronne vir GIS-pedagogie is nie. Die bestaande vakwoordeboeke is, eerstens, nie gemik op enige spesifieke gebruiker nie. 'n Woordeboek moet met 'n bepaalde gebruiker voor oog saamgestel word omdat 'n enkele woordeboek nooit alle behoeftes van alle gebruikers kan bevredig nie. Tweedens vervul hierdie vakwoordeboeke almal net 'n kommunikatiewe funksie, en daar is geen aandag gegee aan 'n operasionele funksie nie. 'n Funksie

wat baie nodig is vir hierdie praktiese veld. Derdens is hierdie GIS-vakwoordeboeke ook verouderd in dié opsig dat hulle nie digitaal baie innoverend is nie. Digitale innovering is belangrik omdat dit woordeboeke in die nuwe medium (aanlyn) meer toeganklik en gebruikersvriendelik kan maak. 'n Moontlike oplossing vir hierdie probleem is 'n leksikografies gefundeerde, aanlyn GIS-vakwoordeboek.

As deel van 'n MA-studie aan die Universiteit Stellenbosch is daar 'n vakwoordeboek-model ontwikkel om hierdie probleem aan te spreek (Theron 2019). Hierdie studie het gepoog om aan die hand van die funksieteorie (wat fokus op die behoeftes van woordeboekgebruikers) en die algemene leksikografieteorie (wat handel oor woordeboekstrukture) 'n model daar te stel wat onder meer voorstelle maak in verband met die funksies, inhoud en strukture vir 'n aanlyn vakwoordeboek. Die voorstelle sorg dat die teikengebruiker 'n optimale ontrekking van inligting uit die aangebode data kan hê.

Hierdie studie het bestaan uit drie take (Theron 2019). Die eerste taak was om die gebruikersbehoefte met behulp van die funksieteorie te identifiseer. Die gebruikers is geïdentifiseer as derdejaarstudente aan die Universiteit Stellenbosch wat die Geografiese Kommunikasie-module ('n beginnerskursus in GIS) neem. Die tweede taak het die bestaande aanlyn GIS-vakwoordeboeke geëvalueer in 'n poging om goeie en swak kenmerke te identifiseer wat leiding kan bied vir toekomstige vakwoordeboeke. Die derde en laaste taak was om die bevindinge van die eerste twee take te gebruik om 'n model vir 'n nuwe, aanlyn GIS-vakwoordeboek te konseptualiseer. Hierdie artikel doen verslag van die laasgenoemde taak.

Die model wat in hierdie artikel daargestel word, is vir 'n aanlyn GIS-vakwoordeboek wat spesifiek op die studente van die Geografiese Kommunikasie-module se behoeftes gemik is. Die gebruikersgroep is die groep derdejaarstudente wat, onder andere, Engels- en Afrikaanssprekend is. Die module word in Engels aangebied, maar Afrikaanse studente het die geleentheid om werkstukke in Afrikaans te skryf. Tog is daar die vermoede dat die meeste van die studente hul werkstukke in Engels skryf omdat hulle nie die regte vertaalekwivalente vir die Engelse vakterme het nie. Die model vir die GIS-vakwoordeboek is veronderstel om die Engelse lemmata te verklaar, wat kognitiewe steun vir studente sal bied, sowel as om Afrikaanse vertaalekwivalente te verskaf sodat die Afrikaanse studente hul akademiese opdragte in hul moedertaal kan skryf. Daarom stel hierdie model 'n vakwoordeboek voor wat as 'n hibriediese woordeboek getipeer kan word, want dit bevat aspekte van beide 'n vertalende en verklarende woordeboek.

Die model weerspieël 'n uiteindelijke woordeboek wat geïntegreerd optree met die Geografiese Kommunikasie-module se studiemateriaal op die Universiteit Stellenbosch se SUNLearn-platform. Die elektroniese platform, SUNLearn, bied die moontlikheid om e-woordeboeke deel van e-leeromgewings te maak. As so 'n woordeboek geïntegreerd met die studiemateriaal optree, kan dit 'n positiewe invloed op die algemene woordeboekcultuur van die studente hê (Gouws

2013: 59). So 'n geïntegreerde aanlyn vakwoordeboek is ook gepas vir die gebruikersgroep omdat die meeste van die GIS-take op 'n rekenaar (en op SUNLearn) voltooi moet word. Die model is geïnspireer deur die aanlyn GIS-vakwoordeboek van Esri (<https://support.esri.com/en/other-resources/gis-dictionary/>), die Duitse woordeboek *elexiko* (<https://www.owid.de/wb/elexiko/start.html>) en ook die PharosAanlyn- woordeboekversameling (<https://www-pharosaanlyn-co-za.ez.sun.ac.za/tuis>).

Hierdie artikel bespreek die belangrikste aspekte van hierdie model. Die term "model" word gebruik om te verwys na daardie voorbeeldvorm, as deel van die woordeboek-konseptualiseringsplan, wat nagevolg word in die opstel van 'n bepaalde woordeboek (Gouws en Prinsloo 2005: 10). Daar sal aan die openingskerm, die toegangstruktuurelemente en die woordeboekartikels aandag gegee word. Daar sal ook na die nodige data-inskrywings, die geskikte ontwerp, die woordeboekstrukture, en digitale innovering gekyk word. Hierdie GIS-vakwoordeboek kan die gaping in GIS-pedagogie vul en dit kan ook dalk 'n positiewe invloed op die algemene woordeboekkultuur van die studente hê.

2. Die model

2.1 Die basiese ontwerp

Die voorstel is dat die GIS-vakwoordeboek deel vorm van die Universiteit Stellenbosch se studenteportaal, SUNLearn. Dit sal die vakwoordeboek meer toeganklik maak vir die studente. As die vakwoordeboek die Universiteit se naam dra, en sekere van die korporatiewe ontwerpstriglyne volg, sal die vakwoordeboek ook as betroubaarder beskou word. Hierdie ontwerpstriglyne sluit, byvoorbeeld, die gebruik van die kleure grys en maroen in. Die voorstel is ook om die vakwoordeboek in Engels te ontwikkel, aangesien dit ooreenstem met die modulewerk en die Universiteit se taalbeleid. Engels is die gekose taal van die Universiteit omdat dit beduidende nasionale, internasionale, akademiese en besigheids-gebruikswaarde het (*Taalbeleid van die Universiteit Stellenbosch* 2016: 2).

Al die afbeeldings is vir 'n rekenaarskerm ontwikkel. Die aanpassings wat vir 'n selfoonskerm gemaak sal moet word, sal nie hier bespreek word nie. Die basiese ontwerp van die vakwoordeboek is 'n webtuiste gestroop van kantpanele met net 'n stel permanente kenmerke bo-aan elke webbladsy (sien afbeelding 1).



Afbeelding 1: Die permanente kenmerke bo-aan elke webbladsy van die vakwoordeboek

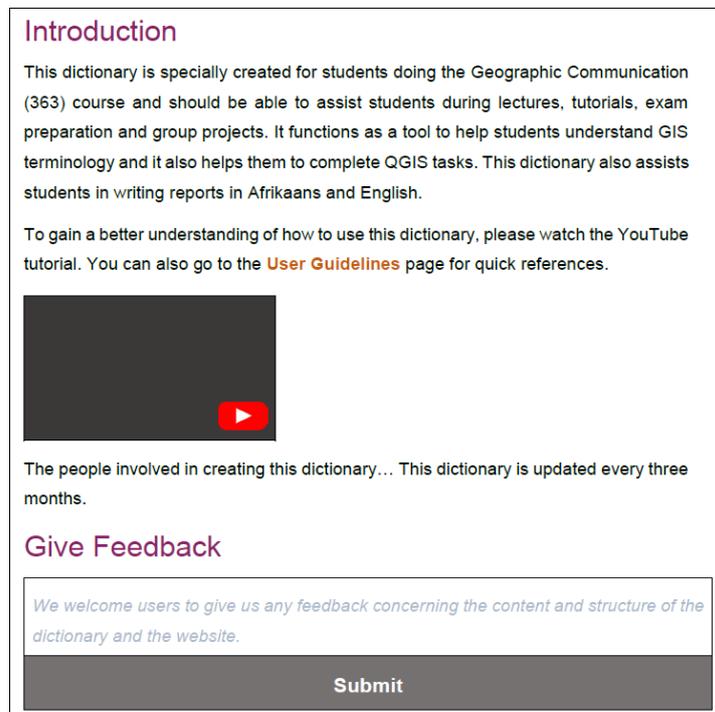
Hierdie permanente kenmerke bestaan uit (1) die vakwoordeboek se naam *Stellenbosch University GIS Dictionary*. Die naam kan in twee kleure aangebied word met "Stellenbosch University" in grys en "GIS Dictionary" in maroen. Hierdie kleurbenadering volg die voorbeeld van die bestaande studenteportaal. Die woorde "Stellenbosch University" kan 'n hiperskakel wees na die Universiteit se algemene webtuiste. Bo die naam kan (2) die SUNLearn-ikoon geplaas word. Die ikoon kan as hiperskakel dien waar die studente na die studenteportaal kan terugkeer. Hierdie hiperskakels help dat die vakwoordeboek en die ander universiteitsplatforms geïntegreerd kan funksioneer.

Onder die naam is daar (3) die drie skakels. Die eerste skakel neem die gebruiker na die openingskerm (*Home*), die tweede skakel neem die gebruiker na die primêre gedeelte van die vakwoordeboek (*A-Z*), en die derde skakel neem die gebruiker na 'n lys van buitekenmerke (*Other Resources*). Hierdie skakels is funksioneel, maar dit benoem ook die tipes data wat aangebied word en waar om dit te vind. Dit is 'n effektiewe uitleg van die dataverspreidingsstruktuur en 'n aanduiding van die data-aanbieding. Dit kan soos 'n inhoudsopgawe in 'n gedrukte woordeboek funksioneer (Gouws 2018b: 230).

Hierdie drie skakels word gevolg deur (4) die soekkassie en (5) die alfabetbalk wat deel van die toegangstruktuur uitmaak. Die soekkassie is nie 'n toegangstruktuurelement nie, maar 'n ruimte waar 'n toegangstruktuurelement geplaas kan word. Die soekkassie en die alfabetbalk werk, uit die ondervinding met die Esri-vakwoordeboek, baie effektief. Die toegangsalfabet maak voorsiening vir 'n numeriese-artikeltrajek en dit word op die alfabetbalk met 'n #-teken aangedui. Die soekkassie is met verdwynskrif gevul wat vir die gebruiker sê dat sy 'n soekwens daar kan intik. Die soekkassie is aangepas om 'n soekwens met 'n enkele klik (⊗) uit te vee. *PharosAanlyn* se soekkassie het hierdie funksie en dit werk baie goed omdat die volgende soektog vinniger geloods kan word. Die ideaal sal wees dat die soekenjin volvorme, akronieme en meerwoordige terme in Afrikaans en in Engels kan herken. Dit sal ook goed wees as die soekenjin speloute kan herken en die soekwens met die korrekte spelwyse kan voorstel.

2.2 Die openingskerm

Die openingskerm bevat die woordeboekinleiding en die terugvoerfunksie. Afbeelding 2 bied 'n uitbeelding van hoe die openingskerm kan lyk. Die woordeboekinleiding moet aandui wie die teikengebruikers is, watter gebruiksituasies die woordeboek aanspreek, en watter funksies die woordeboek kan vervul. In die inleiding kan daar gespesifiseer word wie die samestellers van die woordeboek is en ook watter rol hulle gespeel het (leksikograaf, vakspesialis en IT-technikus). Daar kan ook gespesifiseer word hoe gereeld die woordeboek opgedateer word. Om te spesifiseer wie die samestellers is en hoe gereeld die woordeboek opgedateer word, sal 'n positiewe invloed op die beeld en die gesag van die vakwoordeboek hê.



Afbeelding 2: Die openingskerm

Deel van die inleiding is ook die gebruikershandleiding. Die voorstel is om twee tipes gebruikershandleidings te hê. Die eerste is 'n video waarna die gebruiker kan kyk wanneer sy die vakwoordeboek vir die eerste keer gebruik. Die dosent kan ook hierdie video aan die begin van die module in een van die lesings speel. Die tweede gebruikershandleiding kan 'n gedetailleerde geskrewe

teks wees wat later gebruik kan word as gebruikers sukkel om te onthou hoe 'n sekere deel van die woordeboek funksioneer. Die voorstel is dat hierdie geskrewe gebruikershandleiding op sy eie webblad verskyn en dat daar 'n spesiale soekenjin is wat net op die gebruikershandleiding fokus. Dit sal verseker dat die gebruiker vinnig gehelp kan word indien sy met iets sukkel.

Die terugvoerfunksie kan onder die inleiding geplaas word. Die voorstel is dat die terugvoerfunksie uit 'n kassie bestaan waarin die gebruiker kommentaar kan tik. In die kassie kan daar 'n sin in verdwynskrif staan wat die gebruiker nooit om kommentaar oor die webtuiste en die woordeboekinhoud te lewer. Om die terugvoerfunksie op die openingskerm te plaas, sal 'n aanduiding wees van die samestellers se openlike houding om die woordeboek konstant te verbeter. So 'n openingskerm, met al die bogenoemde fasette, kan die woordeboekvaardighede en die woordeboekkultuur van die studente positief beïnvloed.

2.3 A tot Z

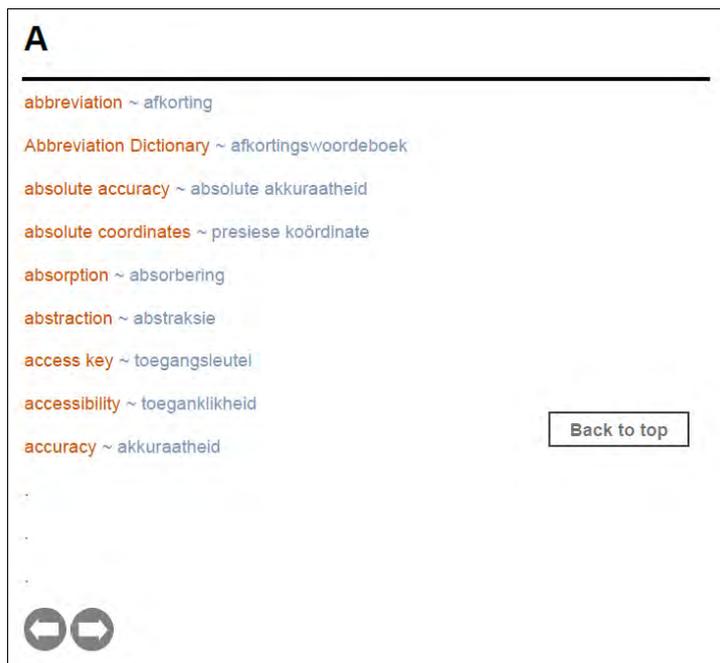
Volgens Gouws (2018b: 229) is die term "sentrale lys" nie meer van toepassing op aanlyn woordeboeke nie omdat daar nie meer noodwendig 'n "lys" is nie. Hy stel voor dat die term "primêre gedeelte" eerder gebruik word. Die A-Z-gedeelte is die primêre gedeelte van die vakwoordeboek en kan wel as 'n lemmalys gegee word. Hierdie lemmalys bevat telkens net die Engelse lemma en die Afrikaanse vertaalekwivalent (sien afbeelding 3).

Die lemma is 'n hiperskakel wat die gebruiker na die betrokke woordeboekartikel lei. Tesame met die hiperskakels word die Afrikaanse vertaalekwivalent ook gegee, want dit sal die gebruiker in sekere situasies tyd kan spaar indien sy nie na die woordeboekartikel hoef te gaan nie. Indien die gebruiker wil sien hoe om die vertaalekwivalent in 'n sin te gebruik, kan sy dan maklik na die artikel gaan. Hierdie aanbieding kan ook moontlik nuttig wees as die gebruiker nie die Engelse term kan onthou nie, aangesien die soekenjin uiteindelik ook Afrikaanse soekwense moet kan opspoor.

Die primêre gedeelte bestaan uit die onderskeie alfabetiese artikeltrajekte. Hierdie lemmalys is een manier waarop die gebruiker deur die vakwoordeboek kan soek as sy iets spesifiek wil naslaan, maar dit kan ook gebruik word as die gebruiker net vir interessantheid deur die vakwoordeboek wil "blaai". Dit is deel van die makrostruktuur, en in effek ook nog deel van die toegangstruktuur.

Die primêre gedeelte bevat ook die permanente kenmerke bo-aan die webbladsy. Aan die onderkant kan daar twee pyltjies wees wat die gebruiker kan gebruik om na die volgende artikeltrajek toe te beweeg. Daar kan ook 'n "terug na bo"-funksie wees wat die gebruiker vinnig kan terugneem na die soekkassie en alfabetbalk bo-aan die webblad. Die "terug na bo"-funksie bestaan uit 'n blokkie aan die regterkant van die skerm waarop die gebruiker kan klik. Hierdie blokkie is soos 'n swewende ballon wat teenwoordig bly selfs wanneer die

gebruiker op en af oor die teks op die skerm rol. Dit verseker dat die gebruiker altyd maklik toegang tot hierdie funksie het. Die Esri-vakwoordeboek maak gebruik van hierdie funksie en dit spaar die gebruiker tyd en moeite. Afbeelding 3 bied 'n illustrasie van hoe die A–Z-gedeelte kan lyk.



Afbeelding 3: Die A–Z-gedeelte

2.4 Woordeboekartikels

In die studie is daar 'n paar tipes data-inskrywings geïdentifiseer wat kan bydra tot die vervulling van die vakwoordeboek se funksie. Tabel 1 en 2 is 'n opsomming van al die verpligte en opsionele data-inskrywings wat geïdentifiseer is (Theron 2019: 71). Verpligte data-inskrywings is dié wat deel van die verpligte mikrostruktuur is, en die opsionele data-inskrywings vorm deel van die uitgebreide verpligte mikrostruktuur. Die verpligte mikrostruktuur bevat al die data-inskrywings wat in ál die artikels belangrik is en die uitgebreide verpligte mikrostruktuur maak voorsiening vir data-inskrywings wat net in sommige artikels verstrekk word. (Gouws en Prinsloo 2005: 141). Die artikel-inskrywings bestaan uit aanduiders wat funksionele data-inskrywings is, en struktuurmerkers wat gebruikers help om tussen die verskillende aanduiders te onderskei (Du Plessis 2016: 56).

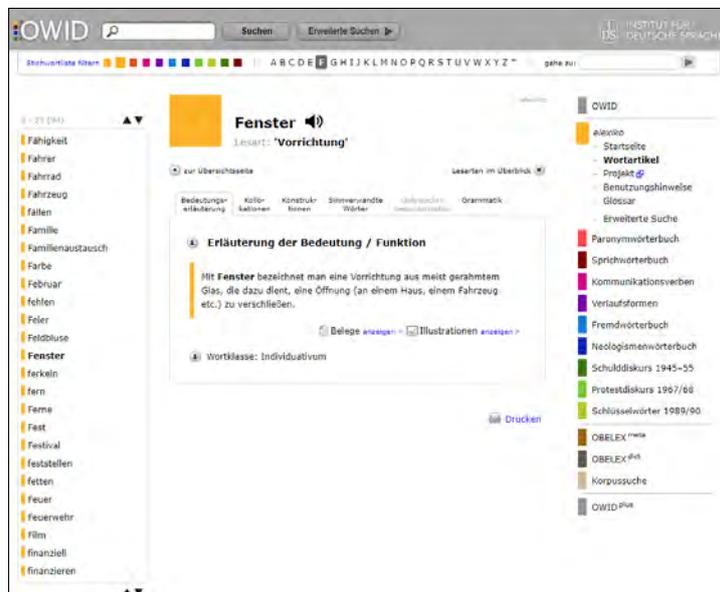
Tabel 1: Verpligte mikrostruktuur

Inskrywing	Motivering
Lemma	Vorm deel van die makrostruktuur en is 'n belangrike element van die soekproses. Meerwoordige terme, volvorme en akronieme sal ook geleemmatiseer word.
Veld van toepassing	Bied kommunikatiewe hulp t.o.v. teksresepsie
Betekenisomskrywing	Vir teksresepsie, sodat studente kan verstaan wat van hulle verwag word.
Afrikaanse vertaalekwivalent	Bied kommunikatiewe hulp t.o.v. teksproduksie en vertaling
Voorbeeldsinne	Bied kommunikatiewe hulp t.o.v. teksproduksie in Afrikaans en Engels.
Woordsoortklassifikasie (Werkwoord of Naamwoord)	Bied kognitiewe hulp, sodat die studente kan verstaan wat van hulle verwag word. Bv. " <i>clip</i> " is 'n werkwoord, want dit is 'n <i>aksie</i> wat uitgevoer moet word.
Verwysing na modulewerk	Kruisverwysings na studiemateriaal kan help sodat die module meer geïntegreerd met die woordeboek sal funksioneer. Sal ondersteuning kan bied tydens toetsvoorbereiding. (Sal ook aanbevelings vir verdere leeswerk gee, indien nodig.)

Tabel 2: Uitgebreide verpligte mikrostruktuur

Inskrywing	Motivering
Akroniem	Vorm deel van die lemma. Dit word in hakies na die volvorm gegee.
Illustrasie	Vir teksresepsie
Kruisverwysings	Bied kognitiewe hulp.
Praktiese toepassings	Bied operasionele hulp, sodat die studente weet hoe om 'n opdrag uit te voer.
Sinonieme	Vir teksresepsie en -produksie

Die woordeboekartikels moet gebruik maak van strategieë om die data-aanbieding meer dinamies te maak. Dinamiese woordeboekartikels kan data-oorlading voorkom deur die skermkootstruktuur effektief te gebruik. Die voorstel vir die GIS-vakwoordeboek is om dieselfde data-aanbiedingstrategie te gebruik as die *lexiko* aanlyn woordeboek (sien afbeelding 4). Dit behels 'n vierkantige teksraam met 'n oortjiekieslys. Die lemma kan direk onder die permanente kenmerke gegee word. Onder die lemma verskyn dan die teksraam met die oortjiekieslys. Soos die gebruiker op 'n oortjie klik, verander die geraamde teks. Hierdie blok-ontwerp help om die data binne die woordeboekartikel georganiseerd te versprei en aan te bied. Dit sal verseker dat daar duidelike, toeganklike soeksones is.



Afbeelding 4: Die data-aanbiedingstrategie van *lexiko*

Daar kan vier soeksones wees op grond van die kommunikatiewe en operasionele funksie. Buiten die lemma, kan al die ander data-aanduiders binne hierdie vier soeksones verdeel word. Drie van die soeksones hou met die kommunikatiewe funksie verband. Die eerste twee soeksones help die gebruiker met teksresepse en die derde soeksone bied hulp met teksproduksie. Die vierde soeksone hou met die operasionele funksie verband omdat dit stap-vir-stap-aanwysings gee om 'n sekere taak met die GIS-programmatuur te verrig. Hierdie vier soeksones kan elk gesien word as 'n beperkte woordeboekartikel en saam vorm hulle 'n omvattende woordeboekartikel (Gouws 2018a: 52). Daar is dus vier beperkte woordeboekartikels hier betrokke en hulle word geïllustreer in figure 5, 6, 7 en 8.

Die data-aanduiders word tussen die vier beperkte woordeboekartikels verdeel, nie net omdat dit die data-tipes funksioneel organiseer nie, maar ook omdat dit die data "filtreer" sodat daar 'n beperkte hoeveelheid data op die skerm vertoon word (Fuertes-Olivera 2016: 237). Dit is nodig sodat die skerm-skoot effektief gebruik word en net die toepaslike en regte hoeveelheid data vir daardie spesifieke gebruikssituasie gegee word (Fuertes-Olivera 2016: 237). In die inligtingsera van vandag is dit dikwels die geval dat gebruikers nie antwoorde op hul vrae vind nie; nie omdat daar 'n gebrek aan inligting is nie, maar juis omdat daar 'n oorfloed daarvan is. Leksikograwe moet vir die gebruiker die naald uit die hooimied haal.

Omdat aanlyn vakwoordeboeke nie van teksverdigting gebruik hoef te maak nie, kan elk van die struktuurmerkers duidelik benoem word en daar kan selfs spasies ooggelaat word tussen die data-aanduiders (sien afbeelding 5). Die eksplisiete struktuurmerkers veroorsaak dat die gebruikershandleiding van effens minder belang is. Die voorstel is dat die struktuurmerkers in 'n ander kleur gegee word en dat hulle hiperskakels is wat die gebruiker na die gebruikershandleiding kan neem indien daar enige iets is waaroor die gebruiker onseker is. Byvoorbeeld, die data-inskrywing vir woordsoortklassifikasies kan die gebruikers na die handleiding neem en eerstens net verduidelik hoekom hierdie aanduiding gegee word, en tweedens ook weer vir hulle verduidelik wat 'n naamwoord en 'n werkwoord is. Die gebruikershandleiding word verder onder afdeling 2.5, wat handel oor die buitekenmerke, bespreek.

2.4.1 Die eerste beperkte woordeboekartikel

Die eerste beperkte woordeboekartikel bevat vier tipes data-inskrywings, waarvan die betekenisparafraze die belangrikste is. Die betekenisparafraze-aanduiding vertoon 'n verhouding van sekondêre lemmatiese adressering, aangesien die betekenisomskrywing verder weg van die lemma geplaas word wat primêre lemmatiese adressering moeilik maak. Hier gaan dit ook om verwyderde sekondêre lemmatiese adressering, in teenstelling tot onmiddellike adressering waar die adresserende aanduiding digby die adres optree. Verwyderde adressering word gebruik omdat die adres (lemma) en die kommentaar wat dit adresseer, 'n groot afstand van mekaar af is.

'n Belangrike aspek om ook in ag te neem, is die vlak van die taal waarin die lemma verklaar word. Die taalgebruik moet nie te moeilik wees vir die gebruiker om te verstaan nie, want die betekenisomskrywing moet nie verdere kommunikatiewe probleme veroorsaak nie. Die samestellers moet ook die beeld van die woordeboek in ag neem as dit by taalgebruik kom. Dit kan problematies wees indien die woordeboek baie formeel is, so ook indien dit té informeel is. Soos reeds genoem is daar drie tipes gebruikers van vakwoordeboeke, te wete die leek, die semideskundige en die deskundige (Du Plessis 2016: 38). Hierdie vakwoordeboek is op semideskundiges gerig en dit moet ook blyk uit die taal waarin die vakwoordeboek geskryf is. Daar is voorts drie taal-

vlakke vir gebruikers, naamlik beginner, intermediêr en gevorderd. Hierdie vakwoordeboek is op gebruikers met 'n intermediêre taalvlak gemik.

Naas die betekenisomskrywing kan (1) die nodige kruisverwysings, (2) die plek(ke) waar die studente die lemma in die modulewerk kan opspoor, (3) die veld waarop die lemma van toepassing is, en (4) enige sinonieme gegee word. Die onderstaande afbeelding 5 bied 'n uitbeelding van die voorbeeldwoordeboekartikel vir die lemma *Digital Elevation Model*. Hierdie voorbeeldwoordeboekartikel bevat al die verpligte en opsionele data-inskrywings behalwe die data-inskrywing vir sinonieme.

Digital Elevation Model (DEM)			
Understanding things	Image	How to do it	Language stuff
<p>Applicable field data models</p> <p>Definition A Digital Elevation Model represents a continuum of elevation values over a topographic surface by a regular array of z-values, referenced to a common vertical datum. DEMs are typically used to represent the bare-earth terrain, void of vegetation and manmade features.</p> <p>Where to find it in the course work lecture 14 tutorial 5</p> <p>Cross-references Digital Terrain Model (DTM)</p>			

Afbeelding 5: Die eerste beperkte woordeboekartikel

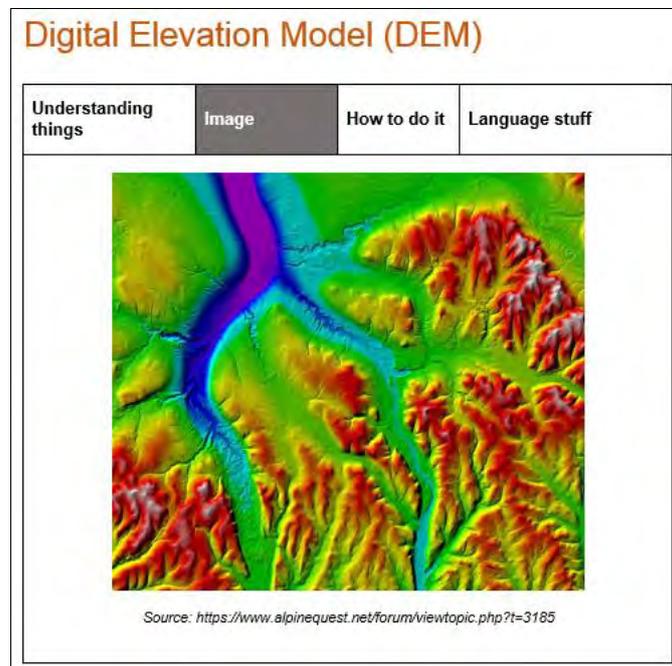
Die vier laasgenoemde data-inskrywings kan ook hiperskakels bevat wat die gebruiker na ander woordeboekartikels of buitekenmeke lei. In die woordeboekartikel vir *Digital Elevation Model* word aangedui dat die lemma deel vorm van die superordinaat *data models*. Indien die gebruiker nie verstaan wat met *data models* bedoel word nie, kan sy dus op die hiperskakel klik om na die woordeboekartikel vir *data models* geneem te word. Die betekenisomskrywing kan woorde bevat wat as lemmas in die woordeboek opgeneem is en hierdie woorde moet as hiperskakels aangebied word. Die voorstel is ook om die woorde wat hiperskakels bevat in ligblou te merk soos wat die tradisie binne die inligtingstechnologie is.

Dit word beskou as goeie praktyk vir 'n woordeboek om voorsiening te maak vir die gebruik van proskriptiewe notas i.v.m. sinonieme en kruisverwysings, waar nodig (Fuertes-Olivera en Tarp 2014: 224). Daar kan gevalle bestaan waar studente twee terme met mekaar verwar of waar hulle geneig is om die

term verkeerd te gebruik. Die voorbeeldwoordeboekartikel se betekenisomskrywing gebruik die definisie wat die Esri-vakwoordeboek gee. In die maak van die GIS-vakwoordeboek sal die vakspesialis en die leksikograaf saam moet werk om oorspronklike data te genereer, maar Esri se voorbeeld kan gevolg word. Die Esri-definisie sê kortliks wat 'n digitale hoogtemodel is, en ook waar dit van toepassing is. Sulke notas kan veral gebruik word indien daar verwarring bestaan tussen die Amerikaanse en Suid-Afrikaanse terme, byvoorbeeld die Amerikaanse term *3D feature* teenoor die Suid-Afrikaanse term *3D object*.

2.4.2 Die tweede beperkte woordeboekartikel

Die tweede beperkte woordeboekartikel bevat net die illustrasieprent en ook die bron waar die illustrasieprent gekry is (sien afbeelding 6). Die illustrasieprent word op sy eie aangebied omdat dit die beste manier is om die skerm-skoot te benut. As dit as deel van die eerste beperkte woordeboekartikel aangebied word, sal daar te veel data in die een skerm-skoot moet pas. Die illustrasieprent is ook 'n opsionele data-inskrywing. Dit beteken dat die hele beperkte woordeboekartikel ook weggelaat kan word, sonder dat dit nodig is om die artikelstruktuur van die eerste beperkte woordeboekartikel aan te pas.



Afbeelding 6: Die tweede beperkte woordeboekartikel

2.4.3 Die derde beperkte woordeboekartikel

Die derde beperkte woordeboekartikel is op die operasionele funksie gerig. Hier kan die data-inskrywings vir die woordsoortklassifikasie en die instruksies vir praktiese toepassing geplaas word. Die beperkte artikel vir die operasionele funksie (*How to do it*) kan net weggelaat word by die woordeboekartikels waar dit nie van toepassing is nie. In die meeste gevalle is dit wel die geval dat selfs naamwoorde 'n beperkte artikel vir die operasionele funksie benodig, aangesien naamwoorde (veral in hierdie veld) immers steeds gepaard gaan met praktiese toepassings. In die geval van die *DEM*-woordeboekartikel (afbeelding 7) kan die aksie "create a DEM" geïdentifiseer word.

Digital Elevation Model (DEM)			
Understanding things	Image	How to do it	Language stuff
<p>Part of speech Noun</p> <p>QGIS instructions In the 363 course, you are not required to create a DEM. However, you are sometimes asked (in tutorial 5 and 6) to create hillshade using a DEM-layer. Instructions to do this are described in the dictionary article for hillshade.</p>			

Afbeelding 7: Die derde beperkte woordeboekartikel.

Die instruksies vir praktiese toepassing in die *DEM*-woordeboekartikel (afbeelding 7), dui aan dat daar word nie van studente in die Geografiese Kommunikasie-module verwag word om 'n digitale hoogtemodel te maak nie, maar dat hulle wel so 'n model moet gebruik om bergskaduwees te teken. Die volgende sin lui: "Instructions to do this are described in the dictionary article for *hillshade*". Die behoud en gebruik van leksikografiese terme (soos "dictionary article") word aangemoedig. Enige leksikografiese terme kan van hiperskakels voorsien word wat na 'n afdeling in die gebruikershandleiding lei waarin die terme verduidelik word. Op hierdie manier kan gebruikers se woordeboekvaardighede verbeter, wat ook 'n positiewe invloed op die woordeboekkultuur van die gebruikers hê.

2.4.4 Die vierde beperkte woordeboekartikel

Die vierde beperkte woordeboekartikel bevat die data-inskrywings wat gebrui-

kers met teksproduksie en vertaling kan help. Die Afrikaanse vertaalekwivalent word gegee, en ook Engelse en Afrikaanse voorbeeldsinne. Sekere besluite oor die hantering van Afrikaanse vertaalekwivalente sal vooraf geneem moet word, aangesien daar terminologie sal wees waaroor daar nog nie Afrikaanse vertaalekwivalente bestaan nie. Daar sal gevalle wees waar die Engelse terme nie noodwendig 'n direkte vertaling in Afrikaans het nie, met ander woorde waar 'n ekwivalentverhouding van zero-ekwivalensie geld. In so 'n geval moet die leksikograaf van surrogaatekwivalente gebruik maak. Surrogaatekwivalensie is wanneer die term in Afrikaans beskryf word en 'n voorstel vir 'n vertaalekwivalent kan gemaak word. Sien 'n volledige bespreking van surrogaatekwivalensie in Gouws en Prinsloo (2010).

Die Engelse voorbeeldsin wat volledig in Afrikaans vertaal word, kan vir die gebruiker illustreer hoe sy die term in take en opdragte kan gebruik. Die voorbeeldsinne moet ook geskryf word om inhoudelik verhelderend ten opsigte van die betekenis van die lemma te wees. Sien die voorbeeldsinne in die DEM-woordeboekartikel in afbeelding 8. Hierdie voorbeeldsinne sou ook deur 'n gekwalifiseerde vertaler vertaal kon word. In hierdie beperkte artikel kan daar ook deurgaans hiperskakels gegee word na die Engelse en Afrikaanse skryfriglyne, wat ook funksioneel tot teksproduksie bydra. Afrikaanse skryfriglyne kan die gebruikers, byvoorbeeld, help met die gebruik van hoof- en kleinletters.

Digital Elevation Model (DEM)			
Understanding things	Image	How to do it	Language stuff
<p>Afrikaanse vertaalekwivalent digitale verhogingsmodel (DVM)</p> <p><i>English example sentence</i> Certain types of <i>raster models</i>, especially <i>Digital Elevation Models (DEMs)</i>, allow you to gain more insight into the <i>terrain</i> that they represent.</p> <p>Afrikaanse voorbeeldsin Sekere rastermodelle, veral digitale verhogingsmodelle (DVM), gee vir jou meer insig in die terrein wat hulle verteenwoordig.</p>			
<p>English Writing Guidelines</p> <p>Afrikaanse skryfriglyne</p>			

Afbeelding 8: Die vierde beperkte woordeboekartikel

Die bogenoemde vier beperkte woordeboekartikels kan nie die volle bewerking van die lemma op een skermkoot gee nie, maar saam vorm hulle 'n omvattende woordeboekartikel wat die lemma volledig bewerk (Gouws 2018a: 53). Tog kan elk van die skermkote beperkte woordeboekartikels genoem word, want hulle voldoen aan die vereistes van 'n artikel. Dit bevat, met ander woorde, 'n lemma en 'n geordende stel aanduiders wat aan daardie lemma geadresseer is.

2.5 Buitekenmerke

Buitekenmerke vorm 'n belangrike deel van die toepassing van die dataver-spreidingstruktuur. In die buitekenmerke word data gegroepeer wat die logiese, eenvormige uitleg van 'n woordeboekartikel sou ontwig, maar wat steeds belangrik is om bepaalde gebruikersbehoefte te bevredig. Tabel 3. is 'n opsomming van die agt buitekenmerke wat na afloop van die eerste twee take van die studie geïdentifiseer is. Al die buitekenmerke moet aan die werklike doel en funksie van die vakwoordeboek voldoen.

Tabel 3: Nuttige buitekenmerke van 'n GIS-vakwoordeboek

Buitekenmerke	Motivering
Gebruikershandleiding	'n Kort video om gebruikers se woordeboekvaardighede te bevorder, en 'n geskrewe handleiding vir naslaanwerk.
Sistematiese inleiding tot GIS	Bied kognitiewe hulp.
Eksamenvoorbereiding	Bied kommunikatiewe hulp en ook ondersteuning met selfstudie. So 'n buitetekst kan dalk uit meer as een deel bestaan.
Navorsingsvoorstelle	Bied kommunikatiewe en kognitiewe hulp.
Engelse skryfriglyne	Bied kommunikatiewe hulp t.o.v. teksproduksie.
Afrikaanse skryfriglyne	Bied kommunikatiewe hulp t.o.v. teksproduksie.
Skakel na aanlyn taalwoordeboeke op die Universiteit se biblioteekwebtuiste.	Bied kommunikatiewe hulp t.o.v. teksresepsie en teksproduksie.
Bronnelys	Verwysings na die korpus.

'n Gebruikershandleiding is nie meer so belangrik vir 'n aanlyn woordeboek soos wat dit vir die gedrukte woordeboek is nie (Klosa en Gouws 2015: 171).

Dit word nietemin ingesluit omdat die gebruikershandleiding die leksikografiese én die tegniese aspekte van die woordeboek kan dek. Volgens een van die dosente in die vakgebied met wie daar 'n onderhoud gevoer is oor die teiken-gebruikers se behoeftes en vaardighede (Theron 2019: 66), is die daar word vermoed dat hulle woordeboekvaardighede ook nie goed is nie. Die inleidingsvideo moet vir die gebruiker demonstreer hoe die vakwoordeboek werk, maar dit moet ook baie duidelik binne die konteks van die gebruiksituasies geplaas word. Op daardie manier kan dit gebruikers aanmoedig om wel die woordeboek te gebruik.

'n Sistematiese inleiding tot GIS bied kognitiewe en kommunikatiewe hulp, want dit kan gebruikers help om op 'n sistematiese manier die vakgebied wat die onderwerp van die woordeboek is, te leer ken en te verstaan (Fuertes-Olivera 2009: 162). Hierdie tipe buitekenmerk kan GIS-studente help om die vakgebied te verstaan as hulle, byvoorbeeld, moet begin voorberei vir die eksamen en ook as hulle opdragte en werkstukke moet voltooi (Fuertes-Olivera 2009: 163). Die dosent kan ook die sistematiese inleiding gebruik as voorgeskrewe leeswerk vir 'n lesing aan die begin van die module. Die woordeboek word dan 'n geïntegreerde leerinstrument. Die inleiding moet spesifiek geskryf word vir die studente as teikengebruikers. Dit sal beteken dat die samestellers die taalgebruik en die omvang van die inhoud in die lug van die behoeftes van die gebruikers moet oorweeg. Die jong gebruikers sal, byvoorbeeld, dalk meer aanklank vind by 'n inleidingsvideo as by 'n geskrewe handleiding.

Daar word voorgestel dat die sistematiese inleiding op twee maniere aangebied word, eerstens as 'n kort oorsig en tweedens as 'n meer omvattende weergawe. Hierdie inhoud moet ook op sy eie webblad aangebied word en van 'n inhoudsopgawe voorsien word. Fuertes-Olivera (2009: 166) dui aan ook dat 'n sistematiese inleiding van genommerde paragrawe of opskrifte gebruik moet maak, dat die betrokke terme in vetdruk verskyn, en dat die hele teks geïntegreerd met die lemmalys en woordeboekartikels moet optree. Daar moet dus kruisverwysings vanuit die woordeboekartikels na die buitekenmerk wees en ook andersom. Die internet en inligtingstechnologie maak dit moontlik om inligting op interessante, dinamiese en aantreklike maniere aan te bied. Die Esri-vakwoordeboek het 'n afdeling "What is GIS" wat gesien kan word as 'n sistematiese inleiding (<https://www.esri.com/en-us/what-is-gis/overview#image1>). Dit stel 'n goeie voorbeeld wat toekomstige vakwoordeboeke gerus kan volg. Dit is wel jammer dat hierdie sistematiese inleiding nie geïntegreerd optree met die Esri-vakwoordeboek nie.

Die eksamenvoorbereiding- en navorsingsvoorstelle-buitekenmerke is geïdentifiseer na die afloop van 'n ondersoek na die gebruikers se behoeftes (Theron 2019: 66). Die ondersoek dui daarop dat is die studente se taal- en skryfvaardighede in eksamensituasies baie swak is. Die eksamenvoorbereiding kan gevolglik 'n teks wees waarin studente riglyne ontvang oor hoe om vrae te beantwoord, en daar kan ook 'n reeks voorbeeldvrae en -antwoorde wees. Studente kan moontlik beter punte behaal en ook die werk beter verstaan as hulle gewys word hoe 'n goeie antwoord op 'n moontlike vraag lyk. Dieselfde geld

vir die navorsingsvoorstelle (Theron 2019: 67). Die betrokke studente het nog nie voorheen 'n navorsingsvoorstel gedoen nie, dus sou riglyne en selfs voorbeelde van afgehandelde projekte die studente baie help om te begryp presies wat van hulle verwag word. Die eksamenvoorbereiding en navorsingsvoorstelle bied dus ook die nodige akademiese ondersteuning.

Die Afrikaanse en Engelse skryfriglyne kan studente help met navorsingsvoorstelle en ook met die verslag wat as deel van 'n groot projek inlewer moet word. Die Afrikaanse skryfriglyne kan ook van hulp wees vir studente wat kleiner opdragte in Afrikaans wil skryf. Dit sal hulle help om die Afrikaanse vertaalekwivalente effektief te gebruik. Daarby word ook voorgestel dat 'n skakel voorsien word om gebruikers na al die taalwoordeboeke op die Universiteit se biblioteekwebtuiste te neem. Dit kan 'n nuttige hulpbron wees tydens opdragte, maar ook indien daar probleme ondervind word met (algemene) terme wat nie in die vakwoordeboek opgeneem is nie. Byvoorbeeld, 'n Gebruiker met Engels as tweede taal mag dalk sukkel om die woord "continuum" in die betekenisomskrywing van die *DEM*-woordeboekartikel (afbeelding 5) te verstaan, en sou die term moontlik in 'n algemene taalwoordeboek sou naslaan.

Die laaste tipe buitekenmerk is die bronnelys. Hierdie bronnelys sal die betroubaarheid van die woordeboek verhoog. Die bronnelys kan dan 'n skakel wees tussen die gebruiker en die korpus. Hierdie bronnelys kan gebruik word om verdere leeswerk of inligting te soek. Daar sal nog verdere navorsing gedoen moet word om die beste manier te bepaal waarop bonverwysing deel van die woordeboekartikel emaa kan word, al dan nie. Dit moet ook hier genoem word dat, alhoewel hulle nie hier bespreek word nie, die skakels na SUNLearn en na die Universiteit se webtuiste ook as buitekenmerke kan geld.

Die uitleg van die bladsy *Other Resources* kan op so 'n manier aangebied word dat die naam van elke buitekenmerk in alfabetiese volgorde gelys word. Hierdie name dien dan as hiperskakels na 'n ander webblad wat die betrokke inligting bevat. Die voorstel is om die inligting op die webblad aan te bied (html-formaat), maar om die gebruiker die opsie te gee om die hele buitekenmerk af te laai as 'n dokument (pdf-formaat) wat uitgedruk kan word. Hierdie opsie is, byvoorbeeld, net relevant ten opsigte van die eksamenvoorbereiding-buitekenmerk. Ander buitekenmerke kan weer goed werk met 'n soekenjin wat gerig is op daardie spesifieke buitekenmerk, byvoorbeeld die gebruikershandleiding en die bronnelys.

2.6 Enkele opmerkings oor die korpus

Die korpus is seker een van dié belangrikste aspekte van enige woordeboekprojek. Die model is slegs die uiteindelijke koppelvlak waarmee die gebruiker in aanraking kom. Hierdie koppelvlak is die skakel wat daar tussen die gebruiker en die korpus is. Dit werk deur middel van 'n databasis wat data uit die korpus neem en dit vir die gebruiker op 'n manier aanbied wat georganiseerd en maklik verteerbaar is. Die koppelvlak is die aspek waaraan daar in hierdie

artikel die meeste aandag geskenk is. Dit is egter nie die enigste komponent wat verseker dat 'n gehaltewoordeboek daargestel word nie. As die databasis en korpus swak is, affekteer dit die gehalte van die woordeboek baie negatief.

In dié studie is daar twee webtuistes geïdentifiseer, GISGeography en GIS Lounge (Theron 2019: 93 en 98), wat saam met die studiemateriaal van die Geografiese Kommunikasie-module kan dien as die vertrekpunt van 'n korpus. Fuertes-Olivera en Tarp (2014: 201) beskou die internet as 'n goeie bron, maar met die voorwaarde dat daar "veiligheidsmaatreëls" gestel moet word. Met ander woorde, alle bronne moet dus eers individueel deur die vakspesialis geanaliseer word om te verseker dat die inhoud aan die vereisde gehalte voldoen. Daarna stel Fuertes-Olivera en Tarp (2014: 202) ook voor dat lemmas vir vakwoordeboeke nie volgens die gebruikelike metode van frekwensie geselekteer word nie, maar op grond van relevansie. Daar sal dus op die kennis en ervaring van die vakspesialis staatgemaak word. Gevolglik word dit weer beklemtoon dat die vakwoordeboek van meet af beplan moet word aan die hand van die leksikografieteorie deur 'n span wat uit minstens 'n leksikograaf, 'n IT-tegnikus en 'n vakspesialis bestaan.

3. Gevolgtrekking

Hierdie artikel het die belangrikste aspekte van die voorgestelde woordeboekmodel vir 'n GIS-vakwoordeboek bespreek (Theron 2019). Enkele opmerkings oor die korpus en databasis is gemaak, maar daar is veral gekyk na die dataaanbieding en die dataverspreidingstruktuur, asook die toegang- en artikelstruktuur van die vakwoordeboek. Die woordeboekmodel is gebaseer op die gebruikersbehoefte en funksies wat in die voorafgaande take van die studie geïdentifiseer is (Theron 2019). Na aanleiding van hierdie bevindings, is daar gefokus op 'n kommunikatiewe funksie wat teksresepsie en -produksie, en vertaling insluit. Daar is ook op 'n operasionele en kognitiewe funksie gefokus. Die woordeboekmodel kan as 'n hibriediese model getipeer word omdat dit op 'n kombinasie van 'n verklarende en vertalende woordeboek dui (Gouws en Prinsloo 2005: 55). Daar is bevind dat 'n GIS-vakwoordeboek by 'n sterk soekenjin en dinamiese woordeboekartikels baat. 'n Voorstel vir verdere navorsing behels ondersoek na die manier waarop toekomstige woordeboeke meer van woordeboek-individualisering gebruik kan maak.

Bronnelys

Woordeboeke

ellexiko. 2018. [Intyds]. Besikbaar: <https://www.owid.de/artikel/155448/Vorrichtung> [11 September 2018].

Esri Support GIS Dictionary. 2018. [Intyds]. Besikbaar: <http://support.esri.com/other-resources/gis-dictionary> [11 September 2018].

PharosAanlyn. 2018. [Intyds]. Beskikbaar: <https://www-pharosaanlyn-co-za.ez.sun.ac.za/tuis> [11 September 2018].

Ander bronne

- Du Plessis, A.H.** 2016. *'n Analise van die selfoon-WAT: 'n Grondslag vir die verbetering van selfoonwoordeboeke*. Ongepubliseerde M.A.-tesis. Stellenbosch: Universiteit Stellenbosch.
- Fuertes-Olivera, P.A.** 2009. Systematic Introductions in Specialised Dictionaries. Nielsen, S. en S. Tarp (Reds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 161-178. Amsterdam/Philadelphia: John Benjamins.
- Fuertes-Olivera, P.A.** 2016. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. *International Journal of Lexicography* 29(2): 226-247.
- Fuertes-Olivera, P.A. en S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlyn/Boston: De Gruyter.
- Gouws, R.H.** 2013. Establishing and Developing a Dictionary Culture for Specialised Lexicography. Jesenšek, V. (Red.). 2013. *Specialised Lexicography. Print and Digital, Specialised Dictionaries, Databases*: 51-62. Berlyn/Boston: De Gruyter.
- Gouws, R.H.** 2018a. Accessibility, Access Structures and Access Procedures. Jesenšek, V. en M. Enčeva (Reds.). 2018. *Wörterbuchstrukturen zwischen Theorie und Praxis*: 35-56. Berlyn: De Gruyter.
- Gouws, R.H.** 2018b. Expanding the Data Distribution Structure. *Lexicographica* 34(1): 225-238.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. en D.J. Prinsloo.** 2010. Surrogaatekwivalensie in tweetalige woordeboeke met spesifieke verwysing na zero-ekwivalensie in Afrikataalwoordeboeke. *Tydskrif vir Geesteswetenskappe* 50(4): 502-519.
- Klosa, A. en R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31(1): 142-172.
- Taalbeleid van die Universiteit Stellenbosch*. 2016. [Intyds]. Beskikbaar: <https://www.sun.ac.za/english/Documents/Language/Finale%20Junie%20Taalbeleid%20November%202016.pdf> [28 Januarie 2019].
- Tarp, S.** 2012. Specialised Lexicography: 20 Years in Slow Motion. *Ibérica* 24: 117-128.
- Theron, L.** 2019. *'n Model vir 'n aanlyn vakwoordeboek vir Geografiese Inligtingstelsels (GIS)*. Ongepubliseerde M.A.-verhandeling. Stellenbosch: Stellenbosch Universiteit.

On Pronunciation in a Multilingual Dictionary: The Case of Lukumi, Olukumi and Yoruba Dictionary

Joy O. Uguru, *Department of Linguistics, University of Nigeria, Nsukka, Nigeria* (joy.uguru@unn.edu.ng)

and

Chukwuma O. Okeke, *Department of Linguistics, University of Nigeria, Nsukka, Nigeria* (chukwuma.okeke@unn.edu.ng)

Abstract: This study centres on reflecting the pronunciation of lemmas in a proposed multilingual dictionary of Lukumi, Olukumi and Yoruba. It shows how the differences and similarities in their pronunciation can be displayed in the proposed dictionary. Lukumi is spoken in Cuba while Olukumi and Yoruba are spoken in Nigeria. The parent language, Yoruba, was used as a reference point to highlight the etymology of Lukumi and Olukumi as well as to buttress their similarities. Two downloaded Lukumi wordlists making up 134 words were used to elicit information on Olukumi equivalents through oral interview. Twenty-two words are used as sample entries. Following Mashamaite's method of promoting the compilation of bilingual dictionaries between African languages, the study presents Lukumi as the source language while Olukumi and Yoruba are the target languages; English translations of the lemmas are shown. The pronunciation of the lemmas is given alongside their meanings and grammatical categories. No dictionary of any Nigerian language has pronunciation of headwords given; hence this study is a positive innovation; also, the display of pronunciation provides evidence of the similarities shared by the three languages. The transcription of the lemmas serves as a good learning aid for the language learners. The dictionary will go a long way to preserve the endangered Lukumi and Olukumi languages.

Keywords: LEMMA, OLUKUMI, LUKUMI, YORUBA, PRONUNCIATION, COGNACY, LEXICOSTATISTICS, MULTILINGUAL DICTIONARY, NIGERIA

Résumé: De la Prononciation dans un Dictionnaire Multilingue: Le Cas du Dictionnaire Lukumi, Olu kumi et Yoruba. Cette étude met l'accent sur une réflexion à propos de la prononciation des lemmes dans un dictionnaire multilingue planifié incluant le lukumi, l'olukumi et le yoruba. Il montre comment les différences et similitudes dans leur prononciation peuvent être présentées dans le dictionnaire en proposition. Le lukumi est parlé au Cuba tandis que l'olukumi et le yoruba sont parlés au Nigéria. La langue parente, le yoruba, a été utilisée comme point de référence pour mettre en évidence l'étymologie du lukumi et de l'olukumi ainsi que pour renforcer leurs similitudes. Deux lexiques téléchargés du lukumi, comprenant un total de 134 mots, ont été utilisés pour obtenir des informations sur les équivalents olukumi par le biais

d'un entretien oral. Vingt-deux mots sont utilisés comme exemples d'entrées. Suivant la méthode de Mashamaite pour promouvoir la confection de dictionnaires bilingues entre les langues africaines, l'étude présente le lukumi comme langue source tandis que l'olukumi et le yoruba sont les langues cibles; des traductions en anglais des lemmes sont présentées. Les différentes prononciations des lemmes sont présentées avec leurs significations et de leurs catégories grammaticales. Aucun dictionnaire d'aucune langue nigériane n'a de prononciation pour leurs entrées ou lemmes; cette étude est donc une innovation positive. Aussi, l'affichage de la prononciation apporte la preuve des similitudes partagées par les trois langues. La transcription des lemmes est une bonne aide à l'apprentissage pour les apprenants de langues. Le dictionnaire contribuera grandement à préserver les langues lukumi et olukumi qui sont menacées de disparition.

Mots-clés: LEMME, OLUKUMI, LUKUMI, YORUBA, APPARENTEE, LEXICOSTATISTIQUE, DICTIONNAIRE MULTILINGUE, NIGERIA

Introduction

Olukumi and Lukumi are spoken by Yoruba descendants in Delta state of Nigeria and in Cuba respectively. While Olukumi speakers are Yoruba descendants who migrated to the present day Delta state (Oshimili Local Government Area) of Nigeria, Lukumi speakers are descendants of Yoruba slaves taken to Cuba and Brazil. Olukumi is spoken in secluded communities like Ugbodu, Ukwunzu and others, hence most of their linguistic heritage has been maintained. Similarly, Lukumi has been maintained because it is mainly used for religious purposes. According to Mason (1997) Lukumi is an alternative term to Olukumi and means 'my friend'. They are similar, having originated from Yoruba, a major Nigerian language spoken in the South West zone. Lukumi has the code, ISO 639-3 luq. It is of the Niger Congo family, specifically of the Yoruboid subgroup. Olukumi, though it does not have a code yet, is also of Yoruboid subgroup. Both varieties are highly related to Yoruba as shown by scholars (Arokoyo 2012; Okolo-Obi 2014 and Anabaraonye 2018).

This paper focuses on indicating pronunciation in a multilingual dictionary with Lukumi as the source language and Olukumi and Yoruba as target languages. Also, in order to have wider readership English translations are provided since Lukumi and Olukumi are spoken in different continents, and the former is used internationally for religious and research purposes. The phonetic features of lemmas are shown for easy pronunciation by readers, especially second language speakers and learners.

We have adopted the method of Mashamaite (2001) who proposes presenting African languages as source languages. He laments that in South Africa, bilingual dictionaries on African languages have English or Afrikaans as the source language while the African languages serve as the target languages. He shows that there is no case where an African language is used as a source language except for bi-directional bilingual dictionaries. A similar situa-

tion can be found in Nigeria. Hence this study fills this gap, being a positive move to project an African language as a source language. Also, though there will be English translations in the proposed multilingual dictionary, the focus is on the three African languages, Lukumi, Olukumi and Yoruba.

Olukumi is not documented, written or studied in schools. Similarly, Lukumi is solely used orally for religious purposes. These situations will lead to the extinction of both languages. Hence the present study is a good step towards the preservation and use of Olukumi and Lukumi. Onwueme (2015) expresses fears that Olukumi could go extinct since its speakers bear Igbo names; Igbo is spoken in their neighbouring communities. Most Olukumi people are bilingual, speaking Igbo and Olukumi, hence the strong influence of Igbo. Concordia (2012) shows that Lukumi retains Yoruba features because it is mainly used for religious purposes. According to Schepens et al. (2013) more frequently used words tend to lose more characters than less frequently used words. Thus, the case of Olukumi which is used for daily activities is understood.

Importance of indicating phonetic features and cognates in dictionaries

According to Van Keymeulen (2003) lexicographers should not include more than they can handle in a dictionary to avoid inadequate management of the data. He however points out that the dictionaries he worked on have such microstructure as pronunciation, meaning, collocations and example sentences; they only include 'general vocabulary' and not terminologies. Though he opines that including elements of microstructure in the dictionary depends on the purpose of the dictionary, it is pertinent that information on pronunciation of the headwords be included in every dictionary since the users have the need to pronounce the words, irrespective of the purpose of the dictionary. Hence in this work, the pronunciation of the lemmas, alongside other macrostructural information, is provided.

Most dictionaries on African languages do not have phonetic transcriptions. This is evident in Igbo dictionaries (Williamson 1972; Blench 2013; Mbah et al. 2013). Some efforts made to include pronunciation in Yoruba dictionaries have not been efficiently done. Yai (1996) chooses to indicate pronunciation of Yoruba words using 'English spelling' in brackets. For example, the Yoruba word, *Òṣùmarè* has the pronunciation (*oshoomanray*). Similarly, Michelena and Marrero (2010) use Spanish spelling to indicate the pronunciation of Yoruba words in their own dictionary. This method of indicating pronunciation is not standard (not phonetic) and can be very misleading.

Rosenhouse (2018) shows that some Arabic language dictionaries have phonetic transcriptions. Nevertheless, these phonetic transcriptions do not always help the learner since his mother tongue affects the pronunciation of the second language. However, it is better that the dictionary gives the learner a guide on

what the pronunciation of the lemmas is so as to reduce mispronunciation which is a major factor in misinterpretation of meaning. Vishnevskaya (2013) points out that phonetic transcription is important for lexicographic purposes. According to him, modern dictionaries, particularly those meant for bilinguals, should have phonetic information. Thus, the current study centres on developing a multilingual dictionary with phonetic details of the lemmas in the language varieties under study. The phonetic aspects help to identify lexical items that are related in the language varieties; this is a big aid to both the language speaker and the learner.

The indication of the pronunciation of the lemmas in a dictionary aids the learner to internalize the spellings and pronunciation of lexemes. According to Shoba (2001) the phonetic-phonological information on a lexical unit is the essential component of the dictionary entry because it facilitates the pronunciation of the word. In her evaluation of Nguni dictionaries, she discovers that the treatment of pronunciation information is inadequate and inconsistent. She concludes that though phonetic-phonological information is included in almost all types of dictionaries, its presentation is associated with certain problems. It is therefore important that a lexicographer understands the phonetic features of a language so as to be able to present them well.

Sobkowiak (2003) distinguishes between lexicographic phonetics and phonetic lexicography. He explains that lexicographic phonetics is phonetics as applied to the process of dictionary making. It centres on the presentation of accent, stress, dialectal variations and other phonetic features in the dictionary. On the contrary, phonetic lexicography centres on other phonetic issues, particularly the place and the role of pronunciation in dictionary compilation.

Mafela (2005) also reveals that adding etymology (cognacy and lexicostatistical facts) could be used to solve the problem of meaning discrimination in dictionaries. This is very pertinent in the case of Lukumi and Olukumi which sprang from the Yoruba language. Due to the long distance separating the two varieties, some words have lost their phonetic features; in such cases, only etymology, where reference is made to the parent language Yoruba, can establish any relationship. Hence Yoruba has been included in this study.

The purpose, functions, nature and typology of the planned dictionary

The earliest attempt at compiling Yoruba vocabulary was that of Crowther (1865). The present study is a prototype of what is being proposed since the list of words is limited. The main dictionary project will be more detailed, covering all areas of life communication. The proposed dictionary will be a Lukumi–Olukumi–Yoruba multilingual dictionary with English translations. There are examples of such dictionaries. According to Marelllo and Tomatis (2008) some of the earliest efforts in making trilingual dictionaries include that of Inglott Bey (1899) who compiled the dictionary of English homonyms with transla-

tions in Italian and French. This was done to help foreigners studying the English language.

Multilingual dictionaries enable thematic and alphabetic lookup. Mashamaite (2001) points out that the primary purpose of bilingual dictionaries is to assist speakers of various languages to learn one another's languages hence promoting multilingualism. He shows that bilingual dictionaries help users to perform the following: reading and listening; speaking and writing as well as translating. No doubt, if these roles can be achieved in the cases of Olukumi and Lukumi, then their preservation and spread is guaranteed. Mashamaite laments the lack of bilingual dictionaries between African languages; this is one of the gaps that the present study aims at filling.

Any modern dictionary derives its data from a corpus (Prinsloo and De Schryver 2009; De Schryver 2006); hence the compilers have to build and query an electronic corpus for the specific language(s) first. In the case of this study, both the online and offline data ought to be crosschecked for accuracy since the language varieties, Lukumi and Olukumi, are spoken in diaspora and the likelihood of influence by languages spoken in their environments is high. Hence care has been taken to crosscheck the collected data with Yoruba, their parent language.

Identifying the phonemes of the languages under study as well as those of the major languages that influenced them

Nota Bene: phonemes

Olukumi vowels: /i/, /u/, / /, /ɪ/, /ĩ/, /ʊ/, /e/, /o/, /ɛ/, /ẽ/, /ɔ/, /õ/, /a/, / /.

Olukumi consonants:

Plosive: /b/, /t/, /d/, /k/, /g/, /kp/, /gb/.

Nasal: /m/, /n/, /ɲ/, /ŋ/, /ŋʷ/.

Trill: /r/

Fricatives: /f/, /s/, /z/, /ʃ/, /ç/, /gʷ/, /ɦ/.

Affricate: /dʒ/

Approximant: /j/, /w/.

Lateral: /l/

Igbo vowels: /i/, /ɪ/, /u/, /ʊ/, /o/, /ɔ/, /e/, /ɛ/, /a/.

Igbo consonants:

Plosives: /p/, /b/, /t/, /d/, /k/, /g/, /kp/, /gb/, /kʷ/, /gʷ/.

Nasals: /m/, /n/, /ɲ/, /ŋ/, /ŋʷ/.

Affricate: /tʃ/, /dʒ/.

Fricative: /s/ /z/ /f/ /ʃ/ (/ʒ/) /ç/ /ɦ/.

Approximant: /ɹ/, /j/, /w/.

Lateral: /l/

Esan (adapted from Ikoyo-Eweto 2017)

Vowels: /a/, /a/ /e/, /e/, /ɛ/, /ĩ/, /i/, /ĩ/, /o/, /ɔ/, /ĩ/, /u/, /ũ/.

Consonants: /p/, /b/, /t/, /d/, /k/, /g/, /kp/, /gb/, /ɕ/, /m/, /n/, /B/, /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /x/, /ɣ/, /tʃ/, /ɕ/, /m/, /ŋ/, /n/, /ɲ/, /l/, /r/, /j/, /w/.

The phonemic inventory of Lukumi (also known as Anagó) as documented by Olmsted (1953) appears below.

Lukumi/Anagó vowels: /i/, /e/, /ɛ/, /a/, /o/, /u/.

Lukumi/Anagó consonants: /b/, /gb/, /kp/, /d/, /t/, /j/, /tʃ/, /g/, /k/, /f/, /s/, /r/, /l/, /m/, /n/, /ɲ/, /ŋ/.

Spanish phonemes:

Spanish has 24 phonemes, 5 vowels and 19 consonants

Spanish vowels: /i/, /u/, /e/, /o/, /a/.

Consonants:

Plosives: /p/, /b/, /t/, /d/, /k/, /g/.

Nasals: /m/, /n/, /ɲ/.

Affricate: /tʃ/

Fricatives: /f/, /θ/, /s/, /ʃ/, /j/, /x/.

Lateral: /l/, /ɬ/.

Flap: /r/

Trill: /r/

Yoruba vowels: /i/, / /, /u/, / /, /e/, /o/, /ɛ/, /ĩ/, /ɔ/, /ĩ/, /a/.

Yoruba consonants:

Plosive: /b/, /t/, /d/, /ʒ/, /k/, /g/, /kp/, /gb/.

Nasal: /m/, /n/, /ŋ/.

Trill: /r/

Fricative: /f/, /s/, /ʃ/, /h/.

Affricate: /dʒ/

Approximant: /j/, /w/.

Lateral: /l/

Phonemic and lexical similarities in Lukumi, Olukumi and Yoruba

Lukumi, Olukumi and Yoruba have many linguistic items that are similar. Olukumi, however, has some phonemes that are non-existent in Yoruba; for instance the /z/ phoneme was borrowed from Igbo (Arokoyo 2012). However, its affinity with Yoruba is evident from the manifestation of high nasality Okolo-Obi (2014). Similarly, Lukumi, which has been very much influenced by Spanish, lost a lot of Yoruba features, particularly grammatical features, pro-

sodic features and some phonemes which have been replaced with Spanish ones (Anabaraonye 2018).

Methods

Oral interview was a major method of data collection in this study. Also, from the Internet, some Cuban Lukumi wordlists were obtained (cf. references for websites) since there is no Lukumi speaker in the immediate environment of the study. Some of the words (kinship terms, numbers, body parts, pronouns, and other basic terms) align with those in the list of Swadesh (1952) while others (common in Yoruba culture) are concepts that feature in Yoruba traditional religion since Lukumi is mainly used for religious purposes. To get their Olukumi equivalents, the words were used to get information through an oral interview with an adult male Olukumi native speaker (Mr. Ogwu) in Ukwunzu, an Olukumi speaking community. A Yoruba native speaker, (Mr. Komolafe) also gave the Yoruba versions. Subsequently, the phonetic features of the words were determined with a view to reflecting them in the dictionary in a way to show their similarities or dissimilarities.

In compiling the entries of the language, the method of Mashamaite (2001) is adopted with modification. This method, the hub and spoke model, links the lexical items of the spoke (source) languages to a common hub (target language). Our focus is to give dictionary users basic information of the entries, including their phonetic forms, while maximising space.

Presentation and compilation of the lemmas

Generally, dictionary entries are arranged in alphabetical order of headwords, which are usually in bold typeface. Dictionary articles generally present the following information types:

- i. Pronunciation information
- ii. Spelling information
- iii. Part of speech information
- iv. Figures of speech information
- v. Meaning demarcation information
- vi. Cross-reference information and
- vii. Information on construction

In the proposed dictionary, Lukumi, Olukumi and Yoruba lemmas are transcribed phonemically to provide both pronunciation and spelling information to the user. In glossing the Olukumi, Lukumi and Yoruba lexical entries in a multilingual dictionary, some user's guide information are highlighted below. Also, the glossed lexical entries of the three languages are presented in an alphabetical order. Check interline sections.

Furthermore, following the international ISO code format for languages, the ISO code OLU, LUK and YOR are used to specify the three languages. The grammatical category is written in abbreviation to the first letter of the word class. Example: noun = n. For the brief illustration of the proposed multilingual dictionary, only the main grammatical category is used. n.: Check the interline sections.

In our demonstration, because the Nigerian languages' equivalents are complete and near cognates, repetition of words is avoided unless where there is difference in tone. The pronunciation information provided also helps in meaning demarcation. If a headword used in a definition is polysemous, the exact meaning of what is intended is clearly stated in the English description. There are also word and space economy. Check the interline sections.

We downloaded Lukumi words from the internet (website addresses can be seen in the references). The words consist of religious and basic words.

Some of the words and their equivalents in Olukumi and Yoruba are given in Table 1 below. However, twenty-two (22) words are used to exemplify (typify) what the proposed dictionary should look like.

Table 1: Some Lukumi/Olukumi/Yoruba words with English Translations

Lukumi	Olukumi	Yoruba	English meaning
1. Agogó /agogo/	g g /agogo/	g g /agogo/	Bell, Hour
2. Asho /aʃo/	şó /aʃɔ/	şó /aʃɔ/	Cloth
3. Ekpó /ekpo/	kpó /ekpo/	epo /ekpo/	Palm oil
4. Ení /eni/	ní /eni/	Ẹn /eni/	Mat
5. Funfún /fũfũ/	F nfún /fũfũ/	Funfun /fũfũ/	White
6. Iré /ire/	ré /ire/	Iré /ire/	Blessings/gift
7. Obí /obi/	b /obi/	Obí /obi/	Kola Nut
8. Oko /oko/	Ókó /ɔkɔ/	Ọkọ /ɔkɔ/	Man, Husband
9. Omí /omi/	m /omi/	Omi /omi/	Water

10. Ona /ona/	òn /ɔna/	òn /ɔna/	Road
11. Orún /orũ/	Orún r n /orũrũ/	O r n /oorũ/	Sun
12. Orí /ori/	r /ori/	Or /ori/	Head
13. Temí /temi/	T m /temi/	T mi /temi/	My, Mine
14. Wa /wa/	W /wa/	W /wa/	Come
15. ar /ara/	r /ara/	Ara /ara/	Body
16. okuta /okuta/	okuta /okuta/	Okuta /okuta/	Stone
17. oguede /oguede/	ògèdè /ɔgèdè/	gèdè gbagb /ɔgèdèagbagba/	Plantain
18. Abó /abo/	gb /agbo/	gb /agbo/	Ram
19. Agoya /agoja/	Y /ja/	Wol /wole/	Enter
20. Akukó /akuko/	kíkò /akíkò/	k kò /akukò/	Rooster
21. Ala /ala/	èl /èla/	l /ala/	Dream
22. Aná /ana/	òn /ɔna/	òn /ɔna/	Road
23. Ara /ara/	kp r /akpara/	r /ara/	Thunder
24. Babá /baba/	b /ba/	B bá /baba/	Father
25. Babalawo /babalawo/	w /awo/	Babalawo /babalawo/	Father of the Secrets/herbalist
26. Ejá /eǰa/	z /eza/	Ēja /eǰa/	Fish
27. Eje /eǰe/	èzè /èzè/	Èjè /eǰe/	Blood

28. Ejo /eɖʒo/	é'zò /ɛzɔ/	Ejó /eɖʒɔ/	Court Case
29. Ińá /iɲa/	Ún /ɔna/	In /ina/	Fire
30. Ikú /iku/	k /uku/	Ikú /iku/	Death
31. Ilé /ile/	l /ule/	Ilé /ile/	House
32. Ilekún /ilekun/	èkùn /ekun/	Il k n /ilekun/	Door
33. Mi /mi/	T m /temi/	Tèm /tɛmi/	My
34. Oke /oke/	d k /edoke/	k /oke/	Mountain
35. Oko /oko/	Ókò /ɔkɔ/	Okò /ɔkɔ/	Man, Husband
36. Omí Dudu /omi dudu/	m ún /omi ɔna/	mí D d /omidudu/	Coffee
37. Omo /omo/	Óm /ɔma/	Omọ /ɔmɔ/	Child
38. Oyín /ojin/	m ny /omi oɲi/	Oyin /ojin/	Honey
39. Tobí /tobi/	b /obi/	Tal bimo /talo bimo/	Who Gave Birth
40. Yeye, Iyá /jeje/ /ija/	Y /je/	Iy /ija/	Mama/mother
41. ay /aja/	z /aza/	Aj /aɖʒa/	Dog
42. meyi /meji/	mêz n /mezɪn/	m j /medʒi/	Two
43. eńe /ɛɲe/	ny /ɛɲi/	ey n /ɛjin/	Tooth
44. leti /leti/	t /eti/	et /eti/	Ear
45. oguede /oguede/	ògèdè /ɔgɛdɛ/	gèdè gbagb /ogɛdɛagbagba/	Plantain

46. Afefé /afefe/	f r /efuru/	Afefé /afefe/	Wind
47. Alejo /aleɕo/	n d /enode/	n d /enode/	Outsider/stranger
48. Arun /arun/	ròn /ur̃/	Arun /arũ/	Sickness/disease
49. Ashé /aʃe/	s /isee/	Ashé /aʃe/	So Be It
50. Dudu /dudu/	kw kw /ok ^w uk ^w u/	D d /dudu/	Dark

It can be seen that from 1 to 45, the lexical items are highly similar in the three languages. Yoruba and Lukumi maintained similarity for the lexical items from 46 to 50, but Olukumi displayed some dissimilarities in the phonemes of the lexical items involved.

Sample compilation of the lemmas

Although Nesi (2009) points out that the electronic dictionary is better than the paper one, it is pertinent to point out that concerning the languages and the circumstance under study, paper dictionary is more accessible and better to comprehend than the electronic one. Hence compilation is discussed here with the paper dictionary in mind.

Nesi (1999: 56) emphasizes the fact that many useful features such as indexes and cross-reference symbols have been added in the paper-dictionaries to assist the user in multiword searches. However, Nesi (1999: 55-56) concludes that the more information the paper-based dictionary contains, the harder (and more time-consuming) it will become for learner users to find exactly what they need to know. This is due to the presence of so much unwanted information. Meijs (1992: 152) as reported by Nesi (1999: 65) predicts "the imminent demise of the dictionary as a book".

As already pointed out in the previous sections, most dictionaries on indigenous Nigerian languages lack some basic features which can enhance their usage and facilitate learning by second language learners. Igbo dictionary by Williamson (1972) which has English as the target language and Igbo as the source language is a mono-directional dictionary serving the need of English speakers. A more recent work *Igbo Adi*, (Mbah et al. 2013) is bidirectional; part A takes care of Igbo speakers while Part B benefits English speakers. Both works, including others not mentioned, do not feature pronunciation as proposed in the present study.

This study proposes a multilingual dictionary for the three languages under study. Most multilingual dictionaries, particularly online ones, are com-

piled in a tabular format. This study proposes a completely different approach from these. Here, we propose a method where there is only one part and each of the three languages have equal representation of the information provided in that part. In other words, speakers and learners of all the languages in the dictionary benefit from that one holistic part.

In typifying the compilation of the proposed dictionary, we partly adopt the model of Mashamaite (2001); that is, the hub and spoke model, a model he proposes for compiling bilingual dictionaries between African languages. He shows that it has the advantage of being economical to use. However, the adoption of the method is with modification since the phonetic features, which he did not consider, are included here. Hence the major parameters of the model will be retained while we incorporate other features that emphasize both meaning and pronunciation. Mashamaite's model seems to emphasize meaning and grammar to the detriment of pronunciation (phonetic form). This is a gap we hope will be filled in the proposed dictionary. Consider, below, an excerpt from Mashamaite (2001: 118).

Source language: Northern Sotho — thelebišene
Target languages: English — television
Lexical unit: same/different
Form unit: same/different
Phonetic form: same/different
Conceptual equivalence: complete

Since the above format will not be good for our dictionary, the following format will be adopted in the proposed dictionary:

Lemma (pronunciation): (Grammar) source language
Lemma (pronunciation): (Grammar) target language1
Lemma (pronunciation): (Grammar) target language 2

} English translation

As usual, the entries are entered alphabetically but presented diagrammatically for clarity and space economy. The form unit, phonetic form (transcription), the lexical unit and grammatical category of the lemmas are displayed with economy of space. The English translations of the lemmas are placed after the braces, an indication that English language is not a focus in the dictionary. Generally, the reversibility principle is applied even though the dictionary will have only one section. Every information that would have been obtained from a second or third section (since three languages are involved) is evident in the sole section. The reversibility principle, which is important for the full understanding of the lemmas from one language to another, is evidently displayed since the translations of the lemmas in the three languages are included in the sole section. Also, the phonemic inventories of the three languages should be displayed in the preliminary pages. The display of the entries in the proposed dictionary is typified below in Table 2.

Table 2: Illustrative Display of the Entries in the Proposed Dictionary

Word	Pronunciation	Language	English translation
1. abo	/abo/	n. LUK	Ram
gb	/agbo/	n. OLU	
gb	/agbo/	n. YOR	
2. agogó	/agogo/	n. LUK	Bell, hour
agogo	/agogo/	n. OLU	
agogo	/agogo/	n. YOR	
3. akukó	/akuko/	n. LUK	Rooster
kíkò	/akikò/	n. OLU	
k kọ	/akukò/	n. YOR	
4. ar	/ara/	n. LUK	Body
r	/ara/	n. OLU	
ara	/ara/	n. YOR	
5. asho	/aʃo/	n. LUK	Cloth
şó	/aʃó/	n. OLU	
şó	/aʃó/	n. YOR	
6. eja	/eɟa/	n. LUK	Fish
z	/eza/	n. OLU	
eja	/eɟa/	n. YOR	
7. eje	/eɟe/	n. LUK	Blood
èzè	/eze/	n. OLU	
èjè	/eɟe/	n. YOR	
8. ejo	/eɟo/	n. LUK	Court case
é+zó	/ezo/	n. OLU	
ejó	/eɟo/	n. YOR	

9. ekpó	/ekpo/	n. LUK	Oil
kpó	/ekpo/	n. OLU	
epo	/ekpo/	n. YOR	
10. ení	/eni/	n. LUK	Mat
ní	/eni/	n. OLU	
en	/eni/	n. YOR	
11. Funfún	/fũfũ/	adj. LUK	White
Funfún	/fũfũ/	adj. OLU	
Funfún	/fũfũ/	adj. YOR	
12. iré	/ire/	n. LUK	Blessings/gift
ré	/ire/	n. OLU	
iré	/ire/	n. YOR	
13. Obí	/obi/	n. LUK	Kola nut
b	/obi/	n. OLU	
Obí	/obi/	n. YOR	
14. Oko	/oko/	n. LUK	Man, husband
Ókọ	/ɔkɔ/	n. OLU	
Ọkọ	/ɔkɔ/	n. YOR	
15. Omí	/omi/	n. LUK	Water
m	/omi/	n. OLU	
Omi	/omi/	n. YOR	
16. Ona	/ona/	n. LUK	Road
òn	/ɔna/	n. OLU	
òn	/ɔna/	n. YOR	

17. Orún	/or /	n. LUK	Sun
Orúnr n	/or r /	n. OLU	
O r n	/oor /	n. YOR	
18. Orí	/ori/	n. LUK	Head
r	/ori/	n. OLU	
Or	/ori/	n. YOR	
19. Temí	/temi/	adj. LUK	My, mine
T m	/temi/	adj. OLU	
T mi	/temi/	adj. YOR	
20. Wa	/wa/	v. LUK	Come
W	/wa/	v. OLU	
W	/wa/	v. YOR	
21. okuta	/okuta/	n. LUK	Stone
okuta	/okuta/	n. OLU	
okuta	/okuta/	n. YOR	
22. oguede	/oguede/	n. LUK	Plantain
ògèdè	/ɔgedɛ/	n. OLU	
gèdè gbagb	/ogedɛ agbagba/	n. YOR	

The presentation of this dictionary is to be likened to that of a dictionary of synonyms where synonyms of lexemes are supplied without the explanation of what the entry words are. In the main dictionary, more words, than the number shown above, will be documented.

This simple, space compacted easy to read method is hereby proposed for multilingual dictionaries, especially paper multilingual dictionaries. Some authors of African languages opine that tone should not be marked since they make works to be cumbersome. However, the format adopted in indicating the pronunciation of lemmas, alongside other features in the proposed dictionary, has shown that it is not cumbersome; rather, making available this information will facilitate the learning of the languages by second language learners particularly.

Furthermore, the degree of reflection of the reversibility principle in the lemmas shown in this study evidently displays the similarities shared by the three languages in their phonetic, lexical and form units. That is, most of their lexical items (lemmas) are translation equivalents in the three languages. De Schryver (2006) citing a scholar like Gouws (1996) shows that the reversibility principle is the condition where the lemmas or translation equivalents in a bilingual dictionary are translation equivalents in the two sections of the dictionary. In a bilingual dictionary, there are usually two sections, the first section dealing with translations from language A to B while the second section deals with translations from language B to A. The reversibility principle demands that both sections should have translation equivalents. For a multilingual dictionary, there may not be such sections; rather (as in the case of this proposed dictionary) lemmas of the languages for a particular entry are given together. However, the reversibility principle is taken into consideration if the lemmas of the languages concerned in the dictionary have equivalent translations (as is the case in the twenty-two sample entries shown above). De Schryver (2006) points out that if words in one language do not map to words in another language, some complexity and especially ingenuity are applied to present their equivalents so as to maintain the reversibility principle.

According to Svensén (1993) translation equivalence entails expressions in the source language having counterparts which are semantically as near as possible in the target language. The high degree of similarities between the three languages under study is seen in the fact that most of their meaning equivalents (lexical units) coincide with their word equivalents (form units). That is, they are linguistically equivalent; in other words, there is a high degree of homogeneity (word for word translation) between the source language, Lukumi and the target languages, Olukumi and Yoruba. This is easily noticeable in the three headwords (one headword for each language) presented for the lemmas. Any observed differences between them could be as a result of varying degrees of influence from the languages and cultural practices (contexts) of their immediate environments.

Discussion

From the entries in this work, it can be observed that Lukumi has retained more Yoruba phonetic features than Olukumi. The implication here (though some words are mainly based on religion) could be that the two varieties, particularly Olukumi, are diverging from the parent language, Yoruba. The divergence of Olukumi could be as a result of influence from the languages, particularly Igbo, surrounding it. There are fears that it could face extinction because of this influence. There is therefore need to intensify its study and documentation so as to foster its maintenance. Hence the proposal for a multilingual dictionary, (as initiated in this study covering Lukumi, Olukumi and Yoruba, with translations in English) should be adopted.

However, there are similarities in the phonemes of Lukumi and Olukumi. Interestingly, their lexical similarity appears to align with their phonemic similarity. They have mainly phonetic spelling; that is, most of their phonemes bear the same symbols as the letters of their alphabet. This is because according to Coulmas (1989) alphabets for African languages were influenced by the work of phoneticians at the International Institute of African Languages and Cultures in London. This institute established the Practical Orthography of African Languages; which was influenced by the IPA, thus being based on the principle of one letter corresponding to one sound. In spite of their phonetic spelling, transcribing the lemmas to show their pronunciation is important since some of the letters do not have one-to-one relationship with the phonemes they represent.

Implications of including pronunciation in a multilingual dictionary

Showing the pronunciation of entries helps in revealing the phonetic similarities and dissimilarities of lexical items in a multilingual dictionary. This in turn, aids in revealing the relationship between the language varieties. Also, it reveals their etymology. This can be seen in the distribution of some phonemes in the sample entries shown above. Consider the ones shown below in Table 3:

Table 3: Distribution of Some Phonemes in Sample Entries

Word	Pronunciation	Language	English translation
Ejá	/eɕʒa/	n. LUK	Fish
z	/eza/	n. OLU	
Ẹja	/eɕʒa/	n. YOR	
Eje	/eɕʒe/	n. LUK	Blood
èzè	/eze/	n. OLU	
Ẹ̀jè	/eɕʒe/	n. YOR	
Ejo	/eɕʒo/	n. LUK	Court Case
é+zó	/ezo/	n. OLU	
Ejó	/eɕʒo/	n. YOR	

From the lemmas above, it can easily be deduced that the phoneme /z/ is used in place of /ɕʒ/ in Olukumi. Also, that Lukumi and Yoruba use of the same phoneme not only shows the etymology of Lukumi and Olukumi but also portrays the fact that Lukumi has retained more Yoruba features than Olukumi.

Only the transcription of their pronunciation can explicitly show these facts, making it easy for speakers as well as second language learners to use the dictionary.

Brandon (1993) explains that phonetic transcription gives uniformity in pronunciation and spelling. This is because the language users are not in doubt as to how to pronounce the words. According to Carroll (1992) phonological representations serve as bases for cognate pairing, which is a major step in compiling a multilingual dictionary. Thus phonetic transcriptions aid in ascertaining the words that have equivalents in the languages represented in the dictionary.

Since the varieties in the proposed dictionary are mainly spoken by people on different continents, it is necessary to have sufficient information in it so as to avoid confusing users. Hence sufficient information, including pronunciation, should be made available in the dictionary. The dictionary can be posted online even though its design is for paper dictionary. Lew (2011) opines that indicating pronunciation in online dictionaries is necessary.

Mashamaite (2001) opines that bilingual dictionaries may serve different purposes depending more on the communicative needs of the dictionary users than on the amount of information supplied by the compiler.

Bilingual dictionaries also aid in translation; here displaying the lexical features of the languages helps a lot. Thus this work, having shown the phonetic and lexical features of the varieties under study, has provided enough information to make the proposed multilingual dictionary functional in language use and study. This is because it is multi-directional; speakers of these languages can use it for various language purposes. According to Rojas (2012) there is a need to have multilingual dictionaries for minority languages. He emphasizes that it is not only international languages that should have multilingual dictionaries. The present study is a step towards achieving this for Lukumi, Olukumi and Yoruba. Hence the significance of this study in the conservation of Olukumi and Lukumi (which are endangered) cannot be over emphasised.

Conclusion

Since the major target of language study is speech, this paper has taken time to show how pronunciation can be displayed in a multilingual dictionary being proposed for Lukumi, Olukumi and Yoruba. The aim of this paper is to show the extent to which Olukumi and Lukumi are mutually intelligible (judging from the phonetic and semantic similarities of their lexical items). The paper shows how a multilingual dictionary can display these similarities and aid intelligibility and communication among the speakers/learners of the languages. We conclude that this can be done through the indication of the phonetic/phonemic transcription of entries of each language.

We show that through the indication of pronunciation of the entries, the phonemic similarities and dissimilarities between the languages can be dis-

- De Schryver, Gilles-Maurice.** 2006. Compiling Modern Bilingual Dictionaries for Bantu Languages: Case Studies for Northern Sotho and Zulu. Corino, E. et al. 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 / Proceedings XII Euralex International Congress, Torino, Italia, September 6th-9th, 2006*: 515-525. Alessandria: Edizioni dell'Orso.
- Gouws, R.H.** 1996. Idioms and Collocations in Bilingual Dictionaries and Their Afrikaans Translation Equivalents. *Lexicographica* 12: 54-88.
- Ikoyo-Eweto, E.O.** 2017. Phonetic Differences between Esan and Selected Edoid Languages. *Journal of Linguistics, Language and Culture* 4(1): 65-85.
- Lew, R.** 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.) 2011. *e-lexicography: The Internet, Digital Initiatives and Lexicography*: 230-250. London/New York: Continuum.
- Mafela, M.J.** 2005. Meaning Discrimination in Bilingual Venda Dictionaries. *Lexikos* 15: 276-285.
- Marello, C. and M. Tomatis.** 2008. Macro- and Microstructure Experiments in Minor Bilingual Dictionaries of XIX and XX Century. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 1155-1164. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Mashamaite, K.J.** 2001. The Compilation of Bilingual Dictionaries between African Languages in South Africa: The Case of Northern Sotho and Tshivenda. *Lexikos* 11: 112-121.
- Mason, J.** 1997. Ogun: Builder of the Lukumi's House. Barnes, S.T. (Ed.). 1997. *Africa's Ogun*: 353-368. Bloomington/Indianapolis: Indiana University Press.
- Mbah, B.M., E.E. Mbah, E.S. Ikeokwu, C.O. Okeke, I.M. Nweze, C.N. Ugwuona, C.M. Akaeze, J.O. Onu, E.A. Eze, G.A. Prezi and B.C. Odii.** 2013. *Igbo Adi*. Nsukka: University of Nigeria Press.
- Michelena, M. and R. Marrero.** 2010. *Diccionario de Términos Yoruba. Pronunciación, sinonimias y uso práctico del idioma lucumí de la nación Yoruba*. Mexico/Miami/Buenos Aires: Prana.
- Nesi, H.** 1999. The Specification of Dictionary Reference Skills in Higher Education. Hartmann, R.R.K. (Ed.). 1999. *Dictionaries in Language Learning: Recommendations, National Reports and Thematic Reports from the TNP Sub-Project 9: Dictionaries*: 53-67. Berlin: Freie Universität Berlin.
- Nesi, H.** 2009. Dictionaries in Electronic Form. Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography*: 458-478. Oxford: Oxford University Press.
- Okolo-Obi, B.** 2014. *Aspects of Olukumi Phonology*. M.A. Project Report. Nsukka: University of Nigeria.
- Olmsted, D.L.** 1953. Comparative Notes on Yoruba and Lucumí. *Journal of the Linguistic Society of America* 29(2): 157-163.
- Onwueme, I.C.** 2015. *Questions Not Being Asked: Topical Philosophical Critiques in Prose, Proverbs, and Poems*. Bloomington: AuthorHouse.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 2009. Corpus Applications for the African Languages, with Special Reference to Research, Teaching, Learning and Software. *Southern African Linguistics and Applied Language Studies* 19(1-2): 111-131.
- Rojas, D.** 2012. *Multilingual Dictionary for Minority Languages*. <https://launchpad.net/milandic>. Accessed 04/06/2018.
- Rosenhouse, J.** 2018. Modern Arabic Dictionaries: Phonetic Aspects and Implications *Zeitschrift für Arabische Linguistik* 67: 71-89. DOI: 10.13173/zeitarabling.67.0071.
- Schepens, J., T. Dijkstra, F. Grootjen and W.J.B. van Heuven.** 2013. Cross-language Distributions of High Frequency and Phonetically Similar Cognates. *PLoS One* 8(5): e63006.

- Shoba, F.M.** 2001. *The Representation of Phonetic-Phonological Information in Nguni Dictionaries*. M.A. Thesis. Stellenbosch: Stellenbosch University.
- Sobkowiak, W.** 2003. Pronunciation in Macmillan English Dictionary for Advanced Learners on CD-ROM. *International Journal of Lexicography* 16(4): 423-441.
- Svensén, Bo.** 1993. *Practical Lexicography*. Oxford: Oxford University Press.
- Swadesh, M.** 1952. Lexico-statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96(4): 452-463.
- Van Keymeulen, J.** 2003. Compiling a Dictionary of an Unwritten Language: A Non-corpus-based Approach. *Lexikos* 13: 183-205.
- Vishnevskaya, G.M.** 2013. Language Globalization through the Prism of Lexicography. Karpova, O.M. and F.I. Kartashkova (Eds.). 2013. *Multi-Disciplinary Lexicography: Traditions and Challenges of the XXIst century*: 2-10. Newcastle upon Tyne: Cambridge Scholars Publishing.

Création de ressources lexicographiques Français–Slovène d'aide à la traduction spécialisée

Sonia Vaupot, *Département de traduction, Faculté des Lettres,
Université de Ljubljana, Slovénie*
(sonia.vaupot@ff.uni-lj.si)

Résumé: Afin de pallier le manque de ressources lexicographiques pour la paire de langues français–slovène dans le domaine de la traduction spécialisée, un projet bilatéral a été mis en place dans le but de développer un dictionnaire en ligne destiné à être utilisé comme un outil lexicographique et une aide à la rédaction de textes scientifiques. Cet article présente une méthode qui permet aux étudiants de participer activement au processus de création d'une base de données en ligne en vue du stockage de données terminologiques et phraséologiques. Nous présentons tout d'abord la situation lexicographique en Slovénie, notamment pour la paire de langues français–slovène, puis les objectifs du projet ainsi que la méthode d'enseignement visant la création de corpus comparables et de ressources lexicographiques compilées par les étudiants. Enfin, nous terminons par une synthèse des résultats obtenus. Cette méthode, utilisée depuis 2018 dans le cadre d'un cours de Master, donne d'excellents résultats du point de vue pratique et pédagogique.

Mots-clés: DICTIONNAIRE BILINGUE, DICTIONNAIRE MULTILINGUE, FORMATION EN TRADUCTION SPÉCIALISÉE, CORPUS SPÉCIALISÉS, TERMINOLOGIE, PHRASÉOLOGIE, COLLOCATIONS, TRADUCTION, ÉQUIVALENCE, DIDACTIQUE

Abstract: **Creating a Slovene–French LSP Dictionary for translation purposes.** In order to compensate for the lack of lexicographical resources for the French–Slovenian language pair in the field of specialized translation, a bilateral project has been set up with the aim of developing an online dictionary to be used as a lexicographical tool and as an aid for writing scientific texts. This article presents a method that allows students to actively participate in the process of creating an online database for storing information on specialized terminology and phraseology. We first present the lexicographical situation in Slovenia, especially for the French–Slovenian language pair, and then the objectives of the project as well as the teaching method aimed at creating comparable corpora and lexicographical resources compiled by the students. Finally, we conclude with a synthesis of the results obtained. This method, which has been used since 2018 in a Masters course, provides excellent results from a practical and pedagogical point of view.

Keywords: BILINGUAL DICTIONARY, MULTILINGUAL DICTIONARY, SPECIALISED TRANSLATION TEACHING, SPECIALISED CORPORA, TERMINOLOGY, PHRASEOLOGY, COLLOCATIONS, TRANSLATION, EQUIVALENCE, DIDACTICS

1. Introduction

La Slovénie possède une certaine tradition lexicographique qui s'est mise en place au fur et à mesure des contacts avec la langue allemande dans le contexte culturel de l'Europe centrale. Pays de plus de deux millions d'habitants, ce pays a été sous la domination du Saint Empire romain germanique et des Habsbourg d'Autriche, avant de faire partie de la Yougoslavie de 1918 à 1991, date de son indépendance.

Créée en 1938, l'Académie slovène des sciences et des arts (en slovène, Slovenska akademija znanosti in umetnosti ou SAZU) comprend l'Institut de la langue slovène Fran Ramovš, fondé en 1945, dont l'une des missions concerne la création de bases de données lexicographiques. Cet Institut a participé à plusieurs projets nationaux et internationaux, dont l'Atlas linguistique slave général (OLA), l'Atlas Linguarum Europae (ALE), la Fédération internationale des banques de terminologie (IFTB), l'Encyclopédie onomastique slovène (EOS), etc. Depuis la proclamation de l'indépendance en 1991, les Slovènes ont continué à développer un intérêt croissant pour leur langue. En 2004, l'Institut d'études appliquées Trojina a été fondé à Ljubljana dans le but de promouvoir la linguistique contemporaine. Orienté vers la recherche et le développement de la langue slovène, l'Institut a collaboré à plusieurs reprises avec l'Université de Ljubljana sur des projets lexicographiques. De nombreux projets (dictionnaires monolingue, bilingue, phraséologique, etc.) ont ainsi vu le jour. Parallèlement, des travaux de recherche (Vrbinc et al. 2018: 363) ont été menés sur la lexicographie slovène en général (Snoj 2004; Žagar Karer 2011; Ledinek et al. 2015; Srebnik 2015; Gantar 2015; Gorjanc et al. 2017). Ces travaux sont influencés, entre autres, par les théories lexicographiques: britannique (Sue Atkins, Patrick Hanks, RRK Hartmann, Adam Kilgarriff, Michael Rundell et John Sinclair); francophone (Sylviane Granger, Gilles-Maurice de Schryver); tchèque (František Čermák, Ladislav Zgusta); allemande (Herbert Ernst Wiegand, Ute Römer); américaine (Dwight Bolinger, Don McCreary, Erin McKean et Ben Zimmer) et russe (Juri Apresjan). Quant à la terminographie moderne, elle subit l'influence de l'école de terminologie de Vienne (ibid.: 364-365).

Toutefois, depuis l'adhésion de la Slovénie à l'Union européenne, un besoin conséquent en ressources lexicales et terminologiques bilingues et plurilingues apparaît notamment dans le domaine de la traduction. Malgré les nombreuses ressources disponibles, on constate une pénurie de ressources lexicales générales et spécialisées, plus particulièrement pour la paire de langues français–slovène ou inversement. Actuellement, les dictionnaires suivants sont disponibles, en version papier ou électronique: le dictionnaire français–slovène (Grad 2004, la version papier datant de 1971), le dictionnaire slovène–français (Jesenik et Dembskij 2008; la version papier datant de 1990) ainsi qu'une édition de poche (Perko 2006) et le dictionnaire en ligne Pons. Il existe peu de dictionnaires spécialisés, mentionnons le dictionnaire slovène–français européen (Perko 2007) et le dictionnaire terminologique slovène–anglais–français (version pdf, 2007) reprenant la terminologie de la Présidence slovène et accessible sur

le site du Secrétariat général du gouvernement de la République slovène, ainsi que les bases multilingues Glosbe et Linguee. Il est aussi possible de consulter en ligne les bases de données lexicales multilingues Evrokorpus, Evroterm ou IATE qui se réfèrent essentiellement aux corpus communautaires.

Par ailleurs, face à la taille du marché de la traduction en Slovénie, il est nécessaire de préciser que les jeunes traducteurs peuvent difficilement se spécialiser dans un seul domaine. Ils sont en grande majorité polyvalents et amenés à traduire, dans des domaines très variés, aussi bien dans la langue maternelle que dans une langue étrangère. Les différents bureaux et services de traduction privés ou officiels (bureau de traduction du gouvernement, différents ministères, etc.) exigent parfois de savoir traduire dans une voire plusieurs langues étrangères. La majorité des traductions dans la langue étrangère sont ensuite relues par un réviseur professionnel. Il est donc indispensable de former les étudiants, futurs traducteurs, à traduire différents types de textes vers la langue maternelle et vers la langue étrangère. Le Département de traduction de la Faculté des Lettres de l'Université de Ljubljana (Slovénie) accueille ainsi chaque année des étudiants qui ont pour langues obligatoires, le slovène (langue maternelle A) et l'anglais (langue étrangère B). Les étudiants choisissent également une autre langue B: l'allemand, le français ou l'italien. Au cours de leur cursus, ils sont donc amenés à traduire vers leur langue A et leurs deux langues B. Après leurs études de Master, une fois sur le marché du travail, les jeunes traducteurs doivent être capables non seulement de traduire tout type de textes, mais aussi d'assurer la gestion terminologique et de pouvoir utiliser divers outils de traduction (Trados, par exemple).

Cet article expose, d'une part, les résultats de deux enquêtes menées auprès d'étudiants inscrits en Master de traduction, l'objectif étant d'identifier la situation lexicographique actuelle pour la paire de langues français-slovène. D'autre part, le marché de la traduction intégrant celui de la création de ressources linguistiques (monolingues, bilingues ou multilingues), nous présentons le projet bilatéral Proteus mis en place entre l'UFR EILA de l'Université de Paris (anciennement Paris 7-Diderot) et le département de traduction de l'Université de Ljubljana. Ce projet porte sur la conception de ressources lexicographiques français-slovènes dans le dictionnaire en ligne ARTES. Nous présentons une méthode active et créative visant à familiariser l'étudiant avec la terminologie et la phraséologie tout en créant un dictionnaire spécialisé bilingue en ligne. Les ressources dans la base sont en effet compilées par les étudiants. Nous présentons les objectifs du projet ainsi que la méthode d'enseignement visant la création des corpus comparables et des ressources lexicographiques en ligne. Enfin, nous terminons par une synthèse des résultats obtenus.

2. Enquêtes exploratoires

Face au peu de ressources lexicographiques à disposition pour le couple de langue français-slovène et inversement, deux enquêtes ont été menées durant

deux années consécutives (de 2018 à 2020). Nous présentons ici les résultats d'une première enquête ayant pour objectif de vérifier l'état actuel du dictionnaire français–slovène (1971), puis une autre enquête qui visait à identifier la méthode de travail des étudiants pour traduire un texte spécialisé sans dictionnaire terminologique à leur disposition.

2.1 Faiblesses du dictionnaire bilingue

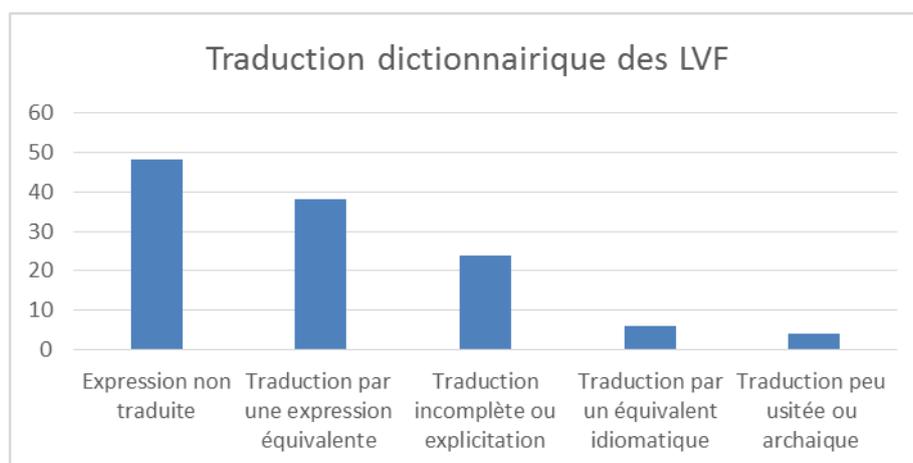
L'importance des locutions dans l'apprentissage d'une langue étrangère n'est plus à prouver (Galisson 1983 et 1991; Tréville 2000; Boers, Demecheleer et Eyckemans 2004; Detry 2008; Pavičić Takač 2008, Cavalla 2009, Szczepaniak et Lew 2011, Simonnet et Tutin 2019, etc.). Nous savons également à quel point la traduction des locutions peut poser des difficultés. Nous avons ainsi choisi de vérifier les traductions slovènes des locutions verbales françaises (LVF), proposées par le dictionnaire français–slovène dont la version papier date de 1971 et la version électronique de 2004.

L'un des premiers réflexes étant de faire une recherche dans le dictionnaire avant de traduire une expression figée inconnue, nous avons proposé un exercice aux étudiants inscrits en Licence 2 et Licence 3 qui consistait à traduire vers leur langue maternelle A (le slovène) 115 locutions verbales françaises avant même de consulter le dictionnaire bilingue. Cette approche visait à vérifier le traitement et la traduction des locutions verbales dans le dictionnaire en ligne slovène–français (Grad 2004). En s'aidant de la définition ou l'explication proposée par le dictionnaire unilingue français de leur choix, les apprenants ont proposé leurs propres traductions slovènes des LVF. Les résultats ont ensuite été recueillis dans un fichier Excel, comportant les LVF, leur définition ou l'explication tirée d'un dictionnaire français, la traduction de l'étudiant et la traduction proposée par le dictionnaire bilingue. En comparant les traductions proposées par le dictionnaire slovène et celles des apprenants, notre objectif visait non seulement à explorer les stratégies de traduction des LVF, mais aussi à vérifier les traductions proposées par le dictionnaire bilingue dont la première version, datant de 1971, a été éditée à 9 reprises. Les résultats sont présentés dans le Tableau 1.

Cette approche nous a permis de constater les faiblesses du dictionnaire bilingue: sur les 115 LVF, 48 LVF sont absentes du dictionnaire (soit un peu moins de la moitié). Parmi les LVF qui apparaissent dans le dictionnaire, on compte 38 LVF traduites en slovène par un simple verbe; 13 LVF sont explicitées; 10 sont traduites par une expression verbale non figée et 6 LVF sont traduites par une expression idiomatique figée. Si nous comparons ces résultats avec les traductions des étudiants, nous remarquons que la majorité des LVF ont été traduites, et non explicitées, par une expression slovène équivalente. Les résultats de cet exercice montrent clairement que la proportion des LVF est non seulement faible dans le dictionnaire bilingue, mais la majorité des expressions sont expliquées au lieu d'être traduites. Nous en concluons que le diction-

naire bilingue s'intéresse aux informations linguistiques et sémantiques, mais ne tient pas suffisamment compte du côté idiomatique de la langue. En ce qui concerne les LVF, la micro-structure est réduite, limitée à une partie explicative et ne répond pas vraiment aux besoins de traduction.

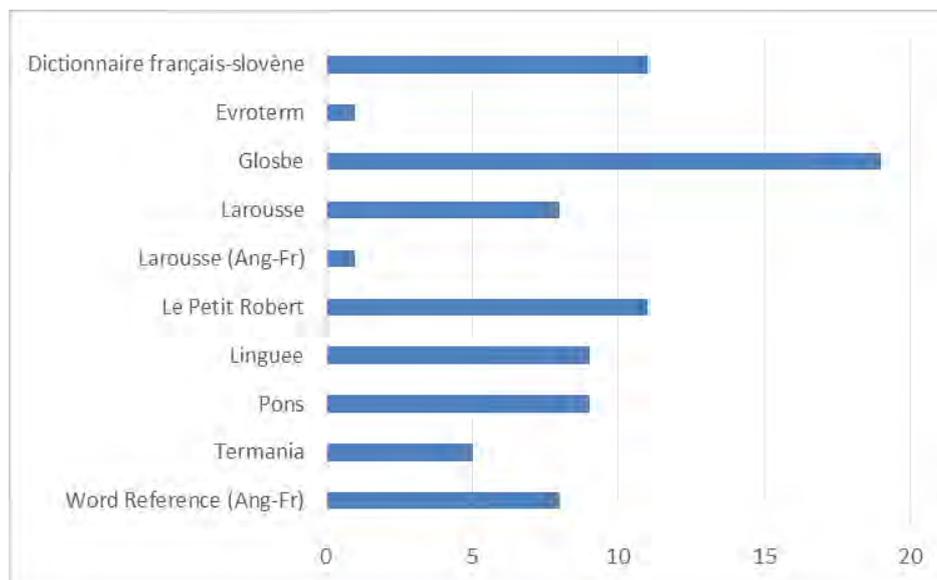
Tableau 1: Les LVF dans le dictionnaire bilingue français–slovène (Grad 2004)



2.2 Manque de ressources terminologiques

Une deuxième approche informatisée a été effectuée à l'aide de la méthode d'enregistrement in vivo Translog afin de collecter les données pendant l'opération traduisante (corrections, révisions, pauses, etc.) et notamment afin de vérifier la fréquence des outils en ligne utilisés par les apprenants (voir aussi Jakobsen 1999, 2000). Les étudiants, inscrits en première et deuxième années de Master, ont été priés de traduire à partir du slovène vers leur seconde langue étrangère B (le français) un court texte comprenant une terminologie spécialisée relative à la région du Karst en Slovénie (Mavrič Bordon 2019). Les données recueillies par le logiciel Translog ont permis de constater que les apprenants ont utilisé, par ordre de préférence: le dictionnaire multilingue Glosbe, le dictionnaire électronique slovène–français (Jesenik et Dembskij 2008), le dictionnaire Pons et la base multilingue Linguee. Certains apprenants sont passés par une langue intermédiaire, présentement l'anglais, et ont utilisé les dictionnaires multilingues Word reference et Linguee ainsi que le portail multilingue Termania. Ils ont également vérifié certains termes français à l'aide du Petit Robert ou du Petit Larousse en ligne. Enfin, le dictionnaire multilingue Evroterm et le dictionnaire anglais–français Larousse n'ont été utilisés qu'à une seule reprise. Le Tableau 2 présente les dictionnaires utilisés par les étudiants selon leur fréquence:

Tableau 2: Outils en ligne selon leur fréquence d'utilisation



Aucun de ces dictionnaires n'étant spécialisé dans la terminologie du Karst, nous pouvons confirmer la tendance des étudiants à utiliser des dictionnaires généraux en ligne et à récolter ensuite les termes spécialisés par l'intermédiaire d'un moteur de recherche (Google, par exemple). En plus des dictionnaires en ligne, les apprenants ont tous consulté Wikipédia, trois apprenants ont parcouru chacun un article scientifique différent et deux apprenants ont consulté la traduction française d'un texte touristique sur le sujet. Ceci leur a permis de trouver une partie de la terminologie requise. Nous remarquons également que les étudiants n'ont pratiquement pas eu recours à un corpus pour traduire leur texte.

Fort de ces constats, la nécessité d'élaborer une base de données lexicographique s'est imposée afin de mettre en place un outil de traduction plus spécialisé pour la combinaison linguistique slovène et française.

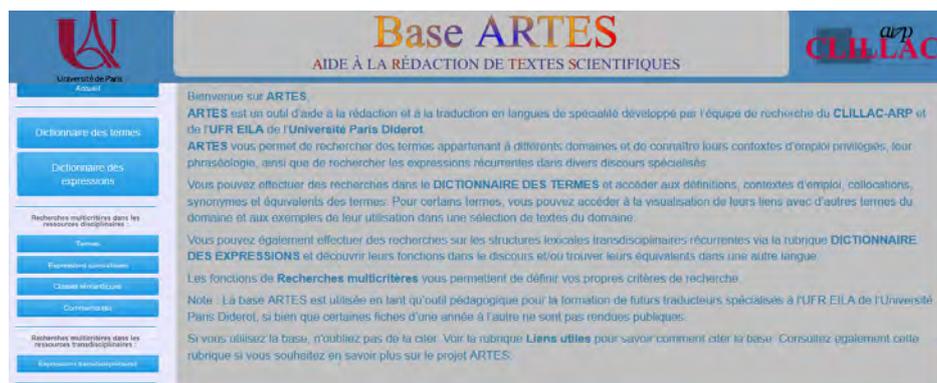
3. Conception du projet lexicographique

Afin de pallier le manque de ressources pour le couple de langue français-slovène, notamment dans le domaine de la traduction spécialisée, un projet bilatéral Proteus a été conduit en 2017 entre le département de traduction de la Faculté des lettres de l'Université de Ljubljana (Slovénie) et l'UFR EILA de l'Université de Paris (autrefois Université de Paris 7-Diderot). Ce projet a pour objectif principal l'élaboration d'un dictionnaire bilingue spécialisé qui serait à

la fois un outil pour les traducteurs et une aide à la rédaction professionnelle (Pecman 2007, 2012; Pecman et Kübler 2011; Pecman et Gledhill 2018; Kübler 2003, 2011; Kübler et Pecman 2012). Dans un premier temps, cela consiste à développer des ressources terminologiques et phraséologiques français–slovènes (et inversement) d'aide à la traduction spécialisée au moyen du dictionnaire en ligne ARTES. L'équipe de l'UFR EILA de l'Université de Paris a aimablement autorisé et permis l'accès à la base ARTES. En plus de l'outil informatique, elle a également apporté les méthodes et son expérience puisqu'elle mène ce projet depuis plusieurs années dans les domaines de la terminologie et phraséologie spécialisées, entre autres, pour la combinaison linguistique français–anglais et inversement. La base en ligne est également conçue pour prévoir des collaborations extérieures multilingues et est adaptée à une cinquantaine de langues. L'accès à la base est assuré par l'équipe de recherche CLILLAC.

Les ressources dans la base ARTES sont en libre accès, sous forme de fiches terminologiques ou d'entrées de dictionnaires, par le biais de son interface de consultation (<https://artes.app.univ-paris-diderot.fr/artes-symfony/web/app.php>). L'interrogation peut porter sur les termes depuis l'onglet «Dictionnaires des termes» ou sur les structures lexico-grammaticales transversales aux domaines, appelées également 'collocations génériques', depuis l'onglet «Dictionnaire des expressions». La fonction «Recherches multicritères», accessible depuis l'interface de consultation (<https://www.eila.univ-paris-diderot.fr/recherche/artes/index>) permet également d'interroger ces ressources au moyen une série de requêtes visant à définir les critères de recherche.

Tableau 3: L'interface de consultation de l'outil ARTES



Dans le cadre du projet bilatéral, la compilation de données dans la base ARTES a été conçue pour répondre aux besoins lexicographiques et pédagogiques spécifiques du département de traduction. Le projet a nécessité la mise en place d'une plateforme pour partager les spécifications et les res-

sources, ainsi que d'un calendrier pour prévoir un échange efficace entre les divers participants. La formation, le soutien à distance et le suivi de la base de données lexicographiques sont assurés par l'équipe française. Des profils d'utilisateurs sont ainsi créés chaque année afin de permettre aux étudiants d'alimenter la base. La base ARTES étant conçue pour permettre aux apprenants de participer au projet, la création des fiches terminologiques et phraséologiques est ainsi confiée, chaque année, aux étudiants inscrits en Master de traduction sous la direction de l'enseignante slovène responsable du projet.

En plus de l'intérêt professionnel, les plus-values didactique et pédagogique du dictionnaire ne sont pas à négliger. Les étudiants sont régulièrement initiés à la recherche d'informations, à la linguistique de corpus et à la création de ressources lexicales. Ils développent des compétences et acquièrent des connaissances souvent indispensables en traduction spécialisée. Le résultat final est présenté sous la forme d'un dossier que les étudiants doivent remettre à l'issue du cours. Les enseignants peuvent ainsi suivre les travaux en cours et évaluer les ressources compilées par les apprenants.

L'objectif final du projet étant d'élaborer une base de données lexicographique, la méthodologie utilisée pour assurer la qualité des données collectées par les apprenants est répartie en trois étapes clés que nous présentons par la suite: (1) la création des corpus comparables, l'analyse et l'extraction des termes ou collocations; (2) la saisie des termes et/ou des collocations dans le dictionnaire en ligne ARTES; (3) la remise des dossiers: des instructions détaillées sont données en début de semestre, environ 3 mois avant la date finale de remise des dossiers.

4. Recherches et constitution des corpus

L'intérêt du recours au corpus, notamment pour la traduction spécialisée, a été souligné à maintes reprises et n'est plus à démontrer (Baker 1993; Olohan 2004; Morin et Daille 2006, 2012; Kübler 2011). Rappelons simplement les propos de L'Homme (2004: 119):

La recherche terminographique repose principalement sur le contenu de textes de spécialité. La collecte d'une documentation représentative du domaine dont on souhaite décrire la terminologie et son exploitation constitue les premières étapes d'une recherche en bonne et due forme. Il est désormais possible de réunir une documentation abondante en format électronique dans des délais raisonnables.

En plus des dictionnaires monolingues ou bilingues et des banques terminologiques, les traducteurs peuvent ainsi extraire de nombreuses données provenant de corpus. Il apparaît cependant que les corpus sont rarement utilisés par les traducteurs ou futurs traducteurs (Durán-Muñoz 2010) dans leur pratique, notamment les corpus en langue cible. Pourtant, la recherche sur corpus offre l'avantage d'une approche textuelle. En outre, ceux qui travaillent sur la langue

de spécialité ne peuvent pas toujours procéder par introspection et sont donc dépendants d'une documentation (Meyer et Mackintosh 1996: 265):

Introspection plays a fundamentally different role in terminography than in lexicography, and this has important implications for corpus terminography. Lexicographers, on the one hand, rely heavily on their own linguistic intuition: they are, after all, producers of the type of language they study. Hence, texts and outside informants are often seen as a means of corroborating personal intuition. Terminographers, on the other hand, are not natural producers of the type of language they study. Faced with a terminological problem, they cannot ask questions such as 'what does X mean to me?', 'would I use X in this way?', but rather 'what does X mean to domain experts?', or 'would domain experts use X in this way?'.

Afin d'élaborer une base de données terminologique et phraséologique dans un domaine de spécialité, le recours au corpus semble donc se justifier. Parallèlement à la conception des corpus, un intervenant extérieur, professeur universitaire et spécialiste des domaines de la diplomatie et des relations internationales, est invité chaque année à donner 15 heures de cours aux étudiants de première année de Master en guise d'introduction à un domaine de spécialité (voir Udovič 2018). Les larges connaissances acquises dans les domaines de la diplomatie et des relations internationales permettent aux étudiants de choisir un domaine de spécialité en connaissance de cause avant de constituer leurs corpus, puis de sélectionner et de traiter correctement les termes et les informations pertinentes à ajouter à la base ARTES. En deuxième année de Master, les étudiants puisent la terminologie dans le domaine des sciences de la vie.

Les corpus étant des aides à l'élaboration de dictionnaires bilingues (Bowker et Pearson 2002), ils sont constitués de textes authentiques. Des instructions détaillées sont fournies aux étudiants concernant la création et l'exploitation du corpus. Pour créer leur propre base de données, les apprenants collectent les ressources à partir de corpus textuels. Ils les versent ensuite pour constituer les données à saisir. Ils construisent ainsi deux corpus comparables: français et slovène. Le corpus comparable est composé d'ensembles de textes, dans des langues différentes, «qui ne sont pas des traductions mutuelles» (Bowker et Pearson: 95; voir aussi L'Homme 2004: 133-134), la comparabilité impliquant un nombre de traits communs entre chaque partie du corpus et le seul trait différent étant la langue de chaque partie. Cette définition impose donc peu de contraintes sur les documents constituant le corpus.

La première année du projet, les étudiants ont commencé à moissonner le web pour en extraire les documents, puis à les convertir en texte brut à l'aide du logiciel d'analyse de corpus Antconc. Mais cette approche s'est avérée longue et fastidieuse. La recherche des corpus s'effectue dorénavant à l'aide du logiciel concordancier Sketch Engine qui permet de regrouper facilement de nombreux documents, son utilisation ne nécessitant pas la conversion du document en texte brut. En outre, cet outil est accompagné d'un langage de requêtes puissant, il fournit rapidement des ressources authentiques et permet de régler

l'affichage des données autour du terme ou de la collocation recherchés. Il s'appuie aussi sur des tableaux regroupant les principaux termes ou cooccurrents classés en fonction de certaines relations (objet, sujet, fréquences, etc.). Les étudiants opèrent donc un tri parmi les ressources proposées par Sketch Engine, et cochent les documents susceptibles d'être intégrés au corpus. Puis, une approche partant des mots clés proposés par le logiciel concordancier est considérée. La recherche se poursuit en fonction des réponses liées à la fréquence d'apparition et/ou à la difficulté, explicite ou implicite, à traduire le terme ou la collocation.

En final, nous notons que les corpus des étudiants (Master 1) portent sur des domaines très variés comme les discours diplomatiques ou politiques, la diplomatie juridique, culturelle, humanitaire ou économique, la politique étrangère, le Brexit, les droits de l'homme, l'impérialisme. etc. Ils sont variables et comptent, en fonction de la langue et du domaine, environ 250 000 à 600 000 mots chacun.

5. **Élaboration du dictionnaire français–slovène en ligne**

La base en ligne ARTES constitue un outil pédagogique destiné à la formation en phraséologie et en terminologie, en plus d'être une ressource lexicographique accueillant de nombreuses informations utiles pour la traduction spécialisée et la rédaction professionnelle. Il est donc indispensable de former les apprenants à l'utilisation de la base qui offre de nombreuses possibilités. L'étudiant peut y construire un simple glossaire ou bien saisir des listes terminologiques ou phraséologiques dans un domaine spécialisé. L'initiation s'effectue en deux temps: d'abord une introduction à la base est effectuée en première année de Master (soit 15h de travaux pratiques sur la base ARTES). En plus de la création des corpus, cela consiste notamment à collecter, saisir et traduire des collocations spécifiques ou génériques. Puis, une application plus poussée est proposée en deuxième année de Master (dans le cadre de 15h de travaux pratiques) où les étudiants produisent des fiches terminologiques dans un domaine de spécialité de leur choix, ces fiches comprenant également la saisie de collocations spécialisées. Ainsi, après la création des corpus, les étudiants recherchent, compilent et traduisent les fiches terminologiques et phraséologiques qu'ils versent dans la base. La conception du dictionnaire débute par la saisie des collocations dans l'onglet «Phraséologie discursive» (Tableau 4). Cela permet une première approche de la base et facilite la saisie ultérieure, plus complexe, de la terminologie dans l'onglet «Terminologie en contexte» (Tableau 6).

5.1 **Création de la base phraséologique discursive**

Des instructions détaillées sont données, au début du semestre, aux étudiants inscrits en première année de Master pour leur apprendre à compiler leurs

ressources, puis à extraire et saisir les collocations qui figureront dans le dictionnaire. L'étudiant doit ainsi réaliser un projet phraséologique à partir d'un domaine relatif à la diplomatie ou aux relations internationales. Après avoir constitué un corpus comparable, cela consiste à verser dans la base ARTES au moins 5 collocations génériques dans chaque langue source et 5 collocations traduites dans chaque langue cible, donc 10 collocations au total pour les deux langues sources ainsi que leurs traductions. Un problème de définition s'est rapidement posé quant à la nature des collocations, les étudiants peinant à distinguer les collocations spécifiques des collocations génériques ou transdisciplinaires. Plusieurs études ont en effet été menées sur la problématique de la traduction des collocations à partir d'un corpus spécialisé (Kübler 2003, Pecman 2007, Volanschi 2007) qui ont conduit à la séparation des collocations spécifiques (associées à la terminologie) et des collocations génériques (associées au discours).

L'onglet «Phraséologie discursive» de la base ARTES (Tableau 4) étant conçu pour enregistrer les collocations génériques qui apparaissent fréquemment dans le discours, l'étape suivante repose sur l'extraction des collocations les plus fréquentes. L'extraction des collocations consiste à compter le nombre de fois où la base et son collocatif apparaissent dans le corpus. Les apprenants effectuent ensuite un étiquetage morpho-syntaxique afin d'identifier le rôle syntaxique à partir de la morphologie de chaque terme, et un filtrage afin d'éliminer les expressions de faibles fréquences, jugées peu informatives ou ne démontrant pas de réel problème de traduction.

Tableau 4: La phraséologie discursive dans la base ARTES

The screenshot shows the 'Dictionnaire ARTES' interface. The main section is titled 'Phraséologie discursive'. It features a search bar and a list of discursive functions. The functions listed include: 'Introduire une transition', 'Préciser l'ordre de son exposition', 'Annoncer un point qui sera discuté plus tard ou plus loin dans le discours', 'Faire un renvoi aux éléments non textuels (tableaux, graphiques, figures...)', 'Faire un renvoi à ce qui vient d'être traité', 'Annoncer le ou les conclusions', 'Evocuer le sujet de son étude', 'Présenter ses objectifs de recherche', and 'Evocuer son positionnement ou le contenu théorique dans lequel s'inscrit le travail'. On the right side, there is a table of 'Collocations associées' with columns for 'Langue', 'Collocation', 'Construction', and 'Discours'. The table contains several entries, such as 'ob upolietaraju nasledenje' (Slovene) for 'discours juridique' and 'informiranje javnosti' (Slovene) for 'discours médiatique'.

Après avoir été saisie dans la base, chaque collocation est présentée dans un dossier par une fiche qui comprend: la collocation dans la langue source et sa traduction en contexte dans la langue cible, le nombre d'occurrences dans le corpus source et dans Google, la fréquence de la collocation dans la langue cible, les structures lexico-grammaticales et les fonctions discursives. Un exemple de fiche est présenté dans le Tableau 5.

Tableau 5: Exemple de fiche comportant la collocation et sa traduction

	Langue source (français)	Langue cible (slovène)
Collocation	permettre de faire vivre	zagotoviti preživetje
Fréquence/vos textes/Google	10/15.500.000	2030
Construction grammaticale	V + prép + groupe verbal	V + N
Portée (Artes)	Discuter un concept ou notion	Discuter un concept ou notion

Le dossier est à remettre à la fin du semestre. Il est composé d'une introduction, des dix fiches correspondant aux collocations génériques intégrées dans la base et d'une conclusion. Dans l'introduction, l'étudiant présente le domaine, justifie l'intérêt de son choix et la méthode de travail sur corpus. Chaque fiche est suivie d'un commentaire qui porte sur le choix et la traduction de la collocation. L'étudiant tire ensuite des conclusions sur le déroulement des tâches qu'il a accomplies.

5.2 Création de fiches terminologiques

L'objectif suivant consiste à créer des fiches terminologiques qui sont versées par les étudiants inscrits en deuxième année de Master dans le dictionnaire en ligne ARTES. La procédure de création des ressources consiste à ajouter manuellement, en tenant compte du contexte, les termes spécialisés et les collocations, puis à identifier leurs équivalents. La saisie du contexte est essentielle pour déterminer le sens et l'usage des unités terminologiques (première étape) qui forment les entrées principales de la base, tandis que les collocations spécifiques (deuxième étape) sont considérées comme des entrées secondaires (Tableau 6). Elles fournissent néanmoins des informations utiles sur les diverses combinaisons de termes dans les situations de communication, notamment dans le domaine des langues de spécialité.

Tableau 6: Terminologie en contexte dans la base ARTES

The screenshot shows the 'DICTIONNAIRE ARTES' interface. At the top, there are three tabs: 'Terminologie en contexte', 'Phraséologie discursive', and 'Recherche multicritères'. Below the tabs, there is a search form with the following elements:

- A dropdown menu for 'Veuillez sélectionner un domaine (facultatif):' with the option '-tous-' selected.
- A checkbox labeled 'Inclure les liens dans les définitions' which is checked.
- A text input field containing 'krizna vez'.
- A button labeled 'Entrer un terme'.
- A dropdown menu for 'Slovène'.

Below the search form, the results are displayed:

1. **Domaine :** *Sciences médicales. Médecine. Hygiène. Pharmacie [CDU 2004]*

krizna vez (Sujet : Krizna vez) - Validé : Non - Auteur de la fiche: Katja Mavrič Bordon, Etudiant(e) en Master de traduction, Université de Ljubljana (2017-2018)

Conformément aux études sur le sujet (L'Homme 2004: 120-123), en établissant leurs corpus, les apprenants sont amenés à collecter et caractériser un ensemble de textes spécialisés jugés pertinents pour l'application visée, à savoir: extraire une liste de termes (simples ou complexes) du domaine de spécialité; repérer les collocations relatives à ces termes et les contextes contenant des éléments définitoires ou des indications sémantiques; repérer les relations sémantiques; vérifier plusieurs équivalents d'un terme; repérer les éléments définitoires, les variantes terminologiques, les indices de relation taxinomique, les synonymes, les co-hyponymes ou des antonymes, les indices de relation méronymique, les termes et leurs cooccurrents.

Par conséquent, l'étudiant doit fournir au moins 10 fiches terminologiques complètes (dites fiches longues) dans chacune des deux langues abordées pour les termes portant sur un domaine spécialisé des sciences de la vie. Le choix du candidat terme est conditionné par le fait que ce terme peut être recherché dans un dictionnaire ou un glossaire pour plusieurs raisons: soit on n'en connaît pas le sens ou on hésite sur l'équivalent, soit il est difficile de prouver que la traduction est attestée ou encore que la phraséologie utilisée est correcte. Les apprenants n'ayant que peu d'expérience dans la construction de fiche terminologique, l'une des grandes difficultés réside dans le choix du terme spécialisé. Nous constatons en effet la tendance chez les apprenants à saisir un terme spécialisé qui est similaire dans les deux langues et ne pose pas de réel problème de traduction: par exemple, *paléanthropologie* et *paleoantropologija*.

Comme pour la partie phraséologie discursive, les termes (simples ou complexes) sont extraits de deux corpus comparables. Ces fiches sont ensuite saisies dans la base. La fiche complète (Tableau 7) comporte les éléments suivants: le terme en langue source; une indication sur la langue source; la catégorie grammaticale du terme; le(s) domaine(s); le(s) contexte(s); les sources du/des contextes; la définition (sa source); les collocations; les concurrents; les antonymes; les hyperonymes, méronymes, etc.; le statut du terme; la note technique; la note linguistique; la note de traduction; l(es) équivalent(s). Le même schéma est saisi pour la langue cible.

Tableau 7: Exemple de fiches terminologiques comparables

Fiche 1 — langue française	Fiche 2 — langue slovène
LIGAMENT CROISÉ <i>nom, masculin, terme</i>	KRIZNA VEZ <i>nom, féminin, terme</i>
Domaine(s): — Sciences médicales. Médecine. Hygiène. Pharmacie [CDU 2004]	Domaine(s): — Sciences médicales. Médecine. Hygiène. Pharmacie [CDU 2004]
Auteur de la fiche: Katja Mavrič Bordon, Étudiant(e) en Master de traduction, 2017–2018 Définition(s): — deux ligaments postérieurs du genou: le ligament croisé antérieur et le ligament croisé postérieur [Source: ligaments croisés — Le grand dictionnaire terminologique (http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=8414667)]	Auteur de la fiche: Katja Mavrič Bordon, Étudiant(e) en Master de traduction, 2017–2018 Définition(s): — vezivo v obliki snopa ali ovojnice, ki povezuje dve kosti ali več kosti, hrustanec ali druge strukture [Source: križna vez — Termania (http://www.termania.net/slovarji/slovenski-medicinski-slovar/5524711/ligament/)]
Contexte(s) / Expressions spécialisées (collocations): — Le ligament croisé postérieur (LCP) est situé au milieu du genou (il fait partie du "pivot central"). Placé dans l'échancrure du fémur, véritable cavité au milieu du genou. [Source: Vincent Chassing — LCP (http://genou.com/lcp/lcpdetaille.ht)] — La <i>reconstruction chirurgicale</i> du ligament croisé antérieur , dont le but est de rétablir la stabilité et la biomécanique du genou lésé, est un sujet central en traumatologie du sport. [Source: V. B. Duthon, G. Messerli, J. Menetrey — Reconstruction du ligament croisé antérieur: indications et techniques — Revue Médicale Suisse — volume 4 (https://www.revmed.ch/RMS/2008/RMS-184/Reconstruction-du-ligament-croise-anterieur-indications-et-techniques) (2008)] — Le ligament croisé antérieur (LCA) est un stabilisateur essentiel de l'articulation du genou. Il empêche la translation antérieure du tibia par rapport au fémur et agit comme frein secondaire à sa rotation Rekonstrukcija sprednje križne vezi — Artros (https://artros.si/operativni_posegi/rekonstrukcija_sprednje_krizne_vezi/)	Contexte(s) / Expressions spécialisées (collocations): — <i>Zadnja križna vez</i> je debelejša in dvakrat močnejša od ACL. Predstavlja os za zunanjo in notranjo rotacijo kolena. [Source: POŠKODBA SPREDNJE KRIŽNE VEZI KOLENA — GoFit (https://www.gofit.si/poskodba-sprednje-krizne-vezi-kolena/)] — <i>Sprednja križna vez</i> veže sprednjo interkondilarno kotanjico golenice z medialno ploskvijo lateralnega kondila stegenice. Ima izredno pomembno stabilizacijsko vlogo, ker preprečuje drsenje golenice navzpred, preveliko zunanjo rotacijo in hiperekstenzijo kolena. [Source: POŠKODBA SPREDNJE KRIŽNE VEZI KOLENA — GoFit (https://www.gofit.si/poskodba-sprednje-krizne-vezi-kolena/)] — Artroskopska <i>rekonstrukcija križnih vezi</i> z avtolognim kitnim presadkom je uveljavljena metoda izbora za zdravljenje pretrganja sprednje križne vezi. Pri posegu kirurg nadomesti pretrgano sprednjo križno vez s kitnim presadkom, običajno z avtologno kito mišice semimembranosus in gracilis, ki

	jo odvzame med posegom iz zadnje stegenske lože. [Source: Rekonstrukcija sprednje križne vezi — Artros (https://artros.si/operativni_posegi/rekonstrukcija_sprednje_krizne_vezi/)]
Traductions: — (Français) <u>ligament croisé</u> nom, masculin, terme	Traductions: — (Slovène) <u>križna vez</u> nom, féminin, terme

Il est également possible de ne fournir qu'une fiche glossaire qui porte sur les termes du texte. La fiche glossaire contient alors le terme en langue source; une indication sur la langue source; la catégorie grammaticale du terme; le domaine; la définition (sa source); les collocations; les concurrents; la note de traduction; l(es) équivalent(s) et inversement pour la langue cible.

En final, l'étudiant remet un dossier à la fin du semestre qui est composé d'une introduction où il présente le domaine, justifie l'intérêt de son choix et la méthode de travail sur corpus. Dans la partie commentaire, il explique et développe sa démarche terminologique, aussi bien pour la traduction que pour la collecte des sources et l'élaboration des ressources terminologiques. Enfin, l'apprenant tire des conclusions sur son travail. D'autres activités pratiques peuvent également être proposées ou consignées dans le dossier. L'étudiant peut, par exemple, tenir un journal durant le semestre où il consigne en détail le déroulement de son travail, les problèmes qu'il a rencontrés, comment il les a résolus ou pas, ou encore ajouter des suggestions d'adaptation de la méthode employée au traitement de son domaine. Ces informations lui permettent de rédiger au fur et à mesure une analyse synthétique des particularités terminologiques du domaine choisi. Il peut également construire l'arborescence du domaine ou/et traduire une dizaine de lignes, voire plus, à partir d'un texte de son choix, etc. Par ailleurs, des présentations orales à l'aide de powerpoint peuvent permettre un échange en langue étrangère et un partage des connaissances acquises dans un domaine particulier.

6. Synthèse des résultats

6.1 Fiches phraséologiques

À l'issue du travail de création des fiches phraséologiques, nous notons que les étudiants ont saisi et traduit principalement des collocations génériques appartenant aux structures lexico-grammaticales suivantes:

- (1) Vb+N: les catégories grammaticales sont les mêmes dans les deux langues (*sprejeti ukrepe/adopter les mesures*).
- (2) Adj+N qui correspond principalement à la structure grammaticale française N+Adj (*pravična obravnava/le traitement équitable*). Toutefois, certaines

collocations de cette catégorie sont également traduites en français par la catégorie grammaticale N+prép (de)+N (*tiskovna konferenca/conférence de presse*).

- (3) La troisième catégorie la plus fréquente concerne la construction grammaticale slovène N+N qui est traduite par quatre catégories grammaticales différentes:
- N+prép+N (*varstvo potrošnikov/protection du consommateur*)
 - N+adj (*združitev družine/regroupement familial*)
 - Vb+prép+N (*odpoved imuniteti/renoncer à l'immunité*)
 - Construction nominale (*sklenitev sporazuma/la mise en place de l'accord*)

La première catégorie (Vb+N) est la plus fréquente, à raison de 51%; la deuxième catégorie comprenant la structure française Adj+N traduite par la structure slovène inversée équivaut à 19% et les autres structures à 30%.

6.2 Fiches terminologiques

Les dossiers remis par les étudiants nous permettent de recenser et de classer quelques différences. La médecine étant un domaine apprécié des étudiants, nous prenons l'exemple des termes médicaux français et slovènes. Les résultats ont été mis en commun et nous en présentons un extrait dans le Tableau 8:

Tableau 8: Récapitulatif des résultats des étudiants dans le domaine de la médecine

Type	Exemples tirés des corpus des étudiants
1. Termes savants presque identiques: différences orthographiques (chute de l'accent aigu ou grave)	
1a. même prononciation, la finale est une consonne en slovène	carcinome / karcinom amyloïde / amiloid auxine / avksin papillomavirus / papilomavirus viroïde / viroid cystatine C / cistatin C cytostatique / citostatik allèle / alel amphotère / amfoter
1b. ajout du — <i>a</i> final en slovène	coccidiose / kokcidioza myxomatose / miksomatoza papillomatose / papilomatoza agnosie / agnozija cyanose / cianoza épilepsie / epilepsija

<p>1c. la terminaison <i>-gie</i> [ʒi] devient <i>-gija</i> [gija] en slovène; la constrictive prépalatale [ʒ] passe à l'occlusive vélaire [g] en position initiale ou interne</p> <p>1d. dénasalisation et terminaison en <i>-a</i> pour le substantif, et en <i>-čen</i> pour l'adjectif slovène</p> <p>1e. suffixes différents <i>-té</i> = <i>-ost</i> <i>-isque</i> = <i>-iskus</i> <i>-ite</i> = <i>-itis</i> <i>-isme</i> = <i>-ika</i> <i>-tion</i> et <i>-sion</i> = <i>-ija</i></p> <p>2. Inversion grammaticale: N+adj. = adj.+N</p> <p>3. Termes différents 3a. termes savants différents</p> <p>3b. termes vernaculaires différents</p> <p>3c. expressions françaises traduites par un ou plusieurs termes slovènes</p>	<p>otorrhée / otoreja néphrectomie / nefrektomija céphalométrie / cefalometrija névrologie / nevroglija</p> <p>génom / genom diplégie / diplegija éthologie / etologija</p> <p>démence / demenca endosymbiose / endosimbioza nevrologique / nevrolingvističen</p> <p>eusocialité / evsocialnost ménisque / meniskus arthrite / artritis eugénisme / evgenika abscission / absizija rémission / remisija démyélinisation / demielinizacija malocclusion / malokluzija</p> <p>ictère nucléaire / nuklearni ikterus cortex préfrontal / prefrontalni korteks neurosciences cognitives / kognitivna nevroznanost</p> <p>axone / nevrit chimiothérapie / kemoterapija thyroïde / ščitnica hémorragie / krvavitev homolatéral / ipsilateralen surrénale / nadledvičnica</p> <p>gale / garje rechute / relaps nutriment / hranilo prématurité / nedonošenost déjection / blato</p> <p>appareil urinaire / sečila enzyme de restriction / restriktaza ligament croisé / križna vez plaque sénile / senilna leha infirmité motrice cérébrale / cerebralna paraliza dégénérescence neurofibrillaire / nevrofibrilarna pentlja surexpression du gène / čezmerno izražanje gena noyaux gris centraux / bazalni gangliji v osrednjem delu možganov substance blanche / bela možganovina</p>
---	---

Nous notons que les étudiants n'ont saisi aucun terme parfaitement identique dans les deux langues. Le Tableau 8 indique que les termes savants sont les plus fréquents dans les deux langues, même si en français la différence est plus significative (90%) qu'en slovène (80%). La majorité des termes savants prennent une orthographe différente tandis que d'autres diffèrent de par leur traduction. La plupart des termes et des collocations sont traduits par un emprunt savant qui rappelle parfois le mot d'origine latine ou grecque: par exemple, le terme *ictère* vient du latin *icterus* ou du grec *ikteros*, ce qui donne en slovène *ikterus*. On en conclut que le slovène médical semble emprunter directement au latin ou au grec.

7. Conclusion

La création du projet lexicographique, la compilation de données, l'activité de traduction, l'approche contrastive, etc. se référant à la paire de langues français–slovène est une démarche innovante en Slovénie. L'originalité du dictionnaire ARTES réside dans son approche globale qui réunit des éléments linguistiques, terminologiques, phraséologiques et discursifs pertinents pour la traduction et la communication en langue de spécialité. La base de données offre un cadre intéressant pour une approche bilingue ou multilingue. Elle permet en outre de mener des recherches sur une variété de phénomènes linguistiques. Dans le cas présent et d'une part, elle nous a permis de développer une nouvelle approche pédagogique en lien avec le futur contexte professionnel de l'étudiant. En deux ans, cette expérience lexicographique a permis de créer dans la base ARTES plus de 200 fiches terminologiques et plus de 400 fiches phraséologiques français–slovènes. D'autre part, cela permet de mener des recherches à partir des données terminologiques et phraséologiques qui sont stockées par les apprenants dans la base. La base peut ainsi être utilisée comme modèle pour l'analyse de la terminologie et de la phraséologie discursive pour la paire français–slovène dans le domaine de la traduction spécialisée.

Cette approche lexicographique s'est avérée être une étape à la fois formatrice et pertinente du point de vue de la traduction et de la communication spécialisée. L'étudiant met en pratique une méthode de traitement lexicographique, développée par une université française, ce qui permet par ailleurs d'approfondir à plus ou moins long terme les liens inter-universitaires. Il apprend à développer une réflexion méthodologique, mais il pourra aussi être en mesure d'adapter si nécessaire les méthodes et les outils à ses propres besoins. Enfin, les dossiers remis par l'étudiant mettent en évidence la progression de la réflexion terminologique et phraséologique, mais aussi sa maîtrise de la langue française et ses compétences en traduction. Cela permet de mettre en évidence les points forts des dossiers, de tirer des conclusions lexicographiques sur un ou plusieurs domaines spécialisés, mais aussi de souligner le bilan individuel et la progression de l'étudiant dans son apprentissage d'une

langue étrangère et d'une méthode de traduction qui diffère de la méthode classique.

Remerciements

La compilation de ressources lexicographiques français-slovène dans la base de données ARTES a bénéficié du soutien du Département de linguistique appliquée et d'études interculturelles — UFR EILA — de l'Université Paris Diderot. Nous tenons à remercier l'équipe de recherche du département EILA, notamment Mojca Pecman, Nicolas Froeliger, Natalie Kübler, Geneviève Bordet, Alexandra Volanschi et Brice Bricaud pour la mise à disposition de la base de données, leurs concours précieux et leur participation active au projet.

Références bibliographiques

- Baker, M.** 1993. *Corpus Linguistics and Translation Studies. Implications and Applications*. Baker, M., G. Francis and E. Tognini-Bonelli (Éds.). 1993. *Text and Technology: In honour of John Sinclair*: 233-250. Philadelphia: John Benjamins
- Boers, F., M. Demecheleer et J. Eyckmans.** 2004. Etymological Elaboration as a Strategy for Learning Idioms. Bogaards, P. and B. Laufer (Éds.). 2004. *Vocabulary in a Second Language: Selection, Acquisition and Testing*: 53-78. Amsterdam: John Benjamins.
- Bowker, L. et J. Pearson.** 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.
- Cavalla, C.** 2009. La phraséologie en classe de FLE. *Les Langues Modernes*. Association des professeurs de langues vivantes (APLV), pp. en ligne. <https://hal.archives-ouvertes.fr/hal-00699916>.
- Detry, F.** 2008. Pourquoi les murs auraient-ils des oreilles? Vers un apprentissage par l'image des expressions idiomatiques en langue étrangère. *Synergies Espagne* 1: 205-218.
- Durán-Muñoz, I.** 2010. Specialized Lexicographical Resources: A Survey of Translators' Needs. Granger, S. and M. Paquot (Éds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of ELEX2009*: 55-66. Louvain-la-Neuve: Presses universitaires de Louvain.
- Galisson, R.** 1983. *Des mots pour communiquer: éléments de lexicométhodologie*. Paris: CLE International.
- Galisson, R.** 1991. *De la langue à la culture par les mots*. Paris: CLE.
- Gantar, P.** 2015. *Leksikografski opis slovenščine v digitalnem okolju*. [Dictionary of Modern Slovene: Problems and Solutions.] Zbirka Sporazumevanje. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Gorjanc, V., P. Gantar, I. Kosem et S. Krek (Éds.).** 2017. *Dictionary of Modern Slovene: Problems and Solutions*. Prevodoslovje in uporabno jezikoslovje. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Grad, A.** (1971) 2004. *Francosko-slovenski slovar*. Ljubljana: Cankarjeva založba.
- Jakobsen, A.L.** 1999. Logging Target Text Production with Translog. Hansen, G. (Éd.). 1999. *Probing the Process in Translation: Methods and Results*: 9-20. Copenhagen Studies in Language 24. Copenhagen: Samfundslitteratur.

- Jakobsen, A.L.** 2000. Understanding the Process of Translation: The Contribution of Time-Delay Studies. Englund Dimitrova, B. (Éd.). 2000. *Översättning och tolkning*: 155-172. Stockholm: ASLA.
- Jesenik, V. et N. Dembskij.** (1990) 2008. *Slovensko–francoski slovar*. Ljubljana: Državna založba Slovenije.
- Kübler, N.** 2003. Corpora and LSP Translation. Zanettin, F., S. Bernardini and D. Stewart (Éds.). 2003. *Corpora in Translator Education*: 25-42. Manchester: St Jerome.
- Kübler, N.** 2011. *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang.
- Kübler, N. et M. Pecman.** 2012. The ARTES Bilingual LSP Dictionary: From Collocation to Higher Order Phraseology. Granger, S. and M. Paquot (Éds.). 2012. *Electronic Lexicography*: 187-209. Oxford: Oxford University Press.
- Ledinek, N, K. Ahačič et A. Perdih.** 2015. *Fran: slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, Vodnik*. [A Guide to Fran: The Dictionaries from the Fran Ramovš Institute of the Slovenian Language in the Research Center of the Slovenian Academy of Sciences and Arts.] Zbirka Fran. Ljubljana: Založba ZRC.
- L'Homme, M.-C.** 2004. *La terminologie: principes et techniques*. Montréal: Les Presses de l'Université de Montréal.
- Mavrič Bordon, K.** 2019. *Uporabni elektronski viri za prevajanje terminologije s področja krasoslovja: magistrsko delo*. Ljubljana: Filozofska fakulteta.
- Meyer, I. et K. Mackintosh.** 1996. The Corpus from a Terminographer's Viewpoint. *International Journal of Corpus Linguistics* 1(2): 257-285.
- Morin, E. et B. Daille.** 2006. Comparabilité de corpus et fouille terminologique multilingue. *Traitement Automatique des Langues. ATALA* 47(1): 113-136.
- Morin, E. et B. Daille.** 2012. Compositionnalité et contextes issus de corpus comparables pour la traduction terminologique. *TALN* 2: 141-154.
- Olohan, M.** 2004. *Introducing Corpora in Translation Studies*. London/New York: Routledge.
- Pavičić Takač, V.** 2008. *Vocabulary Learning Strategies and Foreign Language Acquisition*. Second Language Acquisition 27. Clevedon/Buffalo/Toronto: Multilingual Matters.
- Pecman, M.** 2007. Phraseology and Didactics: Lexical Combinatorics as a Teaching Tool of English for French Students in Scientific Disciplines. Kržišnik, E. and W. Eismann. 2007. *Frazeologija v Jezikoslovju in Drugih Vedah/Phraseologie in der Sprachwissenschaft und anderen Disziplinen/Phraseology in Linguistics and Other Branches of Science*: 349-363. Strunjan: European Society of Phraseology (EUOPHRAS).
- Pecman, M.** 2012. Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques. *Terminology, Phraseology and Translation*. Special Issue of *The Journal of Specialized Translation (JoSTrans)* 18: 113-138.
- Pecman, M. et C. Gledhill.** 2018. How Trainee Translators and Their Teachers Deal with Phraseological Units in the ARTES Database. *Equivalences*. 45(1-2): 235-257.
- Pecman, M. et N. Kübler.** 2011. ARTES: An Online Lexical Database for Research and Teaching in Specialized Translation and Communication. *Proceedings of the First International Workshop on Lexical Resources (WoLeR) 2011 at ESSLLI, August 1-5, 2011*: 87-93. Ljubljana: Ljubljana University Press.
- Perko, G.** 2006. *Francosko–slovenski, slovensko–francoski žepni slovar*. Ljubljana: Cankarjeva založba.

- Perko, G.** 2007. *Slovensko-francoski evropski slovar*. Ljubljana: Cankarjeva založba.
- Simonnet, D. et A. Tutin.** 2019. Utiliser la linguistique de corpus pour renforcer les compétences des étudiants dans l'analyse syntaxique. *Langue française* 203: 83-99.
- Snoj, J.** 2004. *Tipologija slovarske večpomenskosti slovenskih samostalnikov*. [The Lexicographic Treatment of Polysemous Nouns in Slovenian.] Zbirka Linguistica et philologica. Ljubljana: Založba ZRC SAZU.
- Srebnik, A.** 2015. *Jezikovnotehnoški postopek obračanja dvojezičnih slovarjev*. [The Technology and Linguistics behind the Process of Reversing Bilingual Dictionaries.] Praha: Verbum.
- Szczepaniak, R. et R. Lew.** 2011. The Role of Imagery in Dictionaries of Idioms. *Applied Linguistics* 32(3): 323-347.
- Tréville, M.-C.** 2000. *Vocabulaire et apprentissage d'une langue seconde: recherches et théories*. Montréal, Québec: Éditions Logiques.
- Udovič, B.** 2018. *Zgodovina (gospodarske) diplomacije*. Ljubljana: Založba FDV.
- Volanschi, A.** 2007. Outils informatiques et ressources électroniques pour les traducteurs. *Tribune internationale des langues vivantes (TILV)* 43(2): 24-39.
- Vrbinc, A., D.M.T.Cr. Farina et M. Vrbinc.** 2018. Objectivity, Prescription, Harmlessness, and Drudgery: Reflections of Lexicographers in Slovenia. *Lexikos* 28: 361-383.
- Žagar Karer, M.** 2011. *Terminologija med slovarjem in besedilom: analiza elektrotehniške terminologije*. [Terminology from the Text to the Dictionary: Analysis of Electro-technical Terminology.] Zbirka Linguistica et philologica 26. Ljubljana: Založba ZRC.

Normative Labels in Two Lexicographic Traditions: A Slovene–English Case Study

Marjeta Vrbinc, *Faculty of Arts, University of Ljubljana, Slovenia*
(marjeta.vrbinc@ff.uni-lj.si)

Danko Šipka, *School for International Letters and Cultures, Arizona State
University, Tempe Campus, USA* (danko.sipka@asu.edu)

and

Alenka Vrbinc, *School of Economics and Business, University of Ljubljana,
Slovenia* (alenka.vrbinc@ef.uni-lj.si)

Abstract: This article presents and discusses the findings of a study conducted with the users of Slovene and American monolingual dictionaries. The aim was to investigate how native speakers of Slovene and American English interpret select normative labels in monolingual dictionaries. The data were obtained by questionnaires developed to elicit monolingual dictionary users' attitudes toward normative labels and the effects the labels have on dictionary users. The results show that a higher level of prescriptivism in the Slovene linguistic culture is reflected in the Slovene respondents' perception of the labels (for example, a stronger effect of the normative labels, a higher approval for the claim about usefulness of the labels, a considerably lower general level of acceptance for the standard language) when compared with the American respondents' perception, since the American linguistic culture tends to be more descriptive. However, users often seek answers to their linguistic questions in dictionaries, which means that they expect at least a certain degree of normativity. Therefore, a balance between descriptive and prescriptive approaches should be found, since both of them affect the users.

Keywords: GENERAL MONOLINGUAL DICTIONARY, PRESCRIPTIVISM, NORMATIVITY, DESCRIPTIVISM, NORMATIVE LABELS, PRIMARY EXCLUSION LABELS, SECONDARY EXCLUSION LABELS, USE OF LABELS, USEFULNESS OF LABELS, (UN)LABELED ENTRIES

Opsomming: Normatiewe etikette in twee leksikografiese tradisies: 'n Sloweens–Engelse gevallestudie. In hierdie artikel word die bevindings van 'n studie wat uitgevoer is op gebruikers van eentalige Sloweense en Amerikaanse woordeboeke voorgelê en bespreek. Die doel van die artikel was om te bepaal hoe moedertaalsprekers van Sloweens en Amerikaanse Engels sekere normatiewe etikette in eentalige woordeboeke interpreteer. Die data is verkry deur vraelyste te ontwikkel wat die ingesteldheid van gebruikers van eentalige woordeboeke teenoor normatiewe etikette en die effek daarvan op woordeboekgebruikers belig. Die resultate dui daarop dat 'n hoër vlak van preskriptiwiteit in die Sloweense taalkundige kultuur waargeneem word in die Sloweense respondente se begrip van die etikette (byvoorbeeld, 'n sterker invloed van die normatiewe etikette, 'n groter goedkeuring van die stelling oor die bruikbaarheid van die etikette, 'n aansienlik laer algemene aannamevlak by die standaardtaal) wanneer dit ver-

gelyk word met die persepsie van die Amerikaanse respondente, aangesien die Amerikaanse taalkundige kultuur geneig is om meer deskriptief te wees. Gebruikers soek egter dikwels in woordeboeke antwoorde op hul taalkundige vrae, wat daarop dui dat hulle ten minste 'n sekere graad van normatiewiteit verwag. Aangesien albei benaderings die gebruikers beïnvloed, moet daar 'n balans tussen deskriptiewe en preskriptiewe benaderings gevind word.

Sleutelwoorde: ALGEMENE EENTALIGE WOORDEBOEK, PRESKRIPTIEWEIT, NORMATIEWEIT, DESKRIPTIEWEIT, NORMATIEWE ETIKETTE, PRIMÊRE UITSLUITINGSETIKETTE, SEKONDÊRE UITSLUITINGSETIKETTE, GEBRUIK VAN ETIKETTE, BRUIKBAARHEID VAN ETIKETTE, (ON)GEËTIKETTEERDE INSKRYWINGS

1. Introduction

While the cultural tradition of the Slovene language features linguistically strong prescriptive attitudes and practices, in American English, the linguistic prescriptivism has been much more subdued. This disparity leads to questions about the differences in the effect that the means of linguistic prescriptivism have on the speakers of these two languages. In order to find answers to these questions, a study was carried out, which was aimed at discerning the effect the normative labels have on the speakers of the respective languages.

Normative labels are the qualifiers, and they include *slang*, *incorrect usage*, etc., which exclude single words, one or more of its senses, or one or more of its phonetic or inflectional forms from their usage in the formal standard language¹. Usage, in turn, in a broader sense, pertains to relationships of words or their features (their senses, their phonetic or inflectional forms) to their linguistic context. They range from the required morphosyntactic to semantic combinations to the effects the words create in that context. In a narrower sense, usage does not encompass any phonetic, phonological, prosodic, morphosyntactic, and core semantic features. In this sense, usage pertains to a higher, lower, or absolute affinity of the word or its feature toward certain contexts (e.g., toward a certain period in time or in regard to one geographical region) or a certain contextual effect (e.g., offensiveness, obscenity, etc.). Usage will be employed here in this narrower sense. Usage labels, having in mind that the narrower sense of the term usage is employed here, pertain to compact customary references to the affinity of the words or their features, and toward certain contexts or contextual effects. Hausmann (1989: 651) has identified the following criteria for deploying usage labels: time (e.g., *archaism*), place (e.g., *regionalism*), nationality (e.g., *foreign word*), medium (e.g., *colloquial*), sociocultural (e.g., *slang*), formality (e.g., *informal*), text type (e.g., *poetic*), technicality (e.g., *biology*), frequency (e.g., *rare*), attitude (e.g., *facetious*), and normativity (e.g., *non-standard*). The majority of these types of labels have the capacity of serving as normative exclusion labels given that they have the potential of excluding a lexeme or its meaning from the standard language variety in its narrower sense (those contexts such as media outlets, courts, and similar others, but not related

to literature or the spontaneous common parlance, such as creative composition). The labels that exclude most directly are the ones based on normativity. Straightforward exclusion can additionally be achieved by the criteria time and place as well as by sociocultural criteria. All the aforementioned labels exclude any given word from the modern standard language variety in its narrower sense. Some formality labels (e.g., *informal*) and attitude labels (e.g., *derogatory*) have the potential of being indirectly excluded. Their primary role is to mark the attitude, but they also have a certain potential of being interpreted by the users as exclusion labels. Other labels used in dictionaries do not have this potential. They just mark an area or a feature within a standard language variety.

The following two types of normative exclusion labels are present in all monolingual descriptive dictionaries:

- a. Primary exclusion labels, e.g.: *dialectal*, where the dictionary compiler's goal is to clearly exclude the word, one or more of its senses, or one or more of its forms from the standard language variety;
- b. Secondary exclusion labels, e.g.: *obscene*, where the compiler uses the label to mark something else (the attitude in this case), but the users can interpret the label as being excluded from the standard language variety.

In hands-on application, these two types blend into one another. There is a continuum of exclusiveness from a clear exclusion from the standard language variety on one end, to a weak exclusion effect on the other end of the same continuum.

The present study attempts to answer the following research questions:

- a. How does the type of a normative label influence the level of acceptability of labeled words in the standard language variety?
- b. What are the users' attitudes toward normative dictionary labels?
- c. How does a lemma or one of its senses get labeled or how does it not affect the level of acceptability of labeled words in the standard language variety?
- d. What are the differences between Slovene and American English users in items of the content of the above questions a–c?

2. Theoretical background

2.1 The prescriptivist vs descriptivist divide

The battle between prescriptivist and descriptivist approaches has existed in lexicography since the very beginning of dictionary-making activities (Finegan 2020). The prescriptivist tradition has been much more prevalent than the descriptivist one not only in lexicography but also in other branches of linguistics, which is reflected in various types of reference books produced by linguists and con-

sulted by members of a language community to resolve their linguistic issues and dilemmas. The two most common reference books considered essential by any language community are a dictionary and a grammar book. Besides that, the linguistic tradition of Slavonic languages has relied heavily on another type of reference book called the manual of orthography, which is the closest counterpart of a manual of style in the American tradition. This means that not only lexicographers but also grammarians and those who work on orthography are also faced with the same prescriptivist vs descriptivist controversy.

Traditionally, the making of dictionaries was associated with conservatism and historically, dictionaries were compiled with prescriptivism in mind (Balteiro 2011: 278). Because the term prescriptive has a rather negative connotation, metalexicographers prefer the term normative to prescriptive (Béjoint 2010: 80). A prescriptive dictionary is typically a dictionary with only 'acceptable' words included in the wordlist, meaning that there are no words that can be considered 'unacceptable'. If 'unacceptable' words or senses are not omitted altogether, the dictionary clearly labels the disapproval (Béjoint 2010: 80). Lexicographers compiling a general mono- or bilingual dictionary should know that all the normative elements in the dictionaries may affect (some) users who are likely to take them into consideration. This is why a lexicographer should include normative information on purpose, which is possible mainly when it comes to orthography and pronunciation. To a limited extent, this may also apply to some other areas (Zgusta 1971: 290). The lexical meaning of a lemma or its senses is much more problematic, but even here the normative information will be followed by more "hesitating users" (Zgusta 1971: 291). According to Milkowski (2013: 176), not only purists but also users in general are often emotionally involved in the issues concerning the correct use; on the other hand, not all stable linguistic patterns found in a corpus with a high frequency can become standard. At the same time, users often ignore the authority of the experts and it is also true that some norms set by experts are disregarded by most users of a language, which is why traditional prescriptivism has no place in contemporary linguistics. Prescriptivism as practiced in contemporary linguistics is well explained by Armstrong and Mackenzie (2015: 29), who claim that it is "the expression of an ideology, one which is anchored both in human psychology and in a broader, hierarchical conception of how society should be organized."

Nowadays, linguistics in general and especially its corpora-based part, favors descriptivism (Finegan 2003), which is indicative of a broader trend of abandoning prescriptivism in linguistics. This trend is also observed in lexicography, since lexicographers mostly claim that the dictionary they compiled only describes the language, as it is without giving its users advice on what is correct or desired and without making a distinction between good or bad. However, users consulting a dictionary often try to find out what is correct and what is incorrect, which clearly shows that they do expect a certain degree of prescriptivism. Therefore, a dictionary cannot be based on a descriptive

approach only, which is what Béjoint (2010: 80) clearly emphasizes by saying: "Pure descriptivism is impossible." The lexicographers working on a particular dictionary have to decide what to include and what to exclude from it and this is why all dictionaries represent a certain model of linguistic usage. This is also in line with Yong and Peng (2007: 116), who claim that "a dictionary that rids itself of all traces of value-judgement and does not give stylistic and register labels of any kind to particular lexical items is bound to stir up severe criticisms and raise a public outcry."

When it comes to setting the norm, frequency in large databases is of the utmost importance, but lexicographers using corpora for the dictionary compilation process have to be aware of the fact that due to qualitative limits of the corpora used, absolute objectivity of the description is not to be expected (Jarošová and Benko 2012: 259). The most common criteria for the inclusion of material into dictionaries are authenticity and representativeness. However, any dictionary on the market, even if it claims to be descriptive, has a great normative power, which means that even if one particular incorrect spelling of a word appears with a high frequency in a representative corpus, that does not mean that such a spelling would merit inclusion in a general dictionary (Svensén 2009: 67). It is essential to avoid using simple approaches to the description of a language. Both descriptivism based on corpus analyses and prescriptivism based on expert authority seem to disregard the dynamics of language change (Miłkowski 2013: 175). The two approaches use different norms. For descriptivists, the most important criterion is quantity, which means that a form is acceptable if it is used by a considerable number of users. However, it is difficult to determine the minimum number of users or uses of a certain linguistic form to be considered acceptable (Béjoint 2010: 80). Frequency of use and statistically relevant use are not really the only criteria that should be met for a linguistic phenomenon to become an accepted norm. Prescriptive dictionaries use a qualitative norm on the basis of the use of language by the most proficient users. The issue that needs to be addressed is the uncertainty of which model to choose. As Béjoint (2010: 80) states: "It is always written rather than spoken, and literary, and typically from some time before the compilation of a dictionary. The qualitative norm corresponds to the eighteenth-century type of corpus, giving priority to aesthetic judgement over frequency." Very often, users of a language simply disregard the rules set by experts as these may often be outdated in terms of the development of a language (Miłkowski 2013: 177).

The issue with prescriptivism and descriptivism is often conditioned by tradition and cultural factors. Most certainly, culture plays an important role and determines different approaches used in different countries, societies or language communities. In nations with an academy of science, such as the Italian Accademia della Crusca, the French Academy (*Académie française*), the Royal Spanish Academy (*Real Academia Española*) and Slovenian Academy of Sciences and Arts (*Slovenska akademija znanosti in umetnosti*), the authority rests with these academies in setting the norm for a language, or more precisely, in

determining what is correct or not in terms of grammar, syntax, spelling or the lexis². The authority of an institution is also reflected in the fact that language users are given the opportunity to pose questions to linguists about how to use the language correctly. This is obvious not only in Slovenia, but also in other countries with an authoritative institution (e.g., Poland, where numerous linguistic consulting websites exist and are popular (Milkowski 2013: 176)). In Slovenia, the perception of authority was confirmed in a study carried out in 2018 by one of the interviewees who explained that one of her duties as an employee of the Academy was to provide answers to questions language users asked within the framework of the consulting service of the Academy. She stated that the general public had great confidence and trust in the authority of the Academy and its staff because people believed her, even when she herself was not entirely satisfied with the answers she provided (Vrbinč et al. 2018: 374). Here, a question can be raised why language users use consulting services if it is usage that determines linguistic standards (for more on this issue, see Milkowski 2013).

2.2 Prescriptivism and descriptivism in the lexicographic tradition of the English language

The lexicography of English can be traced back to the Old English period (more Anglo-Saxon and Germanic than its modern counterpart), around 1150 when words in Latin manuscripts were glossed into English. Glosses were later collected and appeared in glossaries which can be considered the forerunners of modern bilingual dictionaries (for more on this issue, see Landau 2001: 45; Béjoint 2010: 52-55). The first true English monolingual dictionary entitled *A Table Alphabeticall, conteyning and teaching the true writing, and vnderstanding of hard vsuall English wordes, borrowed from the Hebrew, Greeke, Latine, or French, etc.* was published in 1604 by Robert Cawdr(e)y, who chose the words from Thomas Thomas's Latin-English dictionary (*Dictionarium Linguae et Anglicanae*) published in 1587, as well as from earlier grammar books (Landau 2001: 43). It was not just English whose first monolingual dictionary appeared at the beginning of the 17th century, but also some other European languages got their first monolingual dictionaries at approximately the same time as part of the grammaticalization of vernacular languages, since it became clear that vernacular languages could also have a grammar like Latin and could be codified. The first encyclopedia in English was published in 1728 by Ephraim Chambers (*Cyclopaedia: Or an Universal Dictionary of Arts and Sciences*), who advocated the creation of an authority on language following the examples of France and Italy, where the academies as language regulators had already been established; however, Chambers believed that this was not a lexicographer's job and spoke in favor of a more descriptive approach, since a dictionary should only record usage (Béjoint 2010: 56; 61).

The 18th century was considered a turning point in lexicography due to the development of science and technology as well as due to the fact that more people wrote and more periodicals appeared. In England, the first decade of the 18th century saw the publication of as many as five dictionaries. It is important to stress that it was only in the 18th century when the idea that dictionaries could exert a regulatory influence on a language emerged (Béjoint 2010: 63-64; 79). The main event in English lexicography of the 18th century was the publication of Samuel Johnson's *Dictionary of the English Language: In Which the Words are Deduced from Their Originals, and Illustrated in Their Different Significations, by Examples from the Best Writers, to Which are Prefixed a History of the Language, and an English Grammar* (1755), which was written on the initiative of a group of publishers who asked Johnson to create an authoritative dictionary of English (Lynch 2003). Johnson's *Dictionary* was both descriptive and prescriptive (Landau 2001: 63). It was the pre-eminent British dictionary until the completion of the *Oxford English Dictionary* (1928). It is among the most renowned and influential dictionaries of the English language, its influence extending beyond Britain and beyond English.

Some of the principles used by Johnson (e.g., the use of literary quotations, definitions in historical sequence) are employed by the *Oxford English Dictionary* (OED), which is believed to follow a descriptive approach, but as Brewer (2010: 24) claims the editors of its first edition "included a number of judgements, expressions of opinion, and recommendations on language which now appear to us clearly prescriptive, not descriptive." Béjoint (2010: 102-103) agrees that the OED can be labeled as being a descriptive dictionary, but like all other dictionaries, its lexicographers had to choose words, usages and spellings when different variants existed. In OED1, a paragraph mark (¶) is used by the editors to indicate what they are objecting to in a current (or past) usage of the entry word or one of its senses. The paragraph mark, however, is not always used to indicate value judgements, as sometimes they stand alone in a headnote or other explanatory matter accompanying the definition (Brewer 2010: 25). Another very clear indication of a prescriptive approach used in the OED is that rude words are excluded from the wordlist "motivated by notions of propriety and social nicety akin to those informing the disapproving comments on inelegance and misuse" (Brewer 2010: 27). Besides the linguistic norms, the social aspect exerts considerable influence on usage. Expressions that carry offensive connotations, mostly words relating to sex and excretory functions as well as expressions with unacceptable value judgement regarding sex, race, nationality, social class, religion, etc., belong to the part of the vocabulary of every language that is considered socially sensitive (Svensén 2009: 67-68). Apart from its declared descriptiveness, the OED also included derogatory labels, such as *low*, *vulgar*, *illiterate* (Béjoint 2010: 103), which clearly express advice to the user which words or senses to avoid. The latest edition of OED published online (3rd edition) no longer has paragraph marks and it no longer makes negative judgements when describing usage, which can be regarded as a

trend toward descriptivism. In Brewer's words (2010: 30): "for the first time, OED is truly living up to its claim to be an historian of the language rather than a critic of it."

Johnson's *Dictionary* also had a significant influence on American lexicography, since Johnson was considered the seminal authority on language for Americans in the second half of the 18th century and American lexicography was colored by his fame (Hitchings 2005: 224). Joseph E. Worcester wrote in the *Introduction* to his dictionary, *A Universal and Critical Dictionary of the English Language*, published in 1846 that Johnson's *Dictionary* "from the time of its first publication, has been, far more than any other, regarded as a standard for language." The *Dictionary* was revised by Henry John Todd and came to be known as Todd-Johnson. Since the first dictionaries used in America were produced in Britain and only very few small dictionaries were published in America at the end of the 18th century, the true beginning of American lexicography was the publication of Noah Webster's *American Dictionary of the English Language* in 1828, which was compiled with the intention of replacing Todd-Johnson in America as the standard work (Landau 2001: 66; Béjoint 2010: 84-85). Unlike Johnson, Webster did not want to "fix" the language but welcomed change (Landau 2001: 69). However, Webster also had strong opinions on what was proper and what was not and he also had significant influence on American spelling (Schulman and Lepore 2008: 57; Landau 2001: 70-71). In 1890, *Webster's International Dictionary*, called *The International*, was published with the aim of being an authority on all varieties of English.

Apart from Webster, Worcester was another lexicographer whose dictionaries dominated the American market in the mid-19th century, his most important dictionary being *A Dictionary of the English Language* (1860). If we compare Webster's and Worcester's attitude toward descriptivism and prescriptivism in lexicography, we can claim that Webster tried to standardize the language and was essentially prescriptive, although he later moved away from prescriptivism realizing how hopeless the task was (Laird 1970: 263 ff.). Worcester, on the other hand, believed that it was not the function of the dictionary to standardize any part of the language; in his opinion, the function of the dictionary is to record constant changes in the language (Martin 2019).

In the second half of the 19th century, dictionaries increased in importance among Americans for various reasons: the search for linguistic knowledge grew; industrialization happened and technology advanced, which resulted in new vocabulary; the growth of the population and public education as a means of self-improvement that resulted in a huge demand for books, while also teaching immigrants and others how to speak and write correctly (Adams 2015: 25). Dictionaries began to be perceived as authorities, since their users expected them to provide answers to their questions about usage, pronunciation, etymology, and the like, or in other words, their users expected them to be prescriptive (Landau 2001: 85; Béjoint 2010: 90-91).

The revised edition of *Webster's International Dictionary*, entitled *Webster's*

New International Dictionary was published in 1909. The second edition of *Webster's New International Dictionary* (1934) was the ultimate authority on meaning and usage and it dominated the market for unabridged dictionaries until the 1960s (Morton 1994: 2; 39). In 1961, the third edition of *Webster's New International Dictionary* came out, which was descriptive, reflecting the view of its editor Philip B. Gove (1967: 7), who pointed out that "Lexicography should have no traffic with guesswork, prejudice, or bias, or with artificial notions of correctness and superiority. It must be descriptive and not prescriptive." The descriptive approach of the dictionary was not well accepted by those who believe that a dictionary should serve as a standard of correctness, thus being prescriptive (Morton 1994: 7; Béjoint 2010: 134). In 1966, *The Random House Dictionary* was published which was intended to be fully descriptive and saw a revised edition in 1987 (Morton 1994: 284). The publication of the *American Heritage Dictionary of the English Language* in 1969 should be understood as a reaction against the descriptive approach of *Webster's Third* (Adams 2015: 17), being self-described as more prescriptive than its competitors (Morton 1994: 285). In the *Introduction to the American Heritage Dictionary*, the editor-in-chief, William Morris states that the dictionary "would faithfully record our language, the duty of any lexicographer", but he adds that the dictionary "would add the essential dimension of guidance, that sensible guidance toward grace and precision of the American public" (Morris 1969: vi). However, the first edition also included elements that pointed towards a more descriptive direction. Successive editions (1982, 1992, 2000 and 2011) qualified the dictionary's prescriptive tendencies, yet, in the course of time, the dictionary changed in ways that reflect the descriptive approach (Adams 2015: 17-18; 25; 32). Despite a growing trend toward description, all five editions of the *American Heritage Dictionary* are still regarded as normative (Adams 2015: 41).

It can be said that the prominence of linguistic prescriptivism in English was set at the level considerably lower than in numerous other European languages very early on. One can trace this attitude to the following statement in Samuel Johnson's *Dictionary*:

[...] academies have been instituted, to guard the avenues of their languages, to retain fugitives, and repulse intruders; but their vigilance and activity have hitherto been vain; sounds are too volatile and subtile for legal restraints; to enchain syllables, and to lash the wind, are equally the undertakings of pride, unwilling to measure its desires by its strength. The French language has visibly changed under the inspection of the academy; the stile of Amelot's translation of Father Paul is observed, by Le Courayer to be un peu passé; and no Italian will maintain that the diction of any modern writer is not perceptibly different from that of Boccace, Machiavel, or Caro. (Johnson 1755: Preface: 10)

As we could see from the above overview, prescriptivism is present in the cultures and linguistic traditions of the English-speaking world in the Northern Hemisphere. The authority for prescriptive usage in English is not very clear, the reason being the absence of normatively inclined institutions like language

academies, and there are various and changing referents like the *Oxford English Dictionary*, the *American Heritage Dictionary*, and the like (Balteiro 2011: 285). British dictionaries have been more descriptive than prescriptive, which is due to the influence of Johnson's *Dictionary* and OED (Béjoint 2010: 154). America might be expected to be particularly open to lexicographic description, but surprisingly, Americans have long embraced prescriptive attitude toward language. Prescriptivism is no longer a hot issue in America and the late twentieth century American lexicography is characterized by the descriptive approach (Béjoint 2010: 155; Adams 2015: 18–20).

2.3 Prescriptivism and descriptivism in the lexicographic tradition of the Slovene language

As stated in Section 2.1, Slavonic languages are characterized by a reference book trio: a dictionary, along with a grammar book and a manual of orthography. A look into the history of linguistic development in Slovenia shows that the first reference books go back to the 16th century, to the works of the Protestant writers who were strongly influenced by the German Protestant Reformation movement. The most notable Slovene Protestant Primož Trubar is regarded as the key figure of Slovene cultural history who is also notable for consolidating the Slovene language. The first proper dictionaries were published at the end of the 16th century by the German lexicographer named Hieronymus Megiser. These dictionaries were multilingual, Slovene being one among many target languages. The multilingual dictionaries formed a basis for the development of bilingual dictionaries that followed, whereas monolingual lexicography in Slovenia developed relatively late.

The first Slovene monolingual dictionary was compiled by Joža Glonar and came out as late as 1936. Glonar adopted a more prescriptive approach as he tried to follow linguistic rules when describing the language (Dobrovoljc 2004: 65–66). The compilation of the first institutional monolingual Slovene dictionary began in the 1950s. The Dictionary of Standard Slovene (DSS) appeared in five volumes and the final, fifth, volume came out in 1991. The second edition with minor changes and corrections was published in 2014. These two editions are characterized by strict adherence to normativity and prescriptivism. In 2016, work on an entirely new dictionary of standard Slovene began (*eSSKJ: Slovar slovenskega knjižnega jezika — e-Dictionary of Standard Slovene*). The first and the second editions of the DSS claim to include a variety of variant forms and language peculiarities. The lexicographers admit openly that their approach is normative, but at the same time they strive towards objectivity to the greatest possible degree. The compilers of the eDSS, however, moved away from prescriptivism to language description, but partly they retain the normative role typical of manuals of orthography. For more information on the inclusion of normative information in the DSS and eDSS, see Šipka et al. 2020.

Compared to monolingual dictionaries, grammar books of Slovene can be traced back to the end of the 16th century. The first grammar book was published in 1584 by Adam Bohorič and was didactically oriented (meant to serve moral teaching as well). Numerous other grammar books were published in the centuries that followed, some didactic, some didactic with normative tendencies and others with pure normative orientation. The most influential grammar book was published in 1976 by Jože Toporišič. It is considered the highest authority in the field of grammar and is currently the single most quoted grammar book in Slovenia (Marušič and Žaucer 2015).

The reference book that has been of the utmost importance in setting the norm for the Slovene standard language has been the manual of orthography. Many different manuals of Slovene orthography typically consist of two parts: the first part contains theoretical rules relating to orthography, pronunciation, inflectional paradigms, syntactic rules and style; the second part is the dictionary part that exhibits the rules explained in the first part. The first manual of orthography was published in the second half of the 19th century by Fran Levec and was didactic in nature; this is why it was used as a school textbook (Dobrovoljc 2015a). The 20th century saw the publication of numerous other manuals of orthography that were first published by individual scholars and were characterized by a strong normative note. The first manual of orthography to be published under the auspices of an institution (the Orthographic Committee of the Scholarly Society in Ljubljana) was authored by Anton Breznik and Fran Ramovš. It came out in 1935, and it is descriptive in nature, meaning that it allows the users to make a choice about the variants included in the manual (Dobrovoljc 2015b). In 2001, a completely new *Manual of Slovene Orthography* was published. It was edited by the grammarian Jože Toporišič and others. This *Manual* is based on Toporišič's (1976) grammar book and follows the tradition of normative manuals of orthography (Dobrovoljc 2015c).

Bearing all these historical developments in mind, the study into the effect of normative labels in contemporary dictionaries on their users was conducted. The findings of this study are presented and discussed in the following sections.

3. Research design

The study was carried out with Slovene and American students who are likely to use monolingual dictionaries in general (e.g., in writing their assignments). We wanted to get an insight into how native speakers of Slovene and American English interpret select normative labels in monolingual general dictionaries. For this purpose, we prepared a questionnaire, which was distributed in early November 2019 to the undergraduate students of economics and English at the University of Ljubljana, Slovenia. The US survey was conducted in late January 2020 with the students of English and International Letters and Cultures at Arizona State University. A total of 105 students provided their responses to all

questions in the Slovene survey, while 112 participants provided their responses to all questions in the US survey³.

The questionnaire was developed with the aim of eliciting monolingual dictionary users' attitudes toward normative labels and the effects the labels have on dictionary users. The Slovene questionnaire is provided in Appendix 1. The US questionnaire can be found in Appendix 2. The questionnaire consists of three parts based on the type of information we wanted to obtain. The first part (Question 1) includes the labels *colloquial*, *dialectal*, *incorrect*, and *obsolete* belonging to primary exclusion labels. The purpose of primary exclusion labels is to exclude a word or its meaning from the contemporary standard language. The four primary exclusion labels are mixed with four secondary exclusion labels (*expressive*, *facetious*, *ironic*, *vulgar*) whose primary purpose differs from that of primary exclusion labels. Secondary exclusion labels can still be interpreted as excluding a lemma or one of its senses from contemporary standard language. Our respondents were asked how often a lemma or one of its senses marked with one of the aforementioned labels is used in the contemporary standard language (always, often, sometimes, rarely, never). Question 1 aims at discerning how our respondents interpret these labels as such, i.e., in isolation from the words that are labeled. The second part, which comprises of three questions, inquiries about whether these labels are useful (Question 2), whether respondents pay attention to them (Question 3), and whether a more precise division within the field of colloquialism would be useful (Question 4). The purpose of these questions was to gather information about users' general attitudes toward this type of labels. The final part of the questionnaire (Question 5) asks if the selected words are used in the standard language (using a Likert scale: strongly agree, agree, no opinion, disagree, strongly disagree). There are four pairs of words. In each pair, the words share a common characteristic, e.g., both words in a pair are colloquial but just one of them is labeled *colloquial*, whereas the other one does not have any label. The purpose of Question 5 was to see whether the labels have any effect on dictionary users.

4. Results

As indicated in Section 1, the present research was meant to provide answers to the following questions, comparing Slovene and US data:

- a. How does a type of normative labels influence the level of acceptability of labeled words in the standard language variety?
- b. What are the users' attitudes toward normative dictionary labels?
- c. How does the fact that a lemma or one of its senses marked with a normative label affect the level of acceptability of labeled words in the standard language variety?

The results relating from these three questions will be discussed in turn below.

4.1 Primary vs secondary exclusion labels

For both, the Slovene and the US surveys, an independent-sample t-test was conducted to compare the levels of the standard language usage of primary and secondary exclusion labels (Question 1). In the Slovene survey, there was no significant difference in the level of standard language use for primary labels (M=2.60, SD=1.21) and secondary levels (M=2.75, SD=1.12); $t(838) = -1.84$, $p = .067$. As can be seen, primary exclusion labels have a somewhat lower degree of standard language use, but the difference is not statistically significant, although it is very close to that status. In the US survey, however, there was a significant difference in the level of standard language use for primary exclusion labels (M=2.97, SD=1.19) and secondary exclusion labels (M=3.26, SD=1.21); $t(894) = -3.65$, $p = .000$. As can be seen, primary exclusion labels have a lower degree of standard language use. The standard language use of individual labels is presented in Table 1.

Table 1: Standard language use of individual labels

	Slovene			US		
	Type	N	Mean	N	Mean	Std. Deviation
colloquial	primary	105	3.03	112	3.34	1.03
dialectal	primary	105	2.39	112	3.19	.87
incorrect	primary	105	2.37	112	2.65	1.45
obsolete	primary	105	2.61	112	2.70	1.20
expressive	secondary	105	3.12	112	3.85	.95
facetious	secondary	105	2.79	112	2.92	1.03
ironic	secondary	105	3.10	112	3.45	1.11
vulgar	secondary	105	1.97	112	2.83	1.42

In the Slovene study, the results show that the main contributor to the exclusion among secondary labels is the label *vulgar*. Also, given that Slovene dic-

tionaries contain both *colloquial* and *lower colloquial*, the primary label *colloquial* is considerably less exclusive than other primary labels. This points to the fact that the difference between primary and secondary colloquial labels could be even higher with other labels in lieu of the aforementioned two. The US data show that the main contributors to the exclusion among secondary labels are the labels *vulgar* and *facetious*. Among primary labels, *colloquial* and *dialectal* are considerably less exclusive than other primary labels.

What is most remarkable in comparison of the two datasets is a higher level of acceptance for the use of labeled dictionary entries in the standard language variety in US respondents across the board. The average value for acceptance is 2.67 for Slovene respondents and 3.12 for US respondents.

4.2 Usefulness and use of labels

The data obtained from the responses to Questions 2, 3 and 4 (Table 2) show that there is a difference between the cognitive component of the attitude about usage labels (there is a relatively high acceptance for the idea that they are generally useful) and the behavioral component of that attitude (the respondents use them at a lower degree). One can also see that the belief that the division inside the label *colloquial* (*standard colloquial* vs *non-standard colloquial*) is useful only slightly above the neutral point (with three being the mid-point of the 1–5 scale).

Table 2: Usefulness and use of labels

	Slovene			US		
	Useful	I use them	Higher vs lower colloquial useful	Useful	I use them	Higher vs lower colloquial useful
N	105	105	105	112	112	112
Mean	4.10	3.22	3.24	3.92	3.12	3.24
SD	.843	1.18	.90	.818	3.38	.942

The Slovene and US data are remarkably similar, the only difference being a higher positive attitude of the Slovene respondents about the usefulness of the labels.

4.3 Labeled vs unlabeled entries

For both groups of respondents, an independent-samples t-test was conducted to compare the level of acceptance for the use in the standard language variety in labeled and unlabeled entries (Question 5). In the Slovene dataset, there was a significant difference in the level of acceptability for labeled entries ($M=2.58$, $SD=1.28$) and non-labeled entries ($M=2.94$, $S=1.30$); $t(838) = -4.05$, $p=.000$. As can be seen, labeled entries have a lower degree of acceptability. The acceptability of individual entries is presented in Table 3.

Table 3: Acceptability of individual entries in the Slovene study

	1ybaksis	1nsekirati	2ydeci	2ndeka	3ydeka	3nhohštapler	4ncoprnik	4yfajfa
N	105	105	105	105	105	105	105	1.05
Mean	1.77	3.30	3.27	3.51	2.92	1.88	3.06	2.34
SD	.80	1.38	1.37	1.11	1.28	.90	1.11	1.06

Heading codes: 1,2,3,4 – the number of the pair of labeled and unlabeled entries, y – label present, n – label absent. The remainder is the name of the entry.

There is a weak, yet not statistically significant, positive correlation between the self-assessment of using the labels and the effect they have (the label use question versus the difference in acceptability between labeled and unlabeled entries [$R = .12$, $p = .22$, $N=105$]).

In the US dataset, there was a significant difference in the level of acceptability for labeled entries ($M=3.04$, $SD=1.34$) and non-labeled entries ($M=3.67$, $S=1.21$); $t(894) = 7.38$, $p=.00$. As can be seen, labeled entries have a lower degree of acceptability. The acceptability of individual entries is presented in Table 4.

Table 4: Acceptability of individual entries in the US study

	1ncram	1ychill_out	2yscrewed	2ybamboozle	3yprick	3nhang_out	4nknocked-up	4njerk
N	112	112	112	112	112	112	112	112
Mean	4.04	3.35	3.11	3.03	2.69	4.08	3.04	3.53
SD	.95	1.33	1.31	1.32	1.34	1	1.29	1.27

Heading codes: 1,2,3,4 – the number of the pair of labeled and unlabeled entries, y – label present, n – label absent. The remainder is the name of the entry.

There is a weak, yet not statistically significant, positive correlation between the self-assessment of using the labels and the effect they have (the label use

question versus the difference in acceptability between labeled and unlabeled entries [$R = .18$, $p = .063$]).

What is most remarkable in a comparison between the Slovene and US data is a considerably lower level of acceptance among Slovene students for labeled and unlabeled entries alike. On a scale from 1 (least acceptable) to 5 (most acceptable), the average among Slovene students was 2.76 (2.58 for unlabeled entries and 2.94 for labeled entries) as opposed to the score obtained from the US students of 3.36 (3.04 for unlabeled entries and 3.67 for labeled entries).

5. Discussion and conclusion

Dictionaries are often thought of as a record of language. However, dictionaries — even if they claim to be only descriptive or only prescriptive — mostly combine both descriptive and prescriptive approaches. This is pointed out by Svensén (2009: 24), who says that "... most general-purpose dictionaries should be characterized as partly normative because, in one way or other, a large portion of what they describe is actually made up of linguistic norms already in existence"; this is also in line with Cameron (1995: 5), who considers this absolute binary distinction between descriptivism and prescriptivism a mistake. An important point that should be emphasized is that in the dictionary, the public often seeks answers to questions about the spelling, pronunciation and the correct usage of words, which means that the public expects at least a certain degree of normativity (Adams 2015: 25). Compilers of a general-purpose monolingual dictionary should therefore always try to find a balance between descriptive and prescriptive approaches, since both of them affect the users (Adams 2015: 43).

The results obtained in the present research point to the following hypotheses that are consistent with the aforementioned claim about the relation between descriptivism and prescriptivism.

First, in both observed samples, the type of normative labels matters. Primary normative labels are associated with a lower acceptance for the use in the standard language variety than their secondary counterparts. In this survey, primary labels showed a lower level of acceptability (2.60 in the Slovene and 2.97 in the US survey) than their secondary counterparts (2.75 in the Slovene and 3.26 in the US survey). The fact that their effect seems to be stronger in the Slovene sample could be related to a higher degree of prescriptivism in the mainstream Slovene linguistic culture as opposed to a more descriptive approach in the American linguistic culture.

Second, in both samples, there is a disconnect between the cognitive component of attitude toward normative labels and its behavioral component. The approval for the claim that these labels are useful (4.10 in the Slovene and 3.92 in the US survey) is considerably higher than the value for the statement that the respondent uses the labels (3.22 in the Slovene and 3.12 in the US survey). A higher approval for the claim about usefulness of the labels in the Slovene

sample may be another consequence of a higher level of prescriptivism in that linguistic culture.

Third, the presence of the label makes a difference in terms of the level of acceptance for standard language use in both observed samples. Consequently, labeled entries have a lower level of acceptability (2.58 in the Slovene and 3.04 in the US survey) than their unlabeled counterparts (2.94 in the Slovene and 3.67 in the US survey). A considerably lower general level of acceptance for standard language use among Slovene students (2.76 versus 3.36 in their US counterparts) may, yet again, be related to a higher prominence of prescriptivism in that linguistic culture.

Among the limitations of the study whose results are presented in this paper, we should mention that the study was carried out in the environment of only two languages, i.e., Slovene and American English. Apart from that, the number of respondents and their background (they were all students) was also part of a fairly limited sample, which means that the two gathered were not representative. Consequently, the results should be understood as hypotheses which should be examined in more detail in further studies.

Acknowledgements

The authors acknowledge the project, The Practice of Normative Labeling in Slovene and American Lexicography (grant number: BI-US/18-20-036), which was financially supported by the Slovene Research Agency. They also acknowledge the approval of the Arizona State University (ASU) Institutional Review Board. Our gratitude goes to Elly van Gelderen, Mariana Bahtchevanova, and Anne Walton-Ramirez for their help with organizing the survey of US students. The authors are also very grateful to the study participants from the undergraduate programs in the School of Economics and Business and the Faculty of Arts at the University of Ljubljana as well as to the students of English and International Letters and Cultures at Arizona State University.

Danko Šipka would also like to thank the ASU Melikian Center and the National Council for Eurasian and East European Research for their financial support of his trip to Slovenia.

Endnotes

1. In discussing the complexity of the relation between standard and non-standard language forms, a study carried out by Smakman (2012) is of a particular importance. He surveyed over one thousand non-linguists from England, Flanders (a part of Belgium, France and the Netherlands), Japan, as well as other parts of the Netherlands, New Zealand, Poland and the United States and concluded that although the ideas about the standard language have very little in common, one can recognize a socially distinctive (exclusive) notion of the standard language and socially cohesive (integrative) one. The former notion points to the fact that the language form is only used in formal communication (such as the news, official documents, etc.).

2. This is very much in line with Scherba (1974: 276; cited in Grigoryan 2007: 4–5), who stated: "From a purely linguistic point of view only normative or academic type of dictionaries must be considered scholarly since their subject matter is linguistic reality — i.e. the lexical system of the language in its totality and complexity."
3. IRB approval was secured at Arizona State University in early April 2019 (IRB ID: STUDY00010014).

References

- Adams, Michael.** 2015. Language Ideologies and *The American Heritage Dictionary of the English Language*: Evidence from Motive, Structure, and Design. *Dictionaries: Journal of the Dictionary Society of North America* 36: 17-46.
- Ahačič, Kozma (Ed.).** 2015. *Slovenske slovnice in pravopisi: spletišče slovenskih slovnice in pravopisov od 1584 do danes*. Version 1.0. Accessed on 14 May 2020:
<https://fran.si/slovnice-in-pravopisi/>
- Armstrong, Nigel R. and Ian Mackenzie.** 2015. On Prescriptivism and Ideology. *Representaciones. Revista de Estudios sobre Representación en Arte, Ciencia y Filosofía* 11(2): 27-58.
- Balteiro, Isabela.** 2011. Prescriptivism and Descriptivism in the Treatment of Anglicisms in a Series of Bilingual Spanish–English Dictionaries. *International Journal of Lexicography* 24(3): 277-305.
- Béjoint, Henri.** 2010. *The Lexicography of English: From Origins to Present*. Oxford: Oxford University Press.
- Brewer, Charlotte.** 2010. Prescriptivism and Descriptivism in the First, Second and Third Editions of *OED*. *English Today* 26(2): 24-33.
- Cameron, Deborah.** 1995. *Verbal Hygiene*. London/New York: Routledge.
- Dobrovoljc, Helena.** 2004. *Pravopisje na Slovenskem*. Ljubljana: ZRC SAZU.
- Dobrovoljc, Helena.** 2015a. Fran Levec. *Slovenski pravopis*. Ahačič, Kozma (Ed.). 2015. Accessed on 14 May 2020:
<https://fran.si/slovnice-in-pravopisi/34/1899-levec>
- Dobrovoljc, Helena.** 2015b. Anton Breznik, Fran Ramovš. *Slovenski pravopis*. Ahačič, Kozma (Ed.). 2015. Accessed on 14 May 2020:
<https://fran.si/slovnice-in-pravopisi/39/1935-breznik-ramovs>
- Dobrovoljc, Helena.** 2015c. Toporišič, Jože (Ed.). [Pravila:] Franc Jakopin, Janko Moder, Janez Dular, Stane Suhadolnik, Janez Menart, Breda Pogorelec, Kajetan Gantar, Martin Ahlin. [Slovar:] Franc Jakopin, Janko Moder, Janez Dular, Milena Hajnšek-Holz. *Slovenski pravopis*. Ahačič, Kozma (Ed.). 2015. Accessed on 14 May 2020.
<https://fran.si/slovnice-in-pravopisi/50/2001-skupinski-toporisicev>
- Finegan, Edward.** 2003. Linguistic Prescription: Familiar Practices and New Perspectives. *Annual Review of Applied Linguistics* 23: 213-224.
- Finegan, Edward.** 2020. *What Is 'Correct' Language?*
<https://www.linguisticsociety.org/resource/what-correct-language>

- Gliha Komac, Nataša, Nataša Jakop, Janoš Ježovnik, Simona Klemenčič, Domen Krvina, Nina Ledinek, Tanja Mirtič, Andrej Perdih, Špela Petric, Marko Snoj and Andreja Žele (Eds.).** 2015. *Koncept novega razlagalnega slovarja slovenskega knjižnega jezika*. Ljubljana: Založba ZRC, ZRC SAZU. Accessed on 14 February 2020:
https://fran.si/179/novi-slovar-slovenskega-knjiznega-jezika/datoteke/Potrjeni_koncept_NoviSSKJ.pdf
- Gove, Philip B. (Ed.).** 1967. *The Role of the Dictionary*. Indianapolis: Bobbs-Merrill.
- Gove, Philip B.** 1967. The Dictionary's Function. Gove, Philip B. (Ed.). 1967: 5-8.
- Grigoryan, Ashot A.** 2007. On Gender Aspect in Lexicography. Karpova, Olga and Faina Kartashkova (Eds.). 2007: 2-10.
- Hausmann, Franz Josef.** 1989. Die Markierung im allgemeinen einsprachigen Wörterbuch: eine Übersicht. Hausmann, Franz Joseph, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989: 649-657.
- Hausmann, Franz Joseph, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.)** 1989. *Dictionaries. An International Encyclopedia of Lexicography. Vol. 1*. Berlin/New York: Walter de Gruyter.
- Hitchings, Henry.** 2005. *Dr Johnson's Dictionary*. London: John Murray.
- Jarošová, Alexandra and Vladimír Benko.** 2012. The Dictionary of the Contemporary Slovak Language: A Product of Tradition and Innovation. Vatvedt Fjeld, Ruth and Julie Matilde Torjusén (Eds.). 2012: 257-261.
- Johnson, Samuel.** 1755. *A Dictionary of the English Language*. London: W. Strahan, for J. & P. Knapton.
- Karpova, Olga and Faina Kartashkova (Eds.).** 2007. *Essays on Lexicon, Lexicography, Terminography in Russian, American and Other Cultures*. Newcastle: Cambridge Scholars Publishing.
- Laird, Charlton Grant.** 1970. *Language in America*. New York: World Publishing Company.
- Landau, Sidney I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.
- Lynch, Jack (Ed.).** 2003. *Samuel Johnson's Dictionary*. New York: Walker & Co.
- Lynch, Jack.** 2003. Introduction to this Edition. Lynch, Jack (Ed.). 2003: 1-21.
- Martin, Peter.** 2019. Peter Martin on *The Dictionary Wars*. Interview on August 26, 2019. Accessed on 16 April 2020:
<https://press.princeton.edu/ideas/peter-martin-on-the-dictionary-wars>.
- Marušič, Franc and Rok Žaucer.** 2015. Jože Toporišič. *Slovenska slovnica*. Ahačič, Kozma (Ed.). 2015. Accessed on 14 May 2020:
<https://fran.si/slovnice-in-pravopisi/46/1976-toporisc>
- Miłkowski, Marcin.** 2013. On the Social Nature of Linguistic Prescriptions. *Psychology of Language and Communication* 17(2): 175-187.
- Morris, William (Ed.)** 1969. *The American Heritage Dictionary of the English Language*. Boston: Houghton Mifflin.
- Morton, Herbert C.** 1994. *The Story of Webster's Third: Philip Gove's Controversial Dictionary and Its Critics*. Cambridge: Cambridge University Press.
- Scherba, Lev V.** 1974. Opit obschei teorii leksikografii (Experience of General Theory of Lexicography). *Jazykovaja sistema i rechevaja dejatel'nost'* (Language System and Speech Activity): 265-304.
- Schulman, Arthur and Jill Lepore.** 2008. *Websterisms: A Collection of Words and Definitions Set Forth by the Founding Father of American English*. New York: Free Press.

- Smakman, Dick.** 2012. The Definition of the Standard Language: A Survey in Seven Countries. *International Journal of the Sociology of Language* 218: 25-58.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Šipka, Danko, Alenka Vrbinc and Marjeta Vrbinc.** 2020 (in press). Normative Labels in Slovene Dictionaries: Users' and Lexicographers' Perspectives. *International Journal of Lexicography* 33(4).
- Vatvedt Fjeld, Ruth and Julie Matilde Torjusen (Eds.).** 2012. *Proceedings of the 15th EURALEX International Congress, 7-11 August 2012, Oslo*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Vrbinc, Alenka, Donna M.T.Cr. Farina and Marjeta Vrbinc.** 2018. Objectivity, Prescription, Harmlessness, and Drudgery: Reflections of Lexicographers in Slovenia. *Lexikos* 28: 361-383.
- Yong, Heming and Jing Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam: John Benjamins.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: Mouton.

Appendix 1: Slovene Questionnaire

1. The following dictionary label means that the word or its meaning is used in standard language (such as news read by TV anchors, or court decisions). Select one option for each label.

colloquial	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
dialectal	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
expressive	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
facetious	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
incorrect	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
ironic	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
obsolete	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
vulgar	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never

2. Labels (for example, coll – colloquial, vulg – vulgar) are a useful part of the dictionary.

strongly agree agree no opinion disagree strongly disagree

3. When I use dictionaries I pay attention to the labels (for example, coll – colloquial, vulg – vulgar).

strongly agree agree no opinion disagree strongly disagree

4. Is the difference between colloquial and lower colloquial useful?

strongly agree agree no opinion disagree strongly disagree

5. The word in bold are used in standard language

báksiš -a m pog. *napitnina, podkupnina*

strongly agree agree no opinion disagree strongly disagree

čóprnik -a m *čarovnik*

strongly agree agree no opinion disagree strongly disagree

déci m neskl. pog. *deciliter*

strongly agree agree no opinion disagree strongly disagree

fájfa -e ž pog. *priprava za kajenje, ki se napolni s tobakom; pipa*

strongly agree agree no opinion disagree strongly disagree

sekírati -am nedov. *s pretiranimi zahtevami povzročati komu neprijetnosti, slabo voljo*

strongly agree agree no opinion disagree strongly disagree

déka² -e ž pog. *(volnena) odeja*

strongly agree agree no opinion disagree strongly disagree

hóhštápler -ja m *domišljav človek, gizdalin*

strongly agree agree no opinion disagree strongly disagree

déka¹ -e ž *dekagram*

strongly agree agree no opinion disagree strongly disagree

Appendix 2: US Questionnaire

1. When we look a word up in a dictionary, we sometimes find a label on it like these listed below. Does that mean that we can always use such words in standard English (such as news read by TV anchors, or court decisions) or never, or somewhere in between? Select one option for each label.

colloquial	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
dialectal	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
expressive	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
facetious	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
incorrect	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
ironic	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
obsolete	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never
vulgar	<input type="checkbox"/> always	<input type="checkbox"/> often	<input type="checkbox"/> sometimes	<input type="checkbox"/> rarely	<input type="checkbox"/> never

2. Labels (for example, coll – colloquial, vulg – vulgar) are a useful part of the dictionary.

strongly agree agree no opinion disagree strongly disagree

3. When I use dictionaries I pay attention to the labels (for example, coll – colloquial, vulg – vulgar).

strongly agree agree no opinion disagree strongly disagree

4. It would be useful if dictionaries would differentiate between colloquial and lower colloquial.

strongly agree agree no opinion disagree strongly disagree

5. Is the word in bold used in standard English (such as news read by TV anchors, or court decisions)?

cram v to study for an examination by memorizing facts at the last minute

strongly agree agree no opinion disagree strongly disagree

chill out v *slang* to calm down; to relax

strongly agree agree no opinion disagree strongly disagree

screwed adj *slang* bilked; cheated

strongly agree agree no opinion disagree strongly disagree

bamboozle v *informal* to practice trickery, deception, cozenage, or the like

strongly agree agree no opinion disagree strongly disagree

prick n *slang* an obnoxious or contemptible person

strongly agree agree no opinion disagree strongly disagree

hang out v to loiter in public places

strongly agree agree no opinion disagree strongly disagree

knocked-up adj made pregnant

strongly agree agree no opinion disagree strongly disagree

jerk n a contemptibly naive, fatuous, foolish, or inconsequential person.

strongly agree agree no opinion disagree strongly disagree

Access Routes to BODY PART Multiword Expressions in the 'Big Five' MELDs: Use of Hyperlinks

Sylwia Wojciechowska, *Faculty of English, Adam Mickiewicz University,
Poznań, Poland (swojciechowska@wa.amu.edu.pl)*

Abstract: The treatment of multiword expressions (MWEs) in dictionaries has not received much attention in metalexigraphy, although the significant role of phraseology has been stressed since the advent of corpus linguistics. The paper aims to analyse the lexicographic representation of semantically related MWEs, containing body part names. The study focuses on access routes to these MWEs in the 'Big Five' monolingual English learners' dictionaries online (MELDs). It investigates the presence and positions of hyperlinked MWEs on the page of the body part headword in order to find out if they depend on a given MWE or are dictionary-specific. Double or multiple hyperlinks to the same MWE are frequently found within a single body part entry, and the variety of access routes is evaluated with a view to offering a more homogeneous presentation of hyperlinked related MWEs.

Keywords: MONOLINGUAL ENGLISH LEARNERS' DICTIONARIES, MULTIWORD EXPRESSIONS, METONYMY, COGNITIVE LEXICOGRAPHY, ACCESS ROUTES, ONLINE DICTIONARIES

Opsomming: Toegangsroetes na meerwoordige LIGGAAMSDEEL-uitdrukings in die 'Groot Vyf' MELDs: Die gebruik van hiperskakels. Alhoewel die beduidende rol van fraseologie sedert die ontstaan van die korpuslinguistiek beklemtoon is, het die hantering van meerwoordige uitdrukings (MWU's) in woordeboeke nog nie baie aandag in die metaleksikografie geniet nie. Hierdie artikel het die analise van die leksikografiese voorstelling van semanties verwante MWU's, wat die name van liggaamsdele bevat, ten doel. Die navorsing fokus op toegangsroetes na hierdie MWU's in die "Groot Vyf" aanlyn eentalige Engelse aanleerderswoordeboeke (MELDs). Die voorkoms en ligging van aaneengeskakelde MWU's op die bladsy waarop die liggaamsdeel-trefwoord voorkom, word ondersoek om sodoende te bepaal of hulle afhanklik is van 'n gegewe MWU en of hulle woordeboekspesifiek is. Dubbel- of veelvoudige hiperskakels van dieselfde MWU kom dikwels binne 'n enkele liggaamsdeel-inskrywing voor, en die verskillende toegangsroetes word met die oog op die aanbod van 'n meer homogene voorstelling van aaneengeskakelde verwante MWU's geëvalueer.

Sleutelwoorde: EENTALIGE ENGELSE AANLEERDERSWOORDEBOEKE, MEERWOORDIGE UITDRUKKINGS, METONOMIE, KOGNITIEWE LEKSIKOGRAFIE, TOEGANGSROETES, AANLYN WOORDEBOEKE

1. Background

Already in the first half of the 20th century, the authors of the first English pedagogical dictionaries, Harold Palmer and A.S. Hornby, acknowledged the role of phraseology in learning a foreign language (Palmer 1933). However, the importance of phraseology began to be widely recognised with the emergence of corpus linguistics and the seminal works of John Sinclair (1987, 1991). Since that time (meta)lexicographers have stressed the role of multiword expressions (henceforth MWEs) in determining meaning, for example, Cowie (1998), Moon (1998) and Hanks (2013). This, however, has not translated into studies of the representation of MWEs in dictionaries, and the status of MWEs in lexicography still remains unsettled.

One reason behind neglecting this issue in metalexicography is lack of consensus concerning MWEs terminology. While proposing her own typology of fixed expressions and idioms, Moon (1998: 19-20) observes that there is no generally agreed set of categories in the literature subsumed under MWEs, and clear classifications are impossible. Collocations, idioms, phrases, similes, metaphors and sayings overlap, and it is often hard to assign a MWE to a single category. Another reason is 'the privileged status that the (orthographic) word has traditionally enjoyed in lexicography' (Lew 2012: 349). This is not surprising due to the variable lexical and structural fixedness of MWEs as shown by corpus-based research (e.g. Moon 1998, Fellbaum 2007, Stathi 2007). The unstable nature of MWEs would be an obstacle if they were to serve as the building blocks of the dictionary macrostructure. Moreover, as Lew (2012) notes, dictionary users tend to look up a single word, either not realising that they are dealing with a MWE or merely because this is, or rather 'was' thanks to search interface enhancement in online dictionaries, a familiar strategy.¹ Advanced dictionary users will usually recognise an item as a MWE, but they may not be sure as to the scope of a MWE. However, the fact that their search strings will vary from the canonical form included in the dictionary does not constitute a problem any longer owing to the partial matching functionality in many online dictionaries.

The result of all this is inconsistent treatment of MWEs in many printed dictionaries, as observed by Oppentocht and Schutz (2003: 218): '[t]hese can often be found under more than one entry, in different forms, and even with different explanations'. At the same time, the authors envisage that such anomalies will not occur in electronic dictionaries due to the use of proper computational tools. The current study does not support this claim, as it shows duplication or multiplication of information and cases of inconsistent treatment concerning MWEs in present-day monolingual English learners' dictionaries (henceforth MELDs) online.

2. Study

2.1 Aims

The present study deals with the lexicographic representation of MWEs in

MELDs online, focusing on access routes. It needs to be stressed that the analysis is not concerned with the *search route* or *path* a user follows to look up the target item, but with the *access structure* that determines the search route.² The study aims to examine access routes to semantically related MWEs, containing body part names.

The following research questions are posed:

- (1) Are idioms defined within entries for certain body part names or in separate hyperlinked entries?
- (2) Where are the definitions of MWEs or hyperlinks to them located on the page of the body part headword?
- (3) Is the use of hyperlinks dictionary-specific or item-specific?
- (4) Do the dictionaries offer a consistent approach?

It is assumed that taking into account an extensive use of hyperlinks in electronic lexicography, dictionaries will exploit this lexicographic device as a way of cross-referencing, and hence facilitate access to related lexical items.

2.2 Data and materials

MWEs with body parts names have been chosen for the present study because they constitute a fairly homogenous group, related by means of metonymic motivation, and are therefore expected to be treated in a fairly consistent way. In fact, these expressions often exhibit a mixed motivation where metonymy is intertwined with metaphor, taking the form of 'metaphtonymy' (Goossens 1990).³ In cognitive linguistic literature the notion of idiom motivation is widely acknowledged (e.g. Gibbs 1993, Geeraerts 2002, Dobrovolskij and Piirainen 2005, Langlotz 2006). Idioms are no longer viewed as *long words*, but as constructions that have their internal syntactic and semantic structure. Metonymic motivation can be illustrated by *be all ears* in the current study, where the part of the body stands for the action of listening (BODY PART FOR ACTION metonymy). The interplay of metaphor and metonymy is represented in the examined sample by, for example, *bite the hand that feeds you*. The decomposable nature of this idiom is reflected by two underlying metaphors and one metonymy: harming is understood in terms of attacking someone physically (HARMING IS ATTACKING metaphor), helping in terms of feeding (HELPING IS FEEDING metaphor), and the body part stands for a person (BODY PART FOR PERSON metonymy).

The study analyses access routes to MWEs with three body part names in their nominal forms: *ear*, *hand* and *shoulder* in the 'Big Five' MELDs online: *Cambridge Advanced Learner's Dictionary* (CALD), *COBUILD Advanced English Dictionary* (COBUILD), *Longman Dictionary of Contemporary English* (LDOCE), *Macmillan English Dictionary Online* (MEDO) and *Oxford Advanced Learner's Dictionary* (OALD).⁴ It needs to be added that dictionaries aggregated on the page of a certain dictionary are also taken into account provided they are aimed at advanced learners of English and are not American English versions.⁵ Thus, hyperlinks to

the following dictionaries are included in the study: in CALD, to *Cambridge Academic Content Dictionary* (CACD), and *Cambridge Business English Dictionary* (CBED); and in COBUILD, to *Easy Learning Idioms Dictionary* (ELID). The other three dictionaries do not contain hyperlinks to aggregated dictionaries.⁶

The study has provided representative results, so there has been no need to extend the sample by incorporating MWEs with other body part names. While selecting items for analysis, the term MWE has been applied in the general sense, namely all types of composite expressions have been chosen, more and less fixed: collocations, idioms, compound nouns, phrases and sayings. The examined material comprises all *ear*, *hand* and *shoulder* MWEs covered in the 'Big Five', but is then reduced to include only the ones present in all the five MELDs. MWEs have been excluded from the sample if body part names appear in attributive position, for example, *hand luggage*; if MWEs are based only on metaphor, not metonymy, for example, *minute hand*; and if they have a literal meaning, for example, *shrug your shoulders*, *have your ears pierced*. Slight differences in canonical forms are not relevant for the study, for example, *make/lose money hand over fist* vs. the truncated form *hand over fist*, or *be caught with your fingers/hand in the till* vs. *have your hands/fingers in the till*. These forms are treated as variants of the same MWE. The total number of body part MWEs analysed and tagged is 106 in CALD, 103 in COBUILD, 99 in LDOCE, 107 in MEDO, and 142 in OALD.

2.3 Procedure

The analysis proceeds in a few steps and the relevant data are entered into separate Excel files created for each of the five MELDs. First, *ear*, *hand* and *shoulder* MWEs are tagged as being defined on the page of the body part entry, or hyperlinked from it to a separate entry. Next, if the former is the case, it is checked whether a given MWEs is presented under a separate sense or covered at the bottom of the entry in a dedicated idioms or phrases box or section. If the latter is the status quo, the position of the hyperlink is examined: as a separate sense, in the idioms or phrases box or section, or in a separate panel with more results. The last of these options, a search panel is a solution adopted in current online dictionaries to help the user with potential target items including the search word or string. This is a form of step-wise access available when the search item matches more than a single lemma or is part of a MWE (for further discussion see Lew 2012). In CALD such a search panel is named *More meanings of*, henceforth referred to as MM; in COBUILD, *Related terms of* (RT); in LDOCE, *More results* (MR); in MEDO, *Other entries for* (OE)⁷; and in OALD, *Other results* (OR). Finally, it is investigated how consistent the dictionaries are in terms of access routes to body part MWEs, and how similar their approaches are when compared to one another.

While analysing access routes to body part MWEs, attention is also paid to reflecting the semantic relations which hold between some expressions. Evalua-

tion is carried out from the perspective of cognitive linguistics. Adopting the cognitive approach in this paper is supported by the fact that increasing numbers of (meta)lexicographers and cognitive linguists recognise the advantages of its application in lexicography, for example, Van der Meer (1999), Moon (2004), Geeraerts (2007), Adamska-Salaciak (2008), Atkins and Rundell (2008), Kövecses and Csábi (2014) and Ostermann (2015), the last of whom in fact coined the term cognitive lexicography. Moreover, the cognitive perspective is also in line with the current tendency in pedagogical lexicography towards explicating figurative language usage. Boers and Lindstromberg (2006), for instance, prove in their studies that new lexical items, including idiomatic expressions, are learned and recalled more effectively when their motivation is explained.

3. Results

3.1 Defined or hyperlinked

The results of the study demonstrate that in four out of five dictionaries under scrutiny, body part MWEs are hyperlinked in the overwhelming majority of cases rather than defined in the entry for the body part name. The proportions of hyperlinked MWEs are equal or close to equal in CALD, COBUILD, LDOCE and MEDO: 95%, 92%, 95% and 92% respectively. OALD turns out to offer a completely opposite approach with merely 8% hyperlinked body part MWEs (see Figure 1). In this dictionary, the prevailing treatment is defining MWEs in an alphabetically arranged *Idioms* box at the bottom of the entry: 128 MWEs out of 142 under examination (90%). In LDOCE, a similar box is found, with collocations instead of idioms. Definitions are also provided without recourse to hyperlinks. However, all the expressions included in these *Collocations* boxes have literal meanings, and as such are not taken into account in the study, for example, *wave your hand* and *hunch your shoulders*.

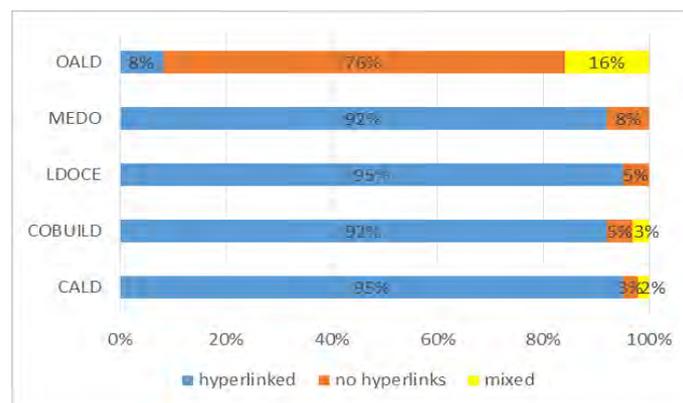


Figure 1: The proportions of hyperlinked MWEs in the 'Big Five'

What is surprising from a semantic perspective, five body part MWEs are defined twice in the *Idioms* box in OALD. Among them is *win something hands down*, which is broken into separate MWEs, *hands down* and *win (something) hands down*, the former featuring *win hands down* in an example sentence (see Figure 2). As MWEs are alphabetically ordered in the *Idioms* box, the two idioms are further separated from each other by other MWEs which come in between. Such a split is redundant and also creates semantic distance between the two related expressions, which goes against the cognitive approach. The other related MWEs which should not be split in the *Idioms* box in OALD are *in somebody's capable, safe, etc. hands* and *in safe hands/in the safe hands of somebody*; *somebody's ears are burning* and *feel your ears burning*; *off your hands* and *on your hands*; and *(at) first hand* and *(at) second, third, etc. hand*. These pairs are clearly variants of the same MWEs, and the user should see the connection between them.



Figure 2: Double access to *win something hands down* in the *Idioms* box in OALD

As Figure 1 shows, in three dictionaries, CALD, COBUILD and OALD, mixed treatment is found. This happens when there are double or multiple routes of accessing a given MWE from the page of the headword, with and without hyperlinks. Such cases make up a small proportion of the investigated items, the highest observed in OALD, 16%. For instance, in OALD *give somebody the cold shoulder* is defined in the *Idioms* box, and also hyperlinked from the OR right-hand panel, both from the *All matches* and *Idioms* sections.

As far as defining body part MWEs under separate senses is concerned, the numbers for all the dictionaries are very low, and include also MWEs that feature only in example sentences under a specific sense (see Figure 3 for MWEs presented under separate senses). In CALD there are 5 out of 106 such cases; in COBUILD, 8 out of 103; in LDOCE, 5 out of 99; in MEDO, 9 out of 107; and in OALD, 4 out of 142. In two dictionaries, CALD and COBUILD, some of these occurrences exhibit the above-mentioned mixed treatment, combining definitions or examples under separate senses on the page of the headword with hyperlinked MWEs. Such cases may result in breaking semantic links between related items. A case in point is separating the expression *give some-*

body a hand from *lend somebody a hand* in COBUILD, with the former featuring in example sentences under sense 4 ('help' meaning), and the latter detached from it and hyperlinked under sense 36. It is worth noting that in the other four MELDs the two variant forms of the same expression are nested under the same sense.

3.2 Positions of hyperlinked body part MWEs

3.2.1 *Dictionary-specific locations.* As has been shown in 3.1, very few of the analysed MWEs are defined in the body part entry, the majority of them being hyperlinked, with the exception of OALD. This section focuses on the different locations of hyperlinks within the entry in each of the 'Big Five'.

2 [SINGULAR] **the ability to hear and judge sounds**
She has a very good ear for music. ←

T Synonyms and related words

The ear and hearing: *anvil, (as) deaf as a post, audio-...*

Explore Thesaurus

3 [COUNTABLE] **the part at the top of a plant such as wheat that contains the grain**
ears of wheat

T Synonyms and related words

Parts of plants: *aerial root, anther, beanstalk...*

Explore Thesaurus

C From our crowdsourced Open Dictionary

4 **listen with half an ear** to not listen carefully ←
I was listening to the music with half an ear as I repaired the car.
Submitted by Emran Khoshrouye Ghiasi from Iran, Islamic Republic of on 16/12/2018

Figure 3: MWEs with *ear* defined as a separate sense (sense 4) and featuring in an example sentence (sense 2) in MEDO

In CALD, the prevailing location is the MM right-hand panel, listing 95% of the examined MWEs (see Figure 4). Body part MWEs are also hyperlinked from the *Idioms* box at the end of the entry, 38%, and there is one case of a hyperlink from a separate sense. This single MWE is the compound *farmhand* (1%), which is cross-referenced from the sense of 'a worker'. It needs to be added that these percentages do not make up together 100% because the same idiom is sometimes hyperlinked twice or more in different places in the entry. This duplication or multiplication of hyperlinks is also observed in the other dictionaries, and will be discussed in 3.2.3.

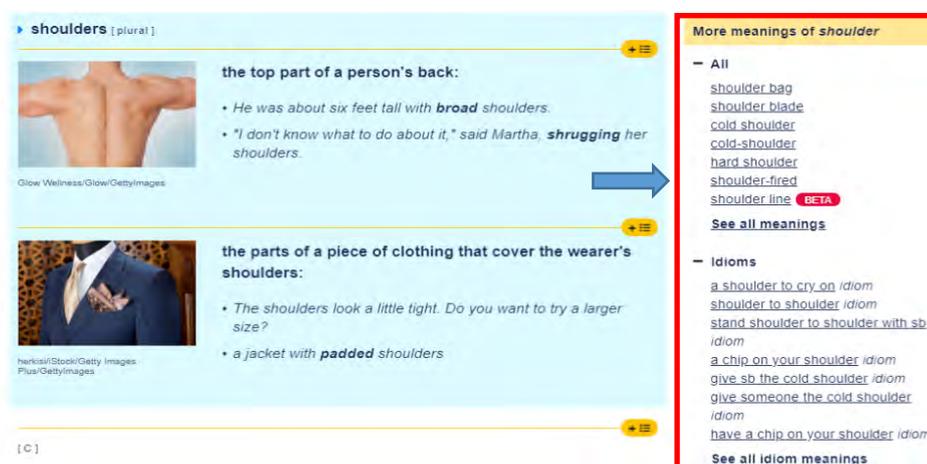


Figure 4: The MM panel on the page of the headword *shoulder* in CALD

COBUILD also assigns hyperlinked body part MWEs to three different locations, two of them being equally frequent: the *More idioms* (MI) section at the bottom of the page after entries from the aggregated dictionaries, and separate numbered senses, 58% and 57% respectively. The third position is also at the bottom of the page in the RT section, 21%.

Separate hyperlinked numbered senses are also the most frequent in LDOCE, constituting 65% of all the examined MWEs. The other two positions are: the MR right-hand panel and hyperlinks grouped at the bottom of the entry before the *Collocations* box, 36% and 18% respectively.

MEDO takes a completely different strategy, and does not place hyperlinks under separate numbered senses except one occurrence, *show your hand*, hyperlinked from the sense of 'advantages in a situation'. Instead, the dominant location for hyperlinked MWEs in MEDO is the OE right-hand panel, 65%, followed by the *Phrases* section at the end of the entry, 47%. There are also sin-

gle occurrences of hyperlinked MWEs in the *See also* section at the very bottom of the entry, making up a mere 10%.

As has already been said in 3.1, in OALD the bulk of body part MWEs are not hyperlinked, but defined in the *Idioms* box (90%). Hyperlinks are only used in the OR right-hand panel, constituting 21% of the investigated expressions, and very rarely under separate senses, 3,5% or under the *See also* cross-references⁸ at the bottom of the entry before the *Idioms* box, 1%. It is observed that the last two locations contain hyperlinks to one specific category of MWEs, body part compound nouns, for example, *chargehand*, *farmhand*, *deckhand*, *dab hand*, *glue ear* and *hard shoulder*.⁹

The above findings indicate that the examined dictionaries offer relatively consistent and dictionary-specific access routes to *hand*, *shoulder* and *ear* MWEs in terms of the presence of hyperlinks and their position on the page of the headword. At the same time, some similarities emerge between the dictionaries. Hyperlinks from separate numbered senses predominate in LDOCE and COBUILD. *Idioms* or *Phrases* sections subsuming hyperlinked MWEs are the distinguishing feature of COBUILD and MEDO, not hyperlinked ones, of OALD. It should be pointed out here that in the *Idioms* section in COBUILD (MI), MWEs are hyperlinked to their entries in the aggregated dictionary ELID. Finally, search panels with more results are found in all the dictionaries, and are placed to the right towards the top of the entry, with the exception of COBUILD, in which they are located at the very bottom of the entry. Here, it should be noted that on mobile phones, the search panels appear at the bottom of the screen, after the text of the main entry including *Idioms* or *Collocations* boxes or sections. The only exception is MEDO, with the OE search panel placed at the very top of the entry. Search panels constitute the preferred location for hyperlinked body part MWEs in CALD and MEDO.

Despite the lack of a clear typology of MWEs in literature, some classification is attempted in the examined MELDs, at least to distinguish idioms from other MWEs, as there is an *Idioms* section or box in CALD and OALD. In MEDO it is called *Phrases* section, and in COBUILD, *More idioms* (MI), suggesting that it only adds to the list of hyperlinked MWEs placed under separate numbered senses. Moreover, in some of the dictionaries, search panels with more results are divided into *All matches* and *Idioms* sections. In both CALD and OALD, the latter selects only idiomatic expressions from the *All matches* section. In LDOCE and COBUILD no distinction is made between idioms and other expressions in the search panels, whereas in MEDO, although the panel is not divided either, idioms are distinguished from the other lexemes such as compound nouns (e.g. *old hand*, *ear candy*) by the label PHRASE assigned to them.

It is, nonetheless, hard to explain the reasoning behind establishing the *See also* section for just a few hyperlinked MWEs, as is the case in LDOCE, MEDO and OALD. In particular, it is puzzling in MEDO, where it is located at the very bottom of the entry, after the *Phrases* section, and contains hyperlinks to expressions that could well be added the *Phrases* section, and are already in-

cluded in the OE search panel, for example, *cold shoulder* or the ones that should be added to either or both of the above-mentioned sections, for example, *all hands on deck*. In LDOCE such additional hyperlinks are grouped after the numbered senses, although they could be added to the list of the numbered hyperlinked senses, the only downside being that it might get too long then. Finally, in OALD, the presence of the *See also* hyperlinks is perhaps more justifiable, because, as has already been mentioned, they are cross-references to one category of MWEs, body part compound nouns.

3.2.2 *The hyperlinked pages.* In the examined MELDs except OALD, the cross-referenced body part MWE tends to be the headword of the hyperlinked entry. In OALD, 23 out of 37 hyperlinked MWEs take the user from the OR panel to the *Idioms* box under a different headword, another content word from the target MWE. In all the cases, the page scrolls down to the searched MWE. For instance, for *bite the hand that feeds you* the hyperlinked page is the entry for *bite* (verb), and the page scrolls down to the target idiom in the *Idioms* box, and for *give somebody the cold shoulder*, the cross-referenced page goes down to the definition of the idiom in the *Idioms* box in the entry for *cold* (adjective).

Automatic scrolling to the target hyperlinked item has, however, not been found in COBUILD. The idiom *hand in hand* is cross-referenced from two consecutive numbered senses under the headword *hand*, sense 24 and sense 25 (see Figure 5), and the hyperlinked page contains two entries, the first for the literal sense and the second for the figurative one. The problem is that both cross-references go to the top of the hyperlinked page, where the literal sense is covered, and the user may overlook the other sense, the figurative one, if they do not scroll down the screen (see Figure 6).

- 22. See [get one's hands on sth](#)
- 23. See [hand in glove](#)
- 24. See [hand in hand](#)
- 25. See [hand in hand](#)
- 26. See [have a hand in sth](#)
- 27. See [a heavy hand](#)



Figure 5: Hyperlinked senses of *hand in hand* in COBUILD

The image shows two screenshots of the COBUILD dictionary entry for the phrase 'hand in hand'. The top screenshot shows the entry with a navigation bar at the top containing 'hand in hand', 'hand in hand', 'English: hand in hand', 'American: hand in hand', and 'Example sentences'. Below the navigation bar, the text reads 'Definition of 'hand in hand'', followed by the phrase 'hand in hand' in a large, bold font, and 'Collins COBUILD' below it. Under the heading 'PHRASE', the definition states: 'If two people are **hand in hand**, they are holding each other's nearest hand, usually while they are walking or sitting together. People often do this to show their affection for each other.' An example sentence follows: 'I saw them making their way, hand in hand, down the path.' Below the example, there is a link 'See full dictionary entry for hand' and a copyright notice: 'COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers'. The bottom screenshot is identical to the top one but with a horizontal line above the 'hand in hand' heading.

Figure 6: Page hyperlinked from *hand in hand* in senses 24 and 25 in COBUILD

The same treatment is noticed for the idiom *shoulder to shoulder*, hyperlinked from senses 12 and 13. Similarly, the prepositional phrase *in hand* is hyperlinked as many as four times from four consecutive senses: 30, 31, 32 and 33, with no automatic scrolling to the second, third or fourth entries representing different meanings of the expression. This practice does not facilitate the look-up process.

While examining access routes, some erroneous hyperlinks with no relevant content have been found. In CALD, there are 12 instances of *hand* MWEs hyperlinked from the MM panel to the top of the entry for *hand* (n) instead of being hyperlinked to an entry with the target MWE as the headword, for example, *at hand*, and some other prepositional phrases with *hand* such as *by hand*, *in hand*, *on hand*, *to hand*, and also *hand in hand*, *the job/matter in hand*, *put your hand in your pocket*. Coincidentally, these erroneous hyperlinks to all of them except the last one appear on the third page of the MM panel; to the last one,

on the fourth page. In fact, the other hyperlinks with the above-mentioned MWEs, if present, work fine in CALD.

3.2.3 Duplication or multiplication of hyperlinks. As has already been stated, it often happens that there are two or more ways of accessing the examined MWEs within a single body part entry. The study aims to find out (1) how many items are accessed both from the main entry (henceforth ME) and the search panel (Scenario 1), and (2) how many are hyperlinked twice or more within the search panel (Scenario 2). In Scenario 1, access from ME can take the form of a separate sense (the MWE being defined, hyperlinked or provided in an example sentence), or a hyperlink or definition in the *Idioms* or *Phrases* box or section. In Scenario 2, the focus is only on the presence of double or multiple hyperlinks within the search panel, no matter if a given expression is at the same time accessed from ME or not.

The highest percentage of double or multiple access routes is found in CALD. A staggering 85% of the body part MWEs are cross-referenced at least twice within the MM search panel (Scenario 2). Moreover, there is double access from ME and the MM panel, constituting 39% (Scenario 1). The abundance of hyperlinks results from aggregating two other dictionaries apart from CALD into the Cambridge Dictionary online: CACD and CBED. Double or multiple access to body part MWEs is also observed in the other dictionaries. In COBUILD, 35% of the examined MWEs are defined or hyperlinked in ME as well as being hyperlinked from the MI and/or RT sections (Scenario 1), whereas 11% of them are cross-referenced from both MI and RT (Scenario 2). The latter case is exemplified in Figure 7, where *(give someone) the cold shoulder* is hyperlinked from both MI and RT. The MI section is treated here separately from ME, because, as has already been mentioned, it is situated far from it, at the bottom of the page after entries from the aggregated dictionaries. Moreover, it contains hyperlinks to entries from one of the aggregated dictionaries, ELID (see Figure 8). In LDOCE, only one of the above scenarios is found, Scenario 1: 24% of the body part MWEs are hyperlinked from ME and the MR search panel. In MEDO: 32% are hyperlinked from ME and the OE search panel (Scenario 1), and a mere 2% have double hyperlinks within OE (Scenario 2). Finally, duplication or multiplication of hyperlinks is the least frequent in OALD due to the lowest number of hyperlinks: 19% of the MWEs receive double access from ME and the OR panel (Scenario 1); 5% are hyperlinked twice within OR (Scenario 2).

More idioms containing

shoulder

[a chip on your shoulder](#)
[come straight from the shoulder](#)
[give someone the cold shoulder](#) ←
[put your shoulder to the wheel](#)
[a shoulder to cry on](#)
[stand shoulder to shoulder](#)

Related terms of

shoulder

[cold shoulder](#) ←
[hard shoulder](#)
[shoulder arms](#)
[shoulder-bag](#)
[shoulder-high](#)

[View more related words](#)

Figure 7: Double hyperlinks to *(give someone) the cold shoulder* from the search panel in COBUILD

give someone the cold shoulder

to deliberately ignore someone

He gave me the cold shoulder at the party; he didn't talk to me at all.

Easy Learning Idioms Dictionary. Copyright © HarperCollins Publishers

Figure 8: The hyperlinked entry for *give someone the cold shoulder* in ELID

The highest number of hyperlinks is found in CALD for *(on the one hand ... on the other hand)*, which is cross-referenced six times in the MM panel: three times under *All meanings* and three times under *Idioms*. As has been observed in 3.2.1, the *Idioms* section in the search panel features the idioms included in the *All meanings* section, so it may as well be claimed that the expression is hyperlinked three times in the MM panel. Moreover, it is cross-referenced from the *Idioms* box at the bottom of ME (see Figure 9). Cross-referencing from the MM panel apart from the hyperlink in ME is fine, and should not be viewed as redundant, as the panel's main function is facilitating access to the searched item (see section 2.3). It is, however, worth noting that there are slightly different forms of this idiom in the hyperlinks from the MM panel (see Figure 10). As has been pointed out in 2.2, they are treated in this study as variants of the same MWE. Under *All meanings* in the MM panel, there is one cross-reference to *on the one hand ... on the other hand* in CALD, and cross-references to two variant forms in CACD *on the one hand but on the other hand* and *on the other hand*. In the *Idioms* box at the bottom of ME, there is a hyperlink to the form *on the one hand ... on the other hand*.

hand noun (MEASUREMENT)

[C]

a unit for measuring the height of a horse up to its shoulder:

• One hand equals four inches (= 10.16 centimetres).

+ SMART Vocabulary: related words and phrases

Idioms

at hand	at the hands of sb
a game in hand	get/lay/put your hands on sb
get/lay/put your hands on sth	go hand in hand with sth
hand in glove	hand in hand
hand over fist	hand to hand
- More idioms	
sb's hands are tied	have your hands full
have sth in hand	have sth on your hands
hold/put your hands up	in hand
the job/matter in hand	keep your hand in
keep a firm hand on sth	live (from) hand to mouth
on hand	on the one hand ... on the other hand
out of hand	put your hand in your pocket
a safe pair of hands	take sb/sth in hand
win (sth) hands down	

Figure 9: Hyperlink to *on the one hand ... on the other hand* from the Idioms box in CALD

[on the one hand ... on the other hand](#) idiom

[out of hand](#) idiom

[put your hand in your pocket](#) idiom

[sb's hands are tied](#) idiom

[take sb/sth in hand](#) idiom

[win \(sth\) hands down](#) idiom

[to hand, at on hand](#) idiom

[a firm hand](#) idiom

[near at hand](#) idiom

[ready to hand](#) idiom

[lend \(someone\) a hand, at give someone a hand](#) idiom

[hand and glove, at hand in glove](#) idiom

[lay a hand on someone, at lay a finger on someone](#) idiom

[tip your hand](#) idiom

[force sb's hand](#) idiom

[on the other hand](#) idiom

Figure 10: Extract from the MM panel in CALD with hyperlinks to variant forms of the idiom (*on the one hand ...*) *on the other hand*

One may wonder if all these hyperlinks take the user to the same entry, with an identical definition and the same example sentences. As it is rightly assumed, this is the case for the hyperlinked entries within CALD, but may not necessarily be so for the ones from CACD, especially as the canonical forms slightly differ. Indeed, as Figure 11 demonstrates, in CACD, the definition and example sentences provided in the cross-referenced entries for *on the one hand but on the other hand* and the truncated form *on the other hand* are different.



Figure 11: Two different hyperlinked entries for two variant forms in CACD

In fact, in the first hyperlinked entry from CACD quoted above, the example sentence is the same as in the cross-referenced entry in CALD. Definitions are, however, different in all the three hyperlinked entries, including the metalinguistic definition in CALD: 'used when you are comparing two different facts or two opposite ways of thinking about a situation'. This multiplication of hyperlinks with differing content appears to be redundant and also misleading for the user, who may think that these are distinct expressions, not just variants of the same form. It is justifiable if differing hyperlinks come from two aggregated dictionaries, like here CALD and CACD, but should be viewed as examples of redundancy and inconsistency if they are from the same dictionary, here from CACD. It is also suggested that there should be one canonical form with the first part of the idiom given in brackets (*on the one hand ...*) *on the other hand*.

Two other of the dictionaries under scrutiny, COBUILD and MEDO, also provide more than one hyperlinked entry for this MWE. In COBUILD *on the one hand* is hyperlinked under sense 41 and *on the other hand* under sense 42. In fact, the second part of the idiom *on the other hand* is embedded in the definition and example sentence under *on the one hand*, which may suggest that one hyperlink will do:¹⁰

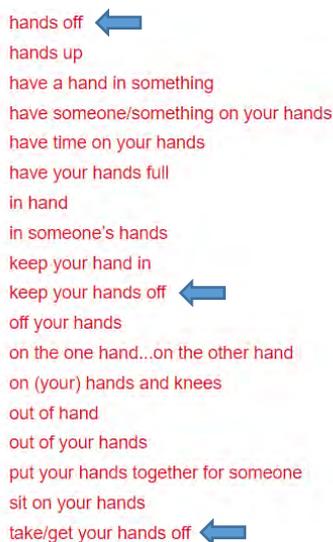
You use **on the one hand** to introduce the first of two contrasting points, facts, or ways of looking at something. It is always followed later by **on the other hand** or 'on the other'.

On the one hand, if the body doesn't have enough cholesterol, we would not be able to survive. On the other hand, if the body has too much cholesterol, the excess begins to line the arteries.

(COBUILD)

In MEDO, there is also double access to the idiom, from the *Phrases* section in ME and the OE panel. Both hyperlinks take the user to the same entry for *on the one hand...on the other hand*. Here, the duplication is justified, OE featuring as the search panel in MEDO. In contrast, LDOCE and OALD cover this idiom only once. OALD defines the form (*on the one hand...*) *on the other (hand)...* in the *Idioms* box in ME, without any cross-reference. LDOCE hyperlinks it from a separate numbered sense *on the other hand*, and provides the full form *on the one hand...on the other hand* as an alternative in brackets.

Other cases of multiple hyperlinks resulting from unnecessary recognition of distinct canonical forms have also been found. In MEDO, there are cross-references to three separate entries *hands off*, *keep your hands off* and *take/get your hands off*, all three of them hyperlinked from the *Phrases* section (see Figure 12), whereas, for example, in LDOCE there is only one hyperlink to the entry for *hands off* with the variant form *keep your hands off* embedded in an example sentence. The definitions in the three hyperlinked entries in MEDO convey the same meaning, but the wording is different (see the extracts below Figure 12). Hence, it is suggested that these multiple cross-references are redundant and should be replaced with one hyperlinked canonical form.



hands off ←
hands up
have a hand in something
have someone/something on your hands
have time on your hands
have your hands full
in hand
in someone's hands
keep your hand in
keep your hands off ←
off your hands
on the one hand...on the other hand
on (your) hands and knees
out of hand
out of your hands
put your hands together for someone
sit on your hands
take/get your hands off ←

Figure 12: Multiplication of hyperlinks in the *Phrases* section in MEDO

hands off

used for telling someone not to touch or take someone or something

keep your hands off

to not touch someone or something

take/get your hands off

to stop touching someone or something

(MEDO)

The duplication or multiplication of entries, however, is not dictionary-specific, since a reverse situation is observed for the idiom *give somebody the cold shoulder*, with more hyperlinks found in LDOCE than in MEDO. In MEDO the idiom is hyperlinked twice: in the *See also* section at the bottom of ME, and in the OE search panel, in both cases to the same entry for *cold shoulder*, which incorporates the phrase *get/be given the cold shoulder*:

cold shoulder

PHRASE

get/be given the cold shoulder

to be treated in an unfriendly way by someone you know

(MEDO)

LDOCE, however, unnecessarily multiplies hyperlinks, placing *give somebody the cold shoulder* among the MWEs grouped at the bottom of ME before the *Collocations* box, and *cold shoulder* and *(give somebody/get) the cold shoulder* in the MR panel. There is one more hyperlinked entry in the MR panel, but it is the same as the one cross-referenced from the MWEs at the bottom of ME. As the dictionary extracts below illustrate, the definitions and most example sentences for the three variant forms are different. Interestingly, the hyperlinked entry for *cold shoulder* provides a cross-reference to the entry for *(give somebody/get) the cold shoulder* instead of a definition. The multiplication of cross-references seems redundant, also because a few example sentences overlap, and some under *cold shoulder* contain the collocate *give* from the other two variant forms (see the underlined examples and the highlighted collocates). Hence, one canonical form with the first part of the idiom *give somebody* in brackets would suffice.

give somebody the cold shoulder

to deliberately ignore someone or be unfriendly to them, especially because they have upset or offended you

- *After I got the promotion, a few of my co-workers started giving me the cold shoulder.*
- *So they have given girlfriends the cold shoulder.*
- *She was sure that at some point she'd given some one the cold shoulder and hurt them badly without noticing.*
- *Vernon regretted giving Harcourt the cold shoulder; he would have been some one to laugh with.*

(give somebody/get) the cold shoulder

to behave in an unfriendly way towards someone that you know

- *So they have given girlfriends the cold shoulder.*
- *Giving the cold shoulder to his usual tippie, Ian Knight raises his coffee cup to Drinkwise Day.*
- *She was sure that at some point she'd given some one the cold shoulder and hurt them badly without noticing.*
- *A declaration of love, or the cold shoulder.*

cold shoulder

→ **(give somebody/get) the cold shoulder**

- *Then from then on, we were treated with an absolute cold shoulder, and no one would speak to us.*
- *Rachaela had turned on Ruth, not just the habitual cold shoulder, but with a firework of dislike and alienation.*
- *So they have given girlfriends the cold shoulder.*
- *She was sure that at some point she'd given some one the cold shoulder and hurt them badly without noticing.*
- *A declaration of love, or the cold shoulder.*
- *Banishment was better than this cold shoulder.*

(LDOCE)

The analysis shows that the differences between the hyperlinked variant forms can really be of little significance. For instance, *lend an ear* is hyperlinked twice in the OE search panel in MEDO, the only difference being the presence of the preposition *to* in brackets in one of the cross-references (see Figure 13).

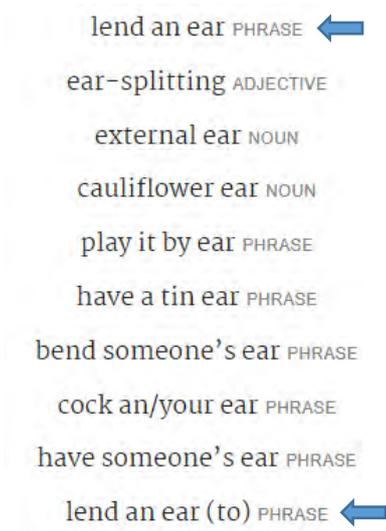


Figure 13: Duplication of hyperlinks in the search panel in MEDO

The definitions, though differently worded, convey the same meaning, and should be merged under one entry. The former entry is also cross-referenced from the *Phrases* section.

lend an ear

to listen to someone talking about their problems

lend an ear (to)

to listen to someone, especially someone who is upset

Just lend a sympathetic ear while she tells you her problems.

(MEDO)

Similarly, *play something by ear* is hyperlinked three times from the OE panel in MEDO, the MWEs having almost identical forms: *play it by ear*, *play something by ear* and *play (something) by ear*. The middle hyperlinked entry on its own would be enough, as it covers both the literal and figurative meanings of the expression (in reverse order). In fact, this is the one also cross-referenced from the *Phrases* section at the bottom of ME.

play it by ear

to decide what to do as a situation develops rather than planning what you will do before you start

We don't know how many people are going to come, so we'll play it by ear.

play something by ear

1. to deal with a situation without having a plan, by reacting to things as they happen
2. to play music by hearing and remembering it, not by reading it

play (something) by ear

to perform a piece of music by remembering what it sounds like, without looking at the notes

(MEDO)

Yet another configuration of double access routes is found for the examined MWEs, but its occurrence is rare in the present study. This is double access within ME. The representation of the expression *in the hands of somebody/in somebody's hands* in LDOCE serves as an example. First, the expression is embedded in an example sentence under sense 3 tagged with the signpost 'control', and then it is hyperlinked under sense 8 (see Figure 14). Semantic distance is unnecessarily created between senses 3 and 8. Double access within ME can also be observed in the *Idioms* box in OALD, where two variant forms of the same expression are sometimes defined separately. This has already been discussed in 3.1.

- 3 **CONTROL** [singular, uncountable] control, power, or influence that someone has
- 🔊 The president has **strengthened the hand of** the gun lobby.
 - 🔊 This matter is too important to be left **in the hands of** (=in the control of) an inexperienced lawyer. ←
 - 🔊 a manager with a **firm hand** (=who controls things strictly)
- 4 → **get out of hand**
- 5 → **on the other hand**
- 6 → **hands off**
- 7 → **in hand**
- 8 → **in the hands of somebody/in somebody's hands** ←

Figure 14: Double access routes within ME in LDOCE for *in the hands of sb*

4. Item-specific treatment

In order to find out if there is any item-specific treatment with regard to access routes, *hand*, *shoulder* and *ear* MWEs included in all the five MELDs were selected. The total amounts to 49 shared items, which constitutes 35–49% of the examined MWEs, depending on the dictionary. Numerous MWEs are covered in four or three of the five MELDs. Leaving the issue of absence of some body part expressions for future research, it was investigated if any MWEs are presented in a similar way.

The study shows that there is a tendency for some MWEs to be defined on the page of the body part headword under one of the first senses. Their top positions in the entry are, obviously, mainly due to the corpus-based frequency. The expression *give/lend somebody a hand* is one of them, together with its variants, *need/want a hand*. These forms are provided in example sentences under sense 3 in OALD, sense 4 in CALD, COBUILD and MEDO, and under hyperlinked sense 2 in LDOCE. Additionally, these senses are tagged with the metonymic meaning 'help' as a navigational device guiding the user to the appropriate sense in CALD, MEDO and OALD. A hyperlink is used only in LDOCE, to the entry *a hand* (see Figure 15). This treatment may mislead the user, who does not see the relevant MWE on the page of the headword, not even as a hyperlink.

hand¹ /hænd/ ●●● S1 W1 noun 🗣️ 🗣️

1 **PART OF BODY** [countable] the part of your body at the end of your arm, including your fingers and thumb, that you use to hold things

- 🗣️ Steve gripped the steering wheel tightly with both hands.
- 🗣️ In her hand was a tattered old photograph.

2 → a hand ←

3 **CONTROL** [singular, uncountable] control, power, or influence that someone has

- 🗣️ The president has **strengthened the hand of** the gun lobby.
- 🗣️ This matter is too important to be left **in the hands of** (=in the control of) an inexperienced lawyer.
- 🗣️ a manager with a **firm hand** (=who controls things strictly)

a hand

help with something – used in the following phrases

need/want a hand

- 🗣️ Do you need a hand packing?

give/lend (somebody) a hand

- 🗣️ Can you give me a hand to lift this?
- 🗣️ If you get stuck, Denise is always willing to lend a hand.

I could do with a hand/use a hand (=it would be useful to have some help)

- 🗣️ We could certainly do with a hand.

→ a helping hand

Figure 15: A hyperlink to the 'help' meaning of *hand* in LDOCE

Another metonymic extension of the literal *hand*, 'somebody who does physical work', in most of the examined MELDs tagged with the navigational device 'worker' or 'person', also tends to be defined under one of the first senses in four of the five MELDs: sense 5 in CALD and COBUILD, sense 2 in MEDO and senses 6 and 7 in OALD, sense 7 separately for the 'sailor' meaning. Only in LDOCE it is defined under sense 39, separated from the other definitions by numerous hyperlinks. The relevant compound nouns such as *factory hand*, *farm hand*, *hired hand*, *deck hand* feature in example sentences, and they are also sometimes hyperlinked from the respective senses (in CALD and OALD). As for MWEs with the other body part names, *have an ear for sth* is defined under one of the first senses in four out of five MELDs under scrutiny (sense 2 in COBUILD and MEDO, sense 6 in LDOCE and sense 3 in OALD). The expression *on sb's shoulders* is also ordered at the top in three of the dictionaries: under sense 2 in CALD, sense 3 in COBUILD and hyperlinked sense 7 in LDOCE. However, it is hyperlinked from the *Phrases* section in MEDO, and defined in the *Idioms* box in OALD.

The analysis shows that apart from the obvious corpus frequency criterion, MWEs which are less fixed and more modifiable, and hence subject to variations in form tend to be defined at the top of the body part entry. The definition under a given sense provides the general (metonymic) meaning of the body part, and the variant forms of a MWE appear in example sentences.

Nonetheless, the study proves that the MWEs present in all the 'Big Five' very rarely exhibit item-specific treatment which would manifest itself in common access routes to certain expressions. The same MWE frequently appears in different locations in each of the MELDs, defined or hyperlinked. A case in point is *have a hand in sth*. It is defined under separate senses in CALD and OALD (senses 6 and 4 respectively), hyperlinked from separate senses in COBUILD and LDOCE (senses 26 and 21 respectively), cross-referenced from the *Idioms* or *Phrases* sections in COBUILD and MEDO, and hyperlinked from the search panels in CALD and MEDO. This and similar examples of different treatment of a given expression in the five dictionaries are probably due to the already mentioned lack of clear classifications of MWEs in literature.

No item-specific treatment has also been found with respect to duplication or multiplication of access routes. Here the number of double or multiple access routes from ME (the MWEs being defined, hyperlinked or provided in example sentences) and the search panel was taken into account, not the number of double or multiple hyperlinks within the search panel, as LDOCE does not exhibit the latter form of duplication. In fact, only one item of the 49 shared MWEs receives double access in all the five dictionaries, the prepositional phrase (*close/near*) *at hand*. Additionally, 10 MWEs (20%) demonstrate duplication or multiplication of hyperlinks in four of the five MELDs. Thus, the majority of the expressions have double or multiple access routes merely in three or fewer MELDs. As Figure 16 shows, the proportions of double or multiple access routes to the 49 MWEs in each of the dictionaries correspond to the ones to all the MWEs presented in 3.2.3. Again the duplication or multiplication of access routes is the most frequent in CALD and COBUILD (55%), and the least frequent in OALD (27%), which proves that it is dictionary-specific, not item-specific.

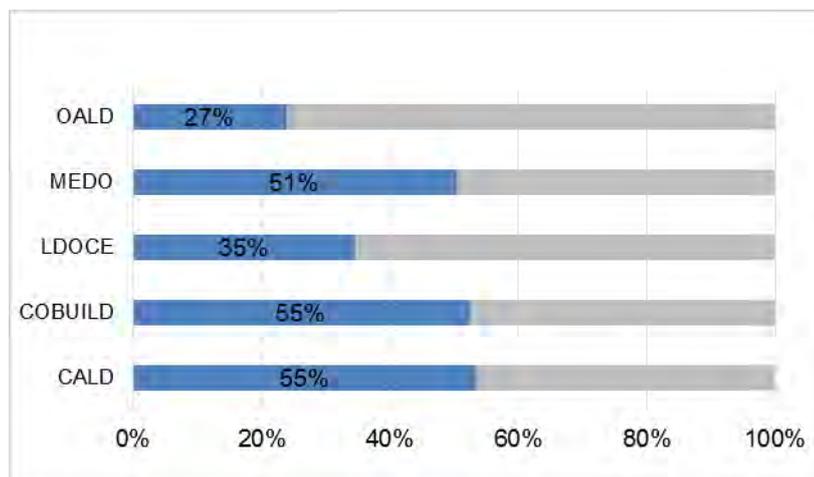


Figure 16: The proportions of double or multiple access routes to the 49 MWEs shared by the 'Big Five'

5. Conclusions

The findings of the study reveal that access routes to body part MWEs are relatively consistent in the MELDs under analysis, and clear preferences have been described for each dictionary. When the dictionaries are compared, it turns out that access routes rarely depend on a specific item. However, some general tendencies in accessing body part MWEs have been observed. It appears that the current practice is to hyperlink MWEs instead of defining them on the page of the headword. The huge majority of the examined body part expressions are hyperlinked to a separate entry. The only exception is OALD, which adopts a reverse approach, defining most of the items in the body part entry, mostly in the *Idioms* box. Another similarity is locating MWEs in three main positions: separate senses, *Idioms* or *Phrases* boxes or sections and search panels, with dictionary-specific distribution.

Duplication or multiplication of access routes within a single body part entry is also observed in all the dictionaries, in some of them to quite a large extent. Double or multiple hyperlinks are found from ME and the search panel or within the search panel itself. The abundance of cross-references is the result of access to aggregated dictionaries and the use of a few slightly differing canonical forms. Double or multiple hyperlinks should not always be seen as something redundant. They seem fine and, in fact, unavoidable, if they come from distinct aggregated dictionaries. They are, however, cases of redundancy and inconsistency if the cross-referenced entries belong to the same dictionary and include different definitions and example sentences, and if the separately hyperlinked MWEs are variants of the same canonical form. In such cases semantic distance between related forms is unnecessarily created. Hyperlinking items both from ME and the search panel is also something self-evident, bearing in mind the role of the latter in helping the user find the target item.

It is hoped that the study shows the need for further research into the area of access routes to MWEs which will incorporate a wider scope of expressions. This might contribute to a more consistent and cognitively oriented representation of MWEs in MELDs.

Endnotes

1. Due to the search technique now widely applied in e-lexicography known as incremental search or search-as-you-type, users can see the target item among the suggestions that appear in the search box already after typing a few characters. Such pull-down lists with suggested items also include MWEs.
2. See Gouws (2018: 46) for the distinction between *access structure* and *search route/path*.
3. For definitions of metaphor and metonymy in cognitive linguistics, see, for example, Lakoff and Johnson ([1980] 2003) and Ungerer and Schmid ([1996] 2006).
4. These are free online versions, where access codes are not required.

5. American versions of the dictionaries are excluded from the study because they are just American counterparts of the same dictionaries.
6. In LDOCE entries from *Longman Business Dictionary* are presented at the bottom of the page, in MEDO no dictionaries are aggregated apart from the American version, and OALD at the bottom of the page provides hyperlinks to the respective entries, not distinct MWEs, from *Oxford Learner's Dictionary of Academic English* available with an access code, and under certain senses offers hyperlinks to *Oxford Collocations Dictionary*, also accessed with a purchase code.
7. Before February 2020, the search panel in MEDO was named *Related words*.
8. In this paper, the term *cross-reference* is used interchangeably with *hyperlink*, which is its electronic form.
9. The compounds *glue ear* and *hard shoulder* do not belong to the examined MWEs, as they are based on metaphor rather than metonymy.
10. In the entries cited in the paper, exact typography has not been replicated and information redundant for the study has been omitted.

References

Dictionaries

- Cambridge Advanced Learner's Dictionary*. (CALD) Accessed on 10 February 2020:
<http://www.dictionary.cambridge.org>
- COBUILD Advanced English Dictionary*. (COBUILD) Accessed on 10 February 2020:
<http://www.collinsdictionary.com>
- Longman Dictionary of Contemporary English*. (LDOCE) Accessed on 10 February 2020:
<http://www.ldoconline.com>
- Macmillan English Dictionary Online*. (MEDO) Accessed on 10 February 2020:
<http://www.macmillandictionary.com>
- Oxford Advanced Learner's Dictionary*. (OALD) Accessed on 10 February 2020:
<http://www.oxfordlearnersdictionaries.com>

Other literature

- Adamska-Sałaciak, A.** 2008. Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, EURALEX 2008, Barcelona, 15–19 July 2008*: 1477-1485. Barcelona: Universitat Pompeu Fabra.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Boers, F. and S. Lindstromberg.** 2006. Cognitive Linguistic Applications in Second or Foreign Language Instruction: Rationale, Proposals, and Evaluation. Kristiansen, G., M. Achard, R. Dirven and F.J. Ruiz de Mendoza Ibáñez (Eds.). 2006. *Cognitive Linguistics: Current Applications and Future Perspectives*: 305-355. Berlin/New York: Mouton de Gruyter.

- Cowie, A.P. (Ed.).** 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.
- Dobrovolskij, D. and E. Piirainen.** 2005. *Figurative Language: Cross-Cultural and Cross-Linguistic Perspectives*. Amsterdam: Elsevier.
- Fellbaum, C.** 2007. Argument Selection and Alternations in VP Idioms. Fellbaum, C. (Ed.). 2007. *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies*: 188-202. London: Continuum.
- Geeraerts, D.** 2002. The Interaction of Metaphor and Metonymy in Composite Expressions. Dirven, R. and R. Pörings (Eds.). 2002. *Metaphor and Metonymy in Comparison and Contrast*: 435-465. Berlin: Mouton de Gruyter.
- Geeraerts, D.** 2007. Lexicography. Geeraerts, D. and H. Cuyckens (Eds.). 2007. *The Oxford Handbook of Cognitive Linguistics*: 1160-1174. Oxford: Oxford University Press.
- Gibbs, R.** 1993. Why Idioms are not Dead Metaphors. Cacciari, C. and P. Tabossi (Eds.). 1993. *Idioms: Processing, Structure, and Interpretation*: 57-77. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goossens, L.** [1990] 2002. Metaphtonymy: The Interaction of Metaphor and Metonymy in Expressions for Linguistic Action. Dirven, R. and R. Pörings (Eds.). 2002. *Metaphor and Metonymy in Comparison and Contrast*: 349-377. Berlin: Mouton de Gruyter.
- Goows, R. H.** 2018. Dictionaries and Access. Fuertes-Olivera, P.A. (Ed.). 2018. *The Routledge Handbook of Lexicography*: 43-58. London/New York: Routledge.
- Hanks, P.** 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, Massachusetts: The MIT press.
- Kövecses, Z. and S. Csábi.** 2014. Lexicography and Cognitive Linguistics. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 27(1): 118-139.
- Lakoff, G. and M. Johnson.** [1980] 2003. *Metaphors We Live By*. Chicago/London: University of Chicago Press.
- Langlotz, A.** 2006. *Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom-Representation and Idiom-Variation in English*. Human Cognitive Processing 17. Amsterdam/Philadelphia: John Benjamins.
- Lew, R.** 2012. How can We Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 343-361. Oxford: Oxford University Press.
- Moon, R.** 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Oxford Clarendon Press.
- Moon, R.** 2004. On Specifying Metaphor: An Idea and its Implementation. *International Journal of Lexicography* 17(2): 195-222.
- Oppentocht, L. and R. Schutz.** 2003. Developments in Electronic Dictionary Design. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 215-227. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.
- Ostermann, C.** 2015. *Cognitive Lexicography: A New Approach to Lexicography Making Use of Cognitive Semantics*. Lexicographica. Series Maior 149. Berlin: De Gruyter.
- Palmer, H.E.** 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Sinclair, J. (Ed.).** 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London/Glasgow: Collins.
- Sinclair, J. (Ed.).** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stathi, K.** 2007. A Corpus-based Analysis of Adjectival Modification in German Idioms. Fellbaum, C. (Ed.). 2007. *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies*: 81-108. London: Continuum.

Ungerer, F. and H.-J. Schmid. [1996] 2006. *An Introduction to Cognitive Linguistics*. London/New York: Longman.

Van der Meer, G. 1999. Metaphors and Dictionaries: The Morass of Meaning, or How to Get Two Ideas for One. *International Journal of Lexicography* 12(3): 195-208.

A Survey of Indonesian Students' Use of Dictionaries

Fallianda, *English Department, Faculty of Humanities,
Airlangga University, Surabaya, Indonesia*
(fallianda.yand@gmail.com)

Abstract: This article aims to define the profile of Indonesian speakers as dictionary users by utilizing a questionnaire-based research. The survey was distributed to university students who were studying English as a foreign language. The students fall into two groups with pre-intermediate and intermediate levels of English proficiency. They were instructed to complete the survey. The question items encourage students to report their consultation frequency of dictionary use, dictionary choices and ratings, and dictionary information needs. The findings confirmed that both pre-intermediate and intermediate students use bilingual dictionaries more often than monolingual dictionaries. The two groups show indifference towards the evaluation of different dictionary types, since learners rate bilingual dictionaries as good as monolingual dictionaries. The collected data offers insights into students' awareness and knowledge of various dictionary types in the Indonesian lexicographic context.

Keywords: DICTIONARY AWARENESS, DICTIONARY USE, INDONESIAN UNIVERSITY STUDENTS, PROFICIENCY LEVELS, QUESTIONNAIRE, PROFILE OF DICTIONARY USERS

Opsomming: 'n Opname van Indonesiese studente se gebruik van woordeboeke. In hierdie artikel word beoog om die profiel van Indonesiese sprekers as woordeboekgebruikers met behulp van vraelysgebaseerde navorsing te definieer. Die vraelys is gegee aan universiteitstudente wat Engels as vreemde taal bestudeer. Die studente kan volgens vaardigheid in Engels in twee groepe verdeel word, nl. dié met pre-intermediêre en dié met intermediêre vaardigheidsvlakke. Hulle het die opdrag gekry om die vraelys te voltooi. Die studente moet o.a. verslag doen oor hul frekwensie van woordeboekgebruik, woordeboekkeuses en -waardebepaling, en hul behoeftes rakende woordeboekinligting. Die bevindings het bevestig dat albei groepe studente (pre-intermediêr en intermediêr) meer dikwels tweetalige woordeboeke as eentalige woordeboeke gebruik. Die twee groepe maak geen onderskeid by die evaluering van verskillende woordeboektipies nie aangesien hulle tweetalige woordeboeke as net so goed as eentalige woordeboeke beskou. Die verkrygte data bied insig in die studente se bewustheid en kennis van verskillende woordeboektipies in die Indonesiese leksikografiese konteks.

Sleutelwoorde: WOORDEBOEKBEWUSTHEID, WOORDEBOEKGEBRUIK, INDONESIESE UNIVERSITEITSTUDENTE, VAARDIGHEIDSVLAKKE, VRAELYS, PROFIEL VAN WOORDEBOEKGEBRUIKERS

1. Introduction

From all the specialisations in dictionary research, the so-called user-perspectives have made a significant contribution to facts about user awareness of the availability and usability of various kinds of dictionaries and other reference works (Müller-Spitzer, Domínguez Vázquez, Nied Curcio, Silva Dias and Wolfer 2018). This allows studies to assess dictionary culture in a particular language community to discover whether users are aware of the value and limitations of the dictionaries they use. There has been increasingly important dictionary research which focused on users' needs for and skills in acquiring dictionary information (Atkins and Varantola 1998; Béjoint 1981; Hatherall 1984; Lew 2004; Tomaszczyk 1979). Unfortunately, the majority of the research has been conducted in English-speaking countries. Against this background, research needs to be done into an Indonesian lexicographic context in which Indonesian speakers might have different ideas on dictionary use and its values (Kwary 2018). However, there has only been little empirical data on dictionary user research by Indonesian speakers.

With this in mind, we collected information about the Indonesian lexicographic tradition. Accordingly, we found that the government considers dictionaries as systems of information storage (Moeljadi, Kamajaya and Amalia 2017) while few researchers took an interest in users' considerations for using a specific type of dictionary (Amalia 2014; Yanto 2016). The reason for the lack of empirical data to define the profile of Indonesian dictionary users is probably that Indonesian speakers have not prioritized dictionaries as one of the necessary reference tools (Badan Bahasa Upayakan Leksikografi Berkembang di Indonesia 2017). Methodological issues emerge for collecting information about dictionary culture in a community where dictionary use must occur. Therefore, researchers can solve the problems by carefully choosing representatives to conduct the survey, such as university students or learners of the English language. In addition, as suggested by Hartmann (2000: 390), researchers should consider including all lexicographic information to raise dictionary awareness. The collected data will be valuable information since the results could contribute to enhance a dictionary culture among particular language communities.

2. The study

Several studies examining users' perspectives have explored specific issues, such as user attitudes, behaviours, preferences, difficulties, dictionary misuse, reference needs and reference skills (Nesi 2014). Examples of important studies include a few empirical studies before the 1980s, such as Tomaszczyk's (1979) and Baxter's (1980) research. They reported preferences for bilingual dictionaries by EFL learners and difficulty in understanding monolingual dictionaries. On the contrary, Béjoint (1981) found that most French students of English

(96% of them) possessed monolingual dictionaries as recommended by their tutors. There was also a massive project, the EURALEX and AILA Research Project into Dictionary Use (Atkins and Varantola 1998). The study presented how users from four linguistic communities regard different types of dictionaries; thus it was found that the users prefer bilingual dictionaries to help them with particular language tasks. The superiority of the bilingual type was also reported by Polish students in Lew's study (2004). The study found a very low consultation frequency of monolingual dictionaries across levels, except the advanced level. It was not surprising, therefore, to find that the students mostly consulted their dictionaries for meaning, English equivalents and Polish equivalents. Although the study of users' perspectives has made impressive progress, the majority of research tended to be conducted in English-speaking countries. None of the above-mentioned studies contributed to show dictionary culture among Indonesian speakers, especially in the higher education context in which dictionaries become one of the important reference tools to support their learning process.

3. Methodology

A questionnaire survey was distributed to eighty-eight students at Airlangga University, Indonesia. The participants were in their first year studying English language and literature. The university students were chosen as representatives of the study because dictionaries are seen to be one of the important reference materials to support their learning process. With this sampling method, the results of the study would provide a better understanding on how Indonesian students acknowledge the value and limitations of dictionaries. The students' levels were equivalent to the A2 and B1 levels of the CEFR (Common Framework of Reference for Languages). The two levels are commonly referred to as pre-intermediate and intermediate levels.

3.1 The user questionnaire

The students were instructed to complete the user questionnaire that was distributed in a paper format. For reasons of different lexicographic traditions, language skills, and exposure to dictionary types, two researchers might have had different research questions and this might have been reflected in question items and their formats. The questionnaire form (see Appendix 1) is almost self-explanatory. On the top right-hand corner, there was a small space for the student identification number which would provide information about the students' English proficiency levels. The students had been told that their participation in the study would not affect their grades and they are expected to answer all the questions truthfully. There were three sections which consisted of several details.

Section A identifies how often the students use certain dictionary categories (Indonesian–English, English–Indonesian, and English–English). The question item gives a range of frequency options from 6–7 days a week, 5–4 days a week, 3–2 days a week, and 1–0 days a week).

Section B asks the students to rank the first, second, and third dictionaries they often use, then to give ratings (excellent, good, OK, poor, and awful) for each dictionary mentioned. For this item, the survey also provides information about dictionaries that are readily available in Indonesia (see Appendix 2). It was possible for the students to mention other kinds of dictionaries and rejected dictionaries they did not actually use. The question item is in the form of open-ended questions. The instruction has made it clear that the participants are encouraged to give as many details as they could (e.g. details about type, title, publisher, author, edition, or medium).

Section C asks the participants to report how often they look for certain information in a dictionary. In this question item, the students were given frequency options of always, often, sometimes, seldom and never. The typical presentation and symbol for each lexicographic information type are given in case the participants cannot relate to different kinds of information given in dictionary entries.

4. Results and discussion

4.1 Frequency of dictionary use

In item A1 of the user questionnaire, the students were asked to identify how often they consult bilingual dictionaries in the direction of Indonesian–English. A detailed summary of the students' responses is presented in Table 1.

Students' Levels					
Pre-intermediate			Intermediate		
Consultation Frequency	n	%	Consultation Frequency	n	%
6–7 days a week	5	11%	6–7 days a week	6	15%
4–5 days a week	8	18%	4–5 days a week	11	27%
2–3 days a week	19	42%	2–3 days a week	15	37%
0–1 day a week	13	29%	0–1 day a week	9	22%
Total	45	100%	Total	41	100%

Table 1: Frequency of Indonesian–English dictionary consultation by pre-intermediate and intermediate students

As shown in Table 1, the frequency use of Indonesian–English dictionaries increases across levels. However, the statistical results show no significant rela-

tionship between the students' levels and the frequency. This means that both groups have a similar consultation frequency when using Indonesian–English dictionaries.

In item A2, the students' responses were collected to find how often bilingual dictionaries in the direction of English–Indonesian are consulted. A detailed breakdown of responses is presented in Table 2.

Students' Levels					
Pre-intermediate			Intermediate		
Consultation Frequency	n	%	Consultation Frequency	n	%
6–7 days a week	3	7%	6–7 days a week	7	17%
4–5 days a week	12	27%	4–5 days a week	13	32%
2–3 days a week	19	42%	2–3 days a week	13	32%
0–1 day a week	11	24%	0–1 day a week	8	20%
Total	45	100%	Total	41	100%

Table 2: Frequency of English–Indonesian dictionary consultation by pre-intermediate and intermediate students

As shown in Table 2, even though there is a tendency of intermediate students to use English–Indonesian dictionaries more often than pre-intermediate students, there is no significant effect of different levels on the frequency of English–Indonesian dictionary consultation. From this fact, it is suggested that the groups show similar consultation frequency when using English–Indonesian dictionaries.

In item A3, the students responded to a question on how often they used monolingual English dictionaries. Table 3 below shows a detailed breakdown of the students' responses.

Students' Levels					
Pre-intermediate			Intermediate		
Consultation Frequency	n	%	Consultation Frequency	n	%
6–7 days a week	3	7%	6–7 days a week	0	0%
4–5 days a week	5	11%	4–5 days a week	7	18%
2–3 days a week	16	35%	2–3 days a week	10	26%
0–1 day a week	22	48%	0–1 day a week	22	56%
Total	46	100%	Total	39	100%

Table 3: Frequency of monolingual English dictionary consultation by pre-intermediate and intermediate students

As shown in Table 3, most of pre-intermediate and intermediate students reported using monolingual English dictionaries 0–1 day a week. This means that the two groups rarely use monolingual dictionaries. An interesting finding appears as the use of monolingual English dictionaries did not seem to increase across levels, indeed the consultation frequency seemed lower among intermediate students. The result of the statistical test, however, shows no significant difference between the students' levels. Therefore, this means that none of the groups used monolingual English dictionaries more often than the other.

It was expected that the students with a higher level would use monolingual dictionaries more often than those with a lower level of English proficiency. Let us now examine the effect of levels on the consultation frequency. The tendency that emerges from the test results has shown no significant effect ($R^2=4\%$, $p=0.29$) of levels on frequency of dictionary use. We can conclude that the two groups of students have similar habits in using dictionaries, in which monolingual dictionaries are rarely used and English–Indonesian dictionaries are consulted as often as Indonesian–English dictionaries.

4.2 Dictionary preference

In section B of the user questionnaire, the students were asked to name three dictionaries they used most frequently. We had set the item by providing three possibilities consisting of three different kinds of dictionaries from which to choose to complete it. Moreover, the three possibilities in the questionnaire are provided to emphasize the students' first choice (item B1), second choice (item B2), and third choice (item B3). A detailed number and percentage of students' responses are separately presented in tables for each level of English proficiency.

From 73 valid responses collected from pre-intermediate students, 42 pre-intermediate students reported using at least one dictionary. Half of pre-intermediate students reported using two dictionaries, while 10 students were identified using three dictionaries to help them with their day-to-day learning process. In Table 4, we provide readers with details of dictionaries mentioned by pre-intermediate students.

No.	Dictionary Code	1st Choice	2nd Choice	3rd Choice	Total	Percent
1.	OALD	16	3	0	19	24%
2.	Google Translate	10	4	0	14	18%
3.	<i>Kamus Indonesia–Inggris/Kamus Inggris–Indonesia</i>	4	3	3	10	13%
4.	<i>Kamusku</i>	6	1	2	9	11%
5.	<i>Kamus Indonesia–Inggris/Kamus Inggris–Indonesia</i> (Echols and Shadily 2014/2017)	2	4	1	7	9%
6.	CALD	0	1	1	2	3%

7.	Line Dictionary	0	2	0	2	3%
8.	LDCE	1	1	0	2	3%
9.	Merriam Webster	1	1	0	2	3%
10.	COBUILD	1	0	0	1	1%
11.	Dictionary.com	0	0	1	1	1%
12.	English monolingual dictionary	0	1	0	1	1%
13.	Flitto	0	0	1	1	1%
14.	<i>Kamus Password English Learner's Dictionary for Speakers of Bahasa Indonesia</i>	0	0	1	1	1%
15.	U-Dictionary	1	0	0	1	1%
16.	Total	42	21	10	73	100%

Table 4: Dictionaries most often used by pre-intermediate students

As shown in Table 4, the OALD is reported to be the most consulted dictionary, as 16 intermediate students claimed to use it as their first choice and 3 additional students preferred to use it as their second choice. The second- and third-ranking lists belong to Google Translate which is better known than unidentified formats of *Kamus Indonesia-Inggris/Inggris-Indonesia*. In the fourth and fifth ranks, there are *Kamusku*, an electronic and off-line bilingual dictionary and *Kamus Indonesia-Inggris/Kamus Inggris-Indonesia* (Echols and Shadily 2014/2017), one of the printed bilingual dictionaries widely available in Indonesia. The current findings confirmed the domination of OALD for many years in both dictionary sales and teaching practices in the EFL context. In Table 5, we present the summary of dictionary preferences, whether dictionaries are categorized as monolingual or bilingual dictionaries.

Dictionary types	1st choice	2nd choice	3rd choice	Total
Bilingual	23 (54.8%)	14 (66.7%)	8 (80%)	45 (61.6%)
Monolingual	19 (45.2%)	7 (33.3%)	2 (20%)	28 (38.4%)
Total	42 (100%)	21 (100%)	10 (100%)	73 (100%)

Table 5: Monolingual versus bilingual dictionaries given as 1st, 2nd, and 3rd choice by pre-intermediate students

As shown in Table 5 above, we can see that more pre-intermediate students choose bilingual rather than monolingual dictionaries from the three possibilities. However, the result shows no significant difference between the prefer-

ences for two dictionary types reported by pre-intermediate students. In other words, we can say that the pre-intermediate students did not favor one dictionary type over the other.

On the other hand, from the total of 72 valid responses by the intermediate students, 39 intermediate students used at least one dictionary. 24 intermediate students reported using two dictionaries, while 9 students claimed to use three dictionaries. In Table 6, we present the breakdown of responses across choices, as well as details of dictionaries mentioned by intermediate students.

No.	Dictionary Code	1st Choice	2nd Choice	3rd Choice	Total	Percent
1.	OALD	13	3	1	17	22%
2.	Google Translate	7	4	3	14	18%
3.	<i>Kamus Indonesia-Inggris/Kamus Inggris-Indonesia</i> (Echols and Shadily 2014/2017)	7	4	1	12	16%
4.	<i>Kamusku</i>	4	5	0	9	12%
5.	<i>Oxford Mini Dictionary & Thesaurus</i>	2	2	0	4	5%
6.	<i>Kamus Indonesia-Inggris/ Inggris-Indonesia</i>	2	1	0	3	4%
7.	English Thesaurus	2	0	0	2	3%
8.	U-Dictionary	1	0	1	2	3%
9.	Babla	0	1	0	1	1%
10.	<i>Kamus Bilingual 600 Miliyar Kata</i>	1	0	0	1	1%
11.	British Dictionary	0	1	0	1	1%
12.	COBUILD	0	1	0	1	1%
13.	Electronic Dictionary (bilingual)	0	1	0	1	1%
14.	LDCE	0	0	1	1	1%
15.	MEDAL	0	0	1	1	1%
16.	Merriam Webster	0	1	0	1	1%
17.	Urbandictionary.com	0	0	1	1	1%
	Total	39	24	9	72	100%

Table 6: Dictionaries most often used by intermediate students

As shown in Table 6 above, the OALD remains the most frequent dictionary used by intermediate students. Of the total 17 responses, 13 students reported choosing the OALD as their first choice dictionary. The two front runner dictionaries are Google Translate and *Kamus Indonesia-Inggris/Kamus Inggris-Indo-*

nesia (Echols and Shadily 2014/2017). *Kamusku* and *Oxford Mini Dictionary & Thesaurus* are the fourth and fifth most consulted dictionaries. In Table 7, readers can find the breakdown of students' preferences based on the classification of dictionary types, which are monolingual and bilingual dictionaries.

Dictionary types	1st choice	2nd choice	3rd choice	Total
Bilingual	22 (56.4%)	16 (66.7%)	5 (55.6%)	43 (60%)
Monolingual	17 (43.6%)	8 (33.3%)	4 (44.4%)	29 (40%)
Total	39 (100%)	24 (100%)	9 (100%)	72 (100%)

Table 7: Monolingual versus bilingual dictionaries given as 1st, 2nd, and 3rd choice by intermediate students

As shown in Table 7 above, intermediate students prefer to use bilingual dictionaries to monolingual ones. However, the statistical test has shown that there is no significant difference in the consultation frequency of consulting bilingual and monolingual dictionaries. In other words, we can say that among intermediate students they did not seem to prefer one type to the other.

An analysis of choices for individual dictionary types above suggests that bilingual dictionaries are consulted more often and are preferred by all of the students as their first choice. However, the effect of students' levels and dictionary choice need to be examined further. From the choices made in the test result, it was found that the effect of dictionary types was not significant, so the effect holds irrespective of the students' levels. From this fact, it was suggested the two groups of students did not favor one type of dictionary over the other. There also appears to be a tendency for the students to use one specific dictionary, that is Oxford.

4.3 Dictionary evaluation

In section B of the user questionnaire, the students were instructed to rate the dictionaries they have named before. The ratings were given in ranking order. The ratings are classified as 'excellent', 'good', 'OK', 'poor', and 'awful'. A detailed number of responses are separately presented in tables for each level of pre-intermediate (see Table 8) and intermediate levels (see Table 10). In Table 8, we collected 68 valid ratings identified by pre-intermediate students.

No.	Dictionary Code	Mean Rating	Rating					n
			Excellent	Good	OK	Poor	Awful	
1.	OALD	4.24	6	9	2	0	0	17
2.	Google Translate	3.54	0	9	2	2	0	13
3.	<i>Kamusku</i>	3.56	1	4	3	1	0	9
4.	<i>Kamus Indonesia-Inggris/Inggris-Indonesia</i>	4.00	0	9	0	0	0	9
5.	<i>Kamus Indonesia-Inggris/Kamus Inggris-Indonesia</i> (Echols and Shadily 2014/2017)	4.33	2	4	0	0	0	6
6.	Line Dictionary	3.00	0	1	0	1	0	2
7.	LDCE	4.00	0	2	0	0	0	2
8.	CALD	4.50	1	1	0	0	0	2
9.	Merriam Webster	4.50	1	1	0	0	0	2
10.	Dictionary.com	3.00	0	0	1	0	0	1
11.	English Dictionary	3.00	0	0	1	0	0	1
12.	COBUILD	4.00	0	1	0	0	0	1
13.	Flitto	4.00	0	1	0	0	0	1
14.	U-Dictionary	4.00	0	1	0	0	0	1
15.	<i>Kamus Password English Learner's Dictionary for Speakers of Bahasa Indonesia</i>	5.00	1	0	0	0	0	1
16.	Total		12	43	9	4	0	68
17.	Mean rating	3.56						

Table 8: Mean ratings and breakdown of rating responses by dictionary category chosen by pre-intermediate students

As shown in Table 8, the five top rows from the collected data give information about the most consulted dictionary. The mean rating of each dictionary is also given in the table. The smaller the number of dictionaries rated by pre-intermediate students, the less reliable the ratings were. However, the identification of the least consulted dictionary is still given for completeness.

From the facts collected and presented in Table 8 above, 17 pre-intermediate students rated the OALD dictionary as 'excellent' and 'good'. The average rating of the OALD is 4.24. This fact complements our previous finding that the OALD is not only preferred, but also highly valued by pre-intermediate students. From the total of 13 pre-intermediate students, nine of them rated Google Translate as 'good'. Google Translate obtained a mean rating of 3.54. The mean ratings received by *Kamusku* and *Kamus Indonesia-Inggris/Inggris-Indonesia* are 3.56

and 4.00 respectively. *Kamus Indonesia–Inggris/Kamus Inggris–Indonesia* (Echols and Shadily 2014/2017) could only be identified by fewer pre-intermediate students and received an average score of 4.33, which is surprisingly higher than the OALD. However, the OALD's rating is considered to be more reliable than that of *Kamus Indonesia–Inggris/Kamus Inggris–Indonesia* (Echols and Shadily 2014/2017). The reason for this is that the less variation a dictionary has, the lower the mean error value of it will be.

There is one interesting case in which the semi-bilingual dictionary (i.e. *Kamus Password English Learner's Dictionary for Speakers of Bahasa Indonesia*) (Kwary 2011) was identified by one of the pre-intermediate students. This finding is quite surprising, as we can see in Yanto's (2016) study that the dictionary was hardly mentioned by Indonesian lecturers in the English department. As there is only one student who could identify the dictionary, this finding is unreliable in which the mean error value could be the highest.

All the dictionaries rated in Table 8 above can be classified into either bilingual or monolingual dictionaries. A summary of the results is given in Table 9 below. The table presents information about the mean ratings, standard errors, and 95% confidence intervals for both bilingual and monolingual dictionaries.

Dictionary types	Mean Rating	Std. Error	-95%	+95%	n
Bilingual	3.79	1.46	3.12	4.45	42
Monolingual	4.15	1.35	3.55	4.76	26

Table 9: Pre-intermediate students' ratings of bilingual and monolingual dictionaries

As shown in Table 9 above, although pre-intermediate students claimed to use bilingual dictionaries more than monolingual ones (see Section 4.2), it emerged that monolingual dictionaries received higher ratings than bilingual ones. The statistical result suggests that there was no significant effect of dictionary types on their ratings. The data somewhat confirms the fact that pre-intermediate students not only use both bilingual and monolingual dictionaries but also value both dictionaries.

Let us examine the ratings given by intermediate students. There were 67 valid responses in the students' evaluation. Table 10 below displays the mean ratings of dictionaries identified by the intermediate students.

No.	Dictionary Code	Mean Rating	Rating					n
			Excellent	Good	OK	Poor	Awful	
1.	OALD	4.31	6	9	1	0	0	16
2.	Google Translate	3.20	1	8	3	2	0	14
3.	<i>Kamus Indonesia–Inggris/ Kamus Inggris–Indonesia (Echols and Shadily 2014/2017)</i>	4.30	4	8	0	0	0	12
4.	<i>Kamusku</i>	2.70	2	4	2	1	0	9
5.	<i>Oxford Mini Dictionary & Thesaurus</i>	4.25	1	3	0	0	0	4
6.	U-Dictionary	4.00	0	2	0	0	0	2
7.	English Thesaurus	4.50	1	1	0	0	0	2
8.	<i>Kamus Bilingual 600 Milyar Kata</i>	3.00	0	0	1	0	0	1
9.	LDCE	3.00	0	0	1	0	0	1
10.	Babla	4.00	0	1	0	0	0	1
11.	Urbandictionary.com	4.00	0	1	0	0	0	1
12.	COBUILD	5.00	1	0	0	0	0	1
13.	MEDAL	5.00	1	0	0	0	0	1
14.	Merriam Webster	5.00	1	0	0	0	0	1
15.	Kamus Aplikasi (Bilingual)	5.00	1	0	0	0	0	1
16.	Total		19	37	8	3	0	67
17.	Mean rating	4.34						

Table 10: Mean ratings and breakdown of rating responses by dictionary category chosen by intermediate students

As shown in Table 10, the OALD is mostly consulted and highly rated by intermediate students. The dictionary was also rated as 'excellent' and 'good' and receives an average rating of 4.31. Google Translate, which is in the second place as the most consulted dictionary, received a mean score of 3.20. *Kamus Indonesia–Inggris/Kamus Inggris–Indonesia* (Echols and Shadily 2014/2017) and *Kamusku* obtained average rating scores of 4.30 and 2.70 respectively.

There is an interesting finding that can be seen in the breakdown of responses given in Table 10. We can see that the five top lists start with a monolingual dictionary by a specific publisher (i.e. *Oxford Advanced Learner's Dictionary*) and end with a smaller dictionary by the same publisher (i.e. *Oxford Mini Dictionary & Thesaurus*). Four intermediate students rated *Oxford Mini Dictionary & Thesaurus* quite highly, with an average rating of 4.25. This dictionary was not even mentioned by the pre-intermediate students. Although the number is small, this provides more information about the students' awareness of different kinds of dictionaries.

A further analysis has to be done to see whether there is a significant effect of dictionary types on the ratings. Table 11 gives detailed information about mean

ratings, standard errors, and 95% confidence intervals for each dictionary type.

Dictionary type	Mean Rating	Std. Error	-95%	+95%	n
Bilingual	3.88	1.40	3.28	4.48	42
Monolingual	4.34	1.31	3.84	4.85	29

Table 11: Intermediate students' ratings of bilingual and monolingual dictionaries

As shown in Table 11, intermediate students valued monolingual dictionaries more than bilingual dictionaries. However, there is no significant effect of dictionary types on the evaluation. We can assume that the two groups evaluate both dictionaries similarly.

5. Reference needs

In section C of the user questionnaire, the students were asked to report how often they consult their dictionaries to look for particular information. We considered including seven main types of dictionary information in the questionnaire. They are meaning, pronunciation, word class, examples, phrases, context, and synonyms. The options of consultation frequency were made to elicit relative frequency.

Types of Information		Levels									
		Pre-intermediate					Intermediate				
		Al	Of	So	Se	Ne	Al	Of	So	Se	Ne
C1: Mng	n	27	11	8	0	0	17	19	4	1	1
	%	58.7%	24%	17.4%	0%	0%	40.5%	45.2%	9.5%	2.4%	2.4%
C2: Pron	n	4	15	18	8	1	6	8	16	11	1
	%	8.7%	32.6%	39%	17.4%	2.2%	14.3%	19.0%	38.1%	26.2%	2.4%
C3: WdCl	n	2	17	14	8	5	4	7	14	15	2
	%	4.3%	37%	30.4%	17.4%	11%	9.5%	16.7%	33.3%	36%	4.8%
C4: Exm	n	4	6	20	13	3	5	11	16	6	4
	%	8.7%	13%	43.5%	28.3%	6.5%	12%	26.2%	38.1%	14.3%	9.5%
C5: Phrs	n	2	4	8	22	10	1	3	26	9	2
	%	4.3%	8.7%	17.4%	47.8%	22%	2.4%	7.3%	63%	22%	4.9%
C6: Con	n	5	8	18	10	5	1	6	12	17	6
	%	10.9%	17.4%	39%	22%	11%	2.4%	14.3%	28.6%	40.5%	14.3%
C7: Syn	n	5	8	18	10	5	6	16	18	2	0
	%	11%	17.4%	39%	22%	11%	14.3%	38%	43%	4.8%	0%
C8: Usage notes	n	0	0	0	0	0	0	1	0	0	0

Table 12: The breakdown of students' responses for reference needs by pre-intermediate and intermediate students

As shown in Table 12, the most consulted information when the students use dictionaries is word meaning. The concept of meaning in the questionnaire referred to English and Indonesian equivalents, as well as an English definition. For this item, we considered that it was not necessary to mention the concept of word meanings separately. Indeed, more than 80% of the students reported to 'always' or 'often' look for meaning. A breakdown of responses for each level is also given here: 27 pre-intermediate learners (58.7%) reported that they 'always' look for meaning in a dictionary and 11 of them (24%) claimed they 'often' look for meaning. In contrast, 17 intermediate students (40.5%) reported that they 'always' look for meaning in a dictionary, while more intermediate students (45.2%) chose the 'often' option. Fewer intermediate students (4%) claimed they 'seldom' or 'never' consult a dictionary for meaning, while none of the pre-intermediate students chose the last two options ('seldom' and 'never'). It is interesting to see that students with a higher level of English proficiency rely on the consultation of meaning in a dictionary, while pre-intermediate students seemed to be confident enough of their vocabulary knowledge to consult dictionaries for meaning.

In item C2 of the questionnaire, around 40% of both pre-intermediate and intermediate students indicated that they need to consult dictionaries to pronounce words correctly. This also means that more than 40% of both pre-intermediate and intermediate students might be confident enough not to consult their dictionaries for the way to pronounce English words. More reasons possibly emerge from the students. In item C3, the students were asked to report how often they consult their dictionaries to find word class information. It is suggested that there was a tendency for pre-intermediate students to choose 'always' and 'often' options. In contrast, intermediate students tend to choose the two lower frequency options, 'seldom' and 'never'. We can thus assume that the frequency of word-class consultation in dictionaries may decrease as students' proficiency levels rise.

In item C4, pre-intermediate students reported lower consultation (>50% of all responses) of example information in dictionary entries. In contrast, more intermediate students reported consulting examples more often. In item C5 and item C6, the students were instructed to report their needs for phrases and context information in dictionaries. The responses from both pre-intermediate and intermediate students have shown that only a few students indicated that they 'always' or 'often' look for phrases and context in their dictionaries. For the last item C7, the students would indicate their consultation frequency of synonym information. It was found that more students in the pre-intermediate level chose the lower frequency options, while more students in the intermediate level chose the higher frequency options. The higher consultation for synonyms can be seen as additional information on word meaning, in addition to the fact that information on meaning is already given in the English definition or in the translation equivalents.

The comparison in the consultation frequency of dictionary information between the two groups of students is presented in Table 13 below. The

responses of the students' reference needs are then ranked by mean scores of the consultation frequency.

Rank	Pre-intermediate Level			Rank	Intermediate Level		
	Information Type	Mean	Std. Dev.		Information Type	Mean	Std. Dev.
1.	C1: Meaning	4.35	0.99	1.	C1: Meaning	4.19	0.89
2.	C2: Pronunciation	3.28	0.93	2.	C7: Synonym	3.62	0.79
3.	C3: Word class	3.07	1.08	3.	C2: Pronunciation	3.17	1.06
4.	C7: Synonym	2.96	1.13	4.	C4: Examples	3.17	1.12
5.	C4: Examples	2.89	1.02	5.	C3: Word class	2.90	1.05
6.	C5: Phrases	2.72	1.18	6.	C5: Phrases	2.74	0.86
7.	C6: Context	2.26	1.04	7.	C6: Context	2.50	0.99

Table 13: Ranking of the seven reference needs by mean consultation frequency

We further examine the relationship between information types and learners' levels, whether there was a significant effect on the frequency in which information types were consulted. Indeed, the type of information most frequently searched by pre-intermediate and intermediate students is meaning, with mean scores of 4.35 and 4.19 respectively for the two groups. This is followed by pronunciation (3.28) and word class (3.07) for the pre-intermediate students. In contrast, synonyms are the second most consulted information by intermediate students (3.62), and pronunciation follows with a mean score of 3.17. The test results have shown that there was an increasing need for synonyms, but not for the remaining types of information.

6. Conclusion

User studies, which are based on the user questionnaire survey, can answer a number of questions that are relevant to the greatest part of practical lexicography. However, for a survey to be maximally useful, researchers need to be really careful about the exact form of the question. They need to tackle the specific questions so that they can present useful information to readers.

The results of this study confirm that both pre-intermediate and intermediate students use bilingual dictionaries more often than monolingual dictionaries. The two groups further show indifference when asked to identify dictionary choices and ratings. Moreover, the attempt to bring the lexicographic information to the students proved useful to some extent. It is interesting to see that the students are able to identify language difficulties and use dictionaries to look for particular information.

Although this study is capable of providing the necessary information about the profile of Indonesian speakers as dictionary users, a questionnaire-

based study has its own shortcomings for not being able to explain the reasons for users' choices. For this reason, adopting a multi-method or mixed approach would be useful in order to better explain the manner of consultation behaviour in this Indonesian context for the future.

References

Dictionaries

Babla. Online Dictionaries by bab.la — Loving Languages. Available at:

<https://en.bab.la/>

(CALD) **McIntosh, C. (Ed.)**. 2013. *Cambridge Advanced Learner's Dictionary*. 4th Edition. Cambridge: Cambridge University Press.

(COBUILD) **Harper Collins — UK**. 2016. *Collins Cobuild Advanced Dictionary of English*. Jakarta: Gramedia Pustaka Utama.

Dictionary.com. Available at:

www.dictionary.com

Echols, J.M. and H. Shadily (Eds.). 2014. *Kamus Indonesia–Inggris*. Edisi Ketiga Yang Diperbarui. Jakarta: Gramedia Pustaka Utama.

Echols, J.M. and H. Shadily (Eds.). 2014. *Kamus Inggris–Indonesia*. Edisi Ketiga Yang Diperbarui. Jakarta: Gramedia Pustaka Utama.

Flitto. For All Your Translation Needs. Available at:

<https://www.flitto.com/>

Kwary, D.A. (Ed.). 2011. *Kamus Password English Learner's Dictionary for Speakers of Bahasa Indonesia*. Edisi Ketiga + CD ROM. Third edition. Jakarta: Kesaint Blanc.

(LDCE) **Delacroix, L. (Ed.)**. 2014. *Longman Dictionary of Contemporary English*. 2014. 6th Edition. London: Pearson Education ESL.

Livingstone, C. (Ed.). 2011. *Oxford Mini Dictionary and Thesaurus*. Oxford: Oxford University Press.

(MEDAL) **Rundell, M. (Ed.)**. 2002. *Macmillan English Dictionary for Advanced Learners*. Second Edition. London: Macmillan Education.

(OALD) **Deuter, M. et al. (Eds.)**. 2015. *Oxford Advanced Learner's Dictionary*. 9th Edition. International Student's Edition. Oxford: Oxford University Press.

U-Dictionary. Available at:

<http://www.u-dictionary.com/home/>

Urbandictionary.com. Available at:

<https://www.urbandictionary.com/>

Other

Amalia, D. 2014. *Formulasi Pendefinisian dan Model Pengentrian Verba dalam Kamus Pemelajar Bahasa Indonesia*. Depok: Universitas Indonesia.

Atkins, B.T.S. and K. Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 21-82. Lexico-

graphica. Series Maior 88. Tübingen: Max Niemeyer.

- Badan Bahasa Upayakan Leksikografi Berkembang di Indonesia.** 2017. Retrieved 25 November 2017, from:
<http://badanbahasa.kemdikbud.go.id/lamanbahasa/berita/2408>
- Baxter, J.** 1980. The Dictionary and Vocabulary Behavior: A Single Word or a Handful? *TESOL Quarterly* 14(3): 325-336.
- Béjoint, H.** 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.
- Bensoussan, M., D. Sim and R. Weiss.** 1984. The Effect of Dictionary Usage on EFL Test Performance Compared with Student and Teacher Attitudes and Expectations. *Reading in a Foreign Language* 2(2): 262-276.
- Campoy-Cubillo, M.C.** 2015. Assessing Dictionary Skills. *Lexicography. Journal of ASIALEX* 2(1): 119-141.
- Frankenberg-Garcia, A.** 2015. Dictionaries and Encoding Examples to Support Language Production. *International Journal of Lexicography* 28(4): 490-512.
- Hartmann, R.R.K.** 2000. European Dictionary Culture. The Exeter Case Study of Dictionary Use among University Students, against the Wider Context of the Reports and Recommendations of the Thematic Network Project in the Area of Languages (1996-1999). Heid, U. et al. (Eds.). 2000. *Proceedings of the Ninth Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000. Vol. 1:* 385-391. Stuttgart: Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.
- Hatherall, G.** 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983:* 183-189. Tübingen: Max Niemeyer.
- Kwary, D.A.** 2018. The Variables for Drawing Up the Profile of Dictionary Users. *Lexicography. Journal of ASIALEX* 4(2): 105-118. Heidelberg: Springer-Verlag GmbH.
- Lew, R.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English.* Poznań, Poland: Motivex.
- Moeljadi, D., I. Kamajaya and D. Amalia.** 2017. Building the Kamus Besar Bahasa Indonesia (KBBI) Database and Its Applications. Xu, Hai (Ed.). 2017. *Proceedings of the 11th International Conference of the Asian Association for Lexicography, ASIALEX 2017, 10-12 June 2017, Guangzhou, China: Lexicography in Asia: Challenges, Innovations and Prospects:* 64-80. Guangzhou, China: ASIALEX.
- Müller-Spitzer, C., M.J. Domínguez Vázquez, M. Nied Curcio, I.M. Silva Dias and S. Wolfer.** 2018. Correct Hypotheses and Careful Reading Are Essential: Results of an Observational Study on Learners Using Online Language Resources. *Lexikos* 28: 287-315.
- Nesi, H.** 2014. Dictionary Use by English Language Learners. *Language Teaching* 47(1): 38-55.
- Tomaszczyk, J.** 1979. Dictionaries: Users and Uses. *Glottodidactica* 12: 103-119.
- Yanto, E.S.** 2016. EFL Teachers' Belief, Knowledge and Practices in Using Dictionaries. *Journal of ELT Research* 1(1): 53-72. Retrieved from:
<http://journal.uhamka.ac.id/index.php/jer/article/view/30>

Appendix 1: Learner's Questionnaire on Dictionary Use

Student identification number:

Instructions: answer truthfully all the questions by crossing (like this) exactly one box in each row. You have 15 minutes for this part.

A. How often do you use the following types of dictionaries?

	6-7 days a week	4-5 days a week	2-3 days a week	0-1 day a week
1. Indonesian–English (bilingual) _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. English–Indonesian (bilingual) _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. English–English (monolingual) _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

B. What dictionaries do you use **most often** (give the title, type, medium, publisher, author, edition — as far as you remember). Several dictionary types and titles have also been provided on the next page. And how do you rate them? (**Answer 1 until 3 dictionaries**)

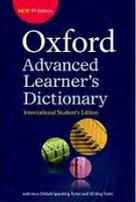
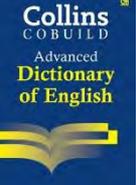
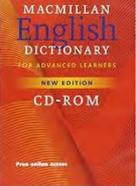
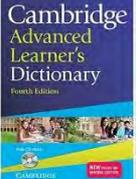
Title/publisher/author/edition (type; medium) e.g., Merriam Webster (<i>mono / bilingual; print / electronic (hand-held dictionary / CD-ROM) / online (web-version / dictionary app)</i>)	excellent	good	OK	poor	awful
1.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

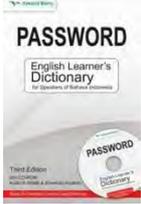
C. How often do you look for the following information in a dictionary:

	always	often	sometimes	seldom	never
1. The meaning (definitions or equivalents) _____	<input type="checkbox"/>				
2. Pronunciation (/ˈædvəkət/) _____	<input type="checkbox"/>				
3. Word class (<i>N, V, etc.</i>) _____	<input type="checkbox"/>				
4. Examples (<i>She had the advantage of...</i>) _____	<input type="checkbox"/>				
5. Phrases (big/great advantage) _____	<input type="checkbox"/>				
6. Context (<i>Law; Formal; Offensive; etc.</i>) _____	<input type="checkbox"/>				
7. Synonyms (syn beneficial) _____	<input type="checkbox"/>				
8. Other (mention): _____	<input type="checkbox"/>				

Instruction: when you are done with this page (double-check that you've answered all the items), turn the page.

Appendix 2: Dictionary Identification

	<p>John M. Echols and Hasan Shadily (Eds.). 2014. <i>Kamus Indonesia–Inggris</i>. Edisi Ketiga Yang Diperbarui. Jakarta: Gramedia Pustaka Utama.</p> <p>John M. Echols and Hasan Shadily (Eds.). 2014. <i>Kamus Inggris–Indonesia</i>. Edisi Ketiga Yang Diperbarui. Jakarta: Gramedia Pustaka Utama.</p>
	<p>Deuter, M. et al. (Eds.). 2015. <i>Oxford Advanced Learner's Dictionary (OALD)</i>. 9th Edition: International Student's Edition. Oxford: Oxford University Press</p>
	<p>Harper Collins — UK. 2016. <i>Collins Cobuild Advanced Dictionary of English (COBUILD)</i>. Jakarta: Gramedia Pustaka Utama.</p>
	<p>Rundell, M. (Ed.). 2002. <i>Macmillan English Dictionary for Advanced Learners (MEDAL)</i>. Second Edition. London: Macmillan Education.</p>
	<p>McIntosh, C. (Ed.). 2013. <i>Cambridge Advanced Learner's Dictionary with CD-ROM (CALD)</i>. 4th Edition. Cambridge: Cambridge University Press.</p>
	<p>Delacroix, L. (Ed.). 2014. <i>Longman Dictionary of Contemporary English. (LDCE)</i>. 6th Edition. London: Pearson Education ESL.</p>

	<p>Kwary, D.A. (Ed.). 2011. <i>Kamus Password English Learner's Dictionary for Speakers of Bahasa Indonesia</i>. Edisi Ketiga + CD ROM. Third Edition. Jakarta: Kesaint Blanc.</p>
. . . .	Others (specify) . . .

John Considine (Editor). *The Cambridge World History of Lexicography*. 2019, xii + 961 pp. ISBN 978-1-107-17886-1. Cambridge: Cambridge University Press. Price: \$155.00.

A dictionary is the authentic record of the relevant language as well as its cultural world. For most languages in the world, the history of the dictionary is almost as long as that of its language and culture. This makes the history of dictionaries a worthy subject of inquiry and systematic research in the field of lexicography. Hartmann and James (1998: 85) state that lexicography 'has two basic divisions: lexicographic practice, or DICTIONARY-MAKING, and lexicographic theory, or DICTIONARY RESEARCH', and that dictionary research is concerned 'with the academic study of such topics as the nature, history, criticism, typology and use of DICTIONARIES and other REFERENCE WORKS' (ibid: 43). Quite a number of volumes on lexicography have been published worldwide to present the history of lexicography. However, most, if not all, of them focus on the lexicography of one single language, as Yong and Peng (2008) on Chinese dictionaries, and Cowie (2009) on English dictionaries, or on one specific type of dictionaries, as Cowie (2002) on English dictionaries for foreign learners.

The Cambridge World History of Lexicography (CWHL) is regarded as the first survey of the global history of lexicography, tracing the historical development of the dictionaries of different types and of various languages from the global perspective. This volume is among a series of academic books on world history produced by Cambridge University Press. Its editor John Considine, Professor of English at the University of Alberta, Canada, has authored, edited or co-edited several books on lexicography, and has been contributing to the *Oxford English Dictionary* for the last thirty years as a library researcher, an assistant editor, and a consultant. *CWHL* is a joint work by 33 leading scholars in lexicography from 15 countries.

With a total of 973 pages, *CWHL* comprises 32 chapters in four parts, plus an introduction, two appendices, two resources and one index. The four parts are arranged chronologically from the ancient world of five thousand years ago to the modern world, and the chapters in each part are further sequenced from the linguistic, geographical, or cultural perspective. This volume is 'the first survey of all the dictionaries which humans have made, from the ancient civilizations of Mesopotamia, Egypt, China, India, and the Greco-Roman world, to the contemporary speech communities of every inhabited continent' (p. i).

Part One 'The Ancient World' explores, in five chapters, the lexicographical traditions of ancient Mesopotamia, ancient Egypt, ancient China, ancient India, and the Greco-Roman world. This part depicts a global map of the ancient lexicography, starting from Mesopotamian lexicography, which originated in cuneiform and has yielded the world's oldest extant wordlists, to ancient Egyptian lexicography, which has the longest of all dictionary traditions in the world. The map then rolls eastward to ancient Chinese lexicography, which has produced the greatest variety of early types of dictionaries

rooted in the philological heritage, and then to ancient Indian lexicography originating from the Sanskrit language and the Brahmanical tradition. The map finally stops at Greco-Roman lexicography, which developed from the great flourishing of Alexandrian scholarship and started by compiling alphabetical lists of lemmata based on a corpus of canonical literary texts.

Part Two 'The Pre-modern World' analyses, in nine chapters, world lexicography in the following millennium, and the continuations up to the seventeenth and eighteenth centuries. This part begins with two chapters on the great traditions of Chinese lexicography and Indian/Tibetan lexicography, which continued from the ancient world into this middle period. Indian and Tibetan lexicography during this period share similar cultural features typically in Sanskrit dictionaries, Buddhist texts and synonym treatment. The next four chapters remain focused on Asia, surveying the Arabic and Hebrew traditions of lexicography, those of the Chinese periphery, and those of the Turkic languages and of Persian. Eastern Inner Asia, Japan, and Korea are grouped into the Chinese periphery as the lexicographical activities there were directly influenced by the Chinese tradition in terms of Chinese characters and typical Chinese dictionaries. The last three chapters turn westwards to the lexicography of the Byzantine world, that of Medieval Latin Christendom, and that of early modern Western Europe. Medieval Latin Christendom refers to the lands where Roman Christianity was practised and the Latin language was read, corresponding roughly to what Western and Central Europe is in the modern day.

As the longest part of this volume, Part Three 'The Modern World: Continuing Traditions' surveys, in eleven chapters, the world lexicography of the past two or three centuries from the East to the West to illustrate the mass production of lexicographical works. It starts with the lexicographical tradition in China which has remained highly active until the present day. It is followed by a description of Japanese, Korean, and Vietnamese lexicography from the beginning of the nineteenth century onwards, a period in which these were no longer considered part of 'the Chinese periphery' but had their own vigorous traditions of lexicography. The next two chapters turn westwards and then southwards to the lexicography of the Turkish and Persian languages, and then to that of South Asia. The following two chapters address the two Semitic lexicographical traditions of Arabic and of Modern Hebrew. Then the last five chapters turn to the lexicographical traditions of the languages belonging to three European language families: one on Slavic and Baltic languages, three on Germanic languages, and one on Romance languages. The three chapters on Germanic comprise one on the Germanic languages other than English, one on standard varieties of English, and one on regional varieties.

Part Four 'The Modern World: Missionary and Subsequent Traditions' traces, in seven chapters, the missionary traditions of lexicography across the modern world: South America, Mesoamerica, North America, East Asia, India and Indonesia, Africa, and Australia. These cultural areas cover South America, North America, Asia, Africa, and Oceania in terms of continents. Only Europe

is excluded, as it functions as the epicentre of the world Christian mission as well as the missionary lexicography.

In the back matter are two appendices: 'the Language Varieties' and 'the Lexicographers'. The first appendix takes a brief account, in 32 pages, of each language, language group and language family of which the lexicography is discussed in this volume. These entries are arranged alphabetically, and each comprises very basic taxonomic information and a note on the lexicographical record. The second gives, in 68 pages, a brief description of each lexicographer whose work is discussed in this volume. Each of these entries, also alphabetically listed, consists of name, date and place of birth and death, principal occupation other than 'lexicographer', and principal contribution to lexicography. This part is deemed as the first biographical survey of lexicographers on a global scale.

CWHL is distinguished from the other volumes of similar topics in at least the following three aspects. Firstly, *CWHL* offers the first comprehensive account of the full history of world lexicography spanning five thousand years and covering three hundred languages. Writing up a volume to cover the comprehensive history of global lexicography is challenging for at least two reasons: one is that 'there have been so many lexicographical traditions in the world over the past five thousand years [...] that their individual traditions are almost ungraspable by a single historian' (p. 1); the other is that 'the concept of "lexicography" is somewhat elusive' (p. 2). The elusive concept of 'lexicography' is actually due to the definition of 'dictionary'. For a long time, the borderline between 'dictionary' and 'encyclopaedia' was vague, and so was that between 'dictionaries of synonyms' and 'thesauruses'. Therefore, 'lexicography' in *CWHL* is defined as 'the making of lists of words and their equivalents or interpretations' (ibid). *CWHL* is the only single volume ever to present the historical description of world lexicography (cf. Considine 2016).

Secondly, it delivers accessible and insightful contributions mainly from the perspective of Eurocentrism. The 33 contributors are mostly from Europe or North America, somewhat indicating a Eurocentristic sense. This is not unusual, as Europe has long been the core of human civilization, dictionary compilation, and lexicographical research worldwide. Accordingly, missionary lexicography plays a vital role in the history of global lexicography, which explains why Part Four is devoted to the missionary traditions of lexicography since the sixteenth century in various cultural areas other than Europe. As Hovdhaugen (1996: 7) observes, 'a satisfactory history of linguistics cannot be written before the impressive contribution of missionaries is recognised'. Asia is a good instance. The efforts of the Jesuits in the fields of linguistics and lexicography have resulted in a number of important dictionaries, grammars, vocabularies, and teaching manuals in Asian countries. Products planned are searchable on-line texts of the European dictionaries used by missionary lexicographers and of the bilingual and multilingual dictionaries which they produced, and a linguistic overview of the grammars and vocabularies published

by the Jesuit mission in Japan and in India.' (Zwartjes 2012: 195-196) A number of pioneering bilingual dictionaries of Chinese were compiled during the sixteenth to nineteenth centuries by European and American missionaries such as Michele Ruggieri (1543–1607), Matteo Ricci (1552–1610), Robert Morrison (1782–1834), and Samuel Wells Williams (1812–1884). These missionary scholars were involved in Chinese lexicography with their vigorous promotion in the Latinization of the Chinese writing system, as well as in the orthographic transferability across dialects of the Chinese language.

Finally, this volume provides a global history of dictionaries both chronologically and synchronically, showing a general picture, and presenting case studies, of how world lexicography has developed in various areas. *CWHL* 'seeks to tell a story' (p. 2) on global lexicography of five thousand years and three hundred languages in a time sequence from the ancient world to the pre-modern world and modern world. This volume can be read 'to follow one thread in the whole story from period to period, looking, for instance, at the three chronologically sequenced chapters on Chinese lexicography as a series' (p. 2). The story is also told in a synchronic way to explore the lexicographical status of a specific language or at a specific period. Readers are offered the alternative to 'read synchronically, across a given part, to get a picture of the lexicographical activity which was taking place at a particular time' (p. 3). In *CWHL*, Chinese lexicography, with its various historical stages, is carefully reviewed and illuminated, although there are no joint contributors from China. As Considine points out, 'the language to which the most chapters are devoted is Chinese' (p. 2), although English is the language in which this volume is written and currently the most important academic language globally. In Appendix 1 'The Language Varieties', the occurrence of 'China/Chinese' and 'England/English' is respectively 11/38 and 3/24; in Appendix 2 'The Lexicographers', 'China/Chinese' and 'England/English' are respectively counted 82/62 and 12/123 times; and in 'Index', 'China/Chinese' and 'England/English' are respectively 2/41 and 2/44. Chinese lexicography has a very long and continuous history, and its magnificent dictionaries are worthy of further research in the Western world. 'Dictionary-making in China, for example, was already well advanced some two thousand years ago, but the extent of this achievement was not appreciated by the West for more than a thousand years.' (Collison 1982: 20)

Admittedly, there are still some possibilities for the improvement of the whole volume. First, there seems not to be an obvious and consistent structural framework for each chapter to observe. The whole volume follows the principle that 'different chapters are handled in different ways' (p. 3), thus it is more readable than searchable. As a result, it is hard to make a comparison between different areas or periods in terms of lexicographical status. Second, little emphasis is placed in this volume on the lexicographical status in the contemporary world. The whole history is divided into three periods: ancient, pre-modern, and modern, and the modern period is from the eighteenth or nineteenth century onwards. However, in the late twentieth and early twenty-first

centuries, a substantial number of dictionaries of various types have been produced in different parts of the world; thus it would be better for the contemporary period to be separated from the modern period so as to describe the current flourishing lexicography. Moreover, in this volume there is not a systematic description of lexicographical study in its various stages. Although this volume is designed to provide 'a history of lexicography, not of dictionaries' (p. 3), it touches on theoretical lexicography very lightly. In Appendix 2 'The Lexicographers', there is no 'Zgusta', a great figure in theoretical lexicography, as *Manual of Lexicography* (Zgusta 1971) is honoured as a milestone to initiate the systematic study of lexicography. 'Lexicographer' is defined as 'one who engages in lexicography, either as a compiler or as a metalexicographer' (Hartmann and James 1998: 84), and 'metalexicographer' is 'one who engages in the theory of lexicography' (ibid: 93).

Overall, despite the minor criticisms raised above, *CWHL* deserves recognition as the first volume on a comprehensive history of world lexicography, and as a helpful resource to the international community of lexicographers and linguists, and to the wider users of dictionaries of various types. It fulfils the function as a rich source of historic information on global lexicography. This volume takes readers on a linguistic journey into the dictionaries of the past five thousand years, into the literal worlds of the people who used or observed those dictionaries, and into the enormous variety of the global cultures of humankind. *CWHL* has established a high-quality precedent for a comprehensive historical assessment of global lexicography.

Acknowledgments

We are grateful to Professor Hans C. Boas of the Linguistics Research Centre, the University of Texas at Austin, Texas, USA, for his encouragement and support.

References

- Collison, R.L. 1982. *A History of Foreign-Language Dictionaries*. London: André Deutsch.
- Considine, J. 2016. A Chronology of Major Events in the History of Lexicography. Durkin, P. (Ed.). 2016. *The Oxford Handbook of Lexicography*: 605-615. Oxford: Oxford University Press.
- Cowie, A.P. 2002. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press.
- Cowie, A.P. (Ed.). 2009. *The Oxford History of English Lexicography*. Oxford: Oxford University Press.
- Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hovdhaugen, E. 1996. Missionary Grammars: An Attempt at Defining a Field of Research. Hovdhaugen, E. (Ed.). 1996. *... and the Word was God: Missionary Linguistics and Missionary Grammar*: 9-22. Münster: Nodus Publikationen.
- Yong, H. and J. Peng. 2008. *Chinese Lexicography: A History from 1046 BC to AD 1911*. Oxford: Oxford University Press.

Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Zwartjes, O. 2012. The Historiography of Missionary Linguistics: Present State and Further Research Opportunities. *Historiographia Linguistica* 39(2/3): 185-242. (DOI: <https://doi.org/10.1075/hl.39.2-3.01zwa>)

Xiangming Zhang
School of Foreign Studies
Guangdong University of Finance and Economics
Guangzhou
China
(simon.xmzhang2019@gmail.com; xmzhang@gdufe.edu.cn)

and

Hai Xu
Centre for Linguistics and Applied Linguistics
Guangdong University of Foreign Studies
Guangzhou
China
(Corresponding Author, xuhai1101@gdufs.edu.cn)

María José Domínguez Vázquez, Mónica Mirazo Balsa and Carlos Valcárcel Riveiro (Eds.). *Studies on Multilingual Lexicography*. Lexicographica. Series Maior 157. 2019, vi + 240 pp. ISBN 978-3-11-060467-2, ISSN 0175-9264. Berlin/Boston: Walter de Gruyter. Price: €99.95.

The current globalization and advancement of Information Communication Technology (ICT) have led to more frequent cultural and personal exchanges across the globe than ever before. Therefore, the issue of multilingualism has become prominent. Ruiz de Zarobe and Ruiz de Zarobe (2015) argue that multilingualism is becoming one of the core issues of current communities. Cenoz (2013) points out that "multilingualism nowadays is a very common phenomenon all over the world, ... it's an every reality in cities today". Responding to these new "social situations", this book, *Studies on Multilingual Lexicography*, brings together a series of articles on the development of multilingual e-dictionaries and terminological works influenced by digital technology progress in the era of the Internet. Under the guideline that "the role of the dictionary must necessarily be related to social development and changes" (p. 1) and the core principle offered and exemplified by the volume "that a dictionary should be compiled to satisfy the real needs that the real people experience in real life situations" (p. 30), the primary goal of this volume is to explore ways to meet the needs of new types of dictionary users both from theoretical and practical perspectives. Until now, few works in the field of lexicography have been attempted on multilingual lexicography in the digital age, except volumes edited by Teubert (2007) and Boas (2009). However, Teubert focuses on how to compile multilingual parallel corpora, while Boas concentrates on how to develop multilingual FrameNets-based lexicographical databases. Therefore this is a timely and valuable work for those who want to understand the latest developments in multilingual dictionaries in the digital age.

Beginning with an introduction by the editors, the volume consists of two sections: the first three chapters form section I which focuses on the theoretical perspective of multilingual lexicography concerning the influence of information technology and the emergence of new types of dictionary users; the rest of the chapters comprise section II which centres on practical aspects and is exemplified by current available multilingual electronic dictionaries including various projects and tools.

Fuertes-Olivera and Bergenholtz in chapter 1 offer a new definition of multilingual lexicography in the era of the Internet and illustrate the definition with their accounting dictionaries (p. 27 and p. 28). They argue that the new concept of multilingual lexicography should be based on a "holistic approach", in which three core lexicographical elements (users, lexicographical data and access routes) need to be interconnected. An in-depth review of traditional online multilingual dictionaries (*InterActive Terminology for Europe*, *Logos Dictionary*, and *Diccionari de la Llengua Catalana Multilingüe*) indicates that the traditional concept of multilingual lexicography is erroneous. Problems such as

the lack of conceptualization before compiling, limited data types, and a failure to take a "holistic approach" into consideration are underlined. Reflecting on a business model of publishing houses in the Internet era, the core and specifics of lexicography, they propose a new definition of multilingual lexicography and illustrate it in great length by their accounting dictionaries. Through the new definition, they suggest the thought-provoking concept of a "dynamic dictionary", in which multilingual lexicography should be a flexible and integrated information tool. The online dictionary should be sold as "an on-going service, not as a finished product" (p. 20).

In the following chapter, Gouws deals with the challenges posed for lexicography by the information age in the light of dictionary planning and compiling, and the user's referential skills. With a brief review of "watershed moments" in the English lexicographical history, he emphasizes the importance of the user-directed aim. He further points out the urgency of formulating lexicographical theory in the Internet era and suggests that a sound theory should negotiate the societal dictionary culture, users' needs and users' referential skills. He also highlights the close relationship between lexicography and society and proposes the concept of an "emerging e-society". Then, taking the lexicographical landscape in South Africa as an example, he elaborates on how to establish a comprehensive dictionary culture and how to respond to dictionary users' needs. Additionally, he suggests that the mobile phone could be an ideal lexicographical medium for young learners and illustrates this with a pilot project (MobiLex). Overall, what Gouws emphasizes in this chapter is the situating of the planning and compiling of a dictionary in a larger social context by bringing the new generation of dictionary users into the foreground.

Tarp's chapter entitled "A Dangerous Cocktail: Databases, Information Techniques and Lack of Vision", discusses the challenges to lexicography posed by the current Information Communication Technologies (ICTs) and explains why the combination of new technology and lack of vision could be "a dangerous cocktail". He points out that current lexicography is full of paradoxes and argues that it is lexicographers' lack of vision and reluctance to enhance the opportunities afforded by technology that led to a series of problems. He further explains the disagreement on the disciplinary status and the existence of a possible theory within the circle of lexicography which resulted in the current paradoxes. To address these problems, he explicates how to conduct a dictionary project step by step based on his function theory under the auspices of current available technology. As a strong advocate of lexicographical theory, Tarp offers insightful ideas on how dictionary functions are closely connected with external lexicographical factors such as social situations. Although it is still a controversial issue whether lexicography needs or has a theory, it would be more than welcome if further elaboration on the integration and compatibility can be made between lexicographical theory and theories from other fields such as linguistics and terminology.

Batiukova and De Miguel in chapter 4 present the theoretical project *Multilingual Electronic Dictionary of Motion Verbs* (DICEMTO) and outline its theoretical foundations. The study is motivated by the proliferation of senses of motion verbs in traditional dictionaries, which the authors think is undesirable both in theory and practice. Using the theory of the Generative Lexicon (GL), they believe that a set of lexical-semantic features (Argument Structure, Thematic Structure, Event Structure and Qualia Structure) and a generative mechanism can determine a word's context and its meaning extension. Therefore they argue that a GL-based definition model is not only an important solution for the traditional dictionary definition, but that it can also extend the meaning to real-world knowledge. They use the Spanish "andar" as a meta-entry to instantiate how to improve or optimize the lexicographical meaning representation. They further demonstrate how the design of a dictionary can "reconcile theoretical exhaustiveness and user-friendliness" by presenting two modules (Minimal definition and Lexical entry) to meet users' different needs. Without doubt, the chapter offers an innovative way for dictionary definitions, which will enhance the understanding of the syntactic features of polysemous words for those dictionary users who have a linguistic background. To exploit its full potential, the need to increase its accessibility for those without a linguistic background may still warrant further investigation.

Calvi and López's chapter present two lexicographical works *Linguaturismo* and *Dictionary of Food and Nutrition* (DFN) and demonstrate how the text genre-based approach can be applied to improve the compilation of electronic terminological works in the age of the Internet. As ICT renders both opportunities and challenges for the publishing house, lexicographer and terminologist, the authors stress the importance of collaboration among them. In the case of *Linguaturismo*, they convincingly show the feasibility and advantage of the genre-based approach with the help of a corpus in compiling a bilingual terminology glossary both in theory and practice. The successful use of a genre-based approach in *Linguaturismo* theoretically paved the way for the development of the multilingual DFN. Although particular problems such as lack of equivalence, difficulty in demarcation, and diatopic variation in certain domains may appear in compiling practice, the authors have proved that these can be solved by lexicographical-based approaches. Overall, the authors propose the adoption of the text genre-based approach with the help of a corpus and the integration of a lexicographical and terminological approach could be the solution for user-oriented lexicography. Moreover, the authors suggest this promising methodological approach can in future be further applied to discursive practices of specialists in other fields. In addition, they note the contribution of ICT-enabled multilingual lexicography in the era of the Internet.

To better address translator's needs, Pastor and Durán-Muñoz present *Inteliterm* in the following chapter, which aims at enhancing translation results and minimizing translators' efforts when searching for terminology. Under De Schryver's proposal (2009) and Nielsen's principles of LSP lexicography (2018: 71),

they point out that, although preliminary improvement has been made, the current intelligent dictionaries neglect specific user groups and register-specific language use. To fill the gap, they stress the need to develop an agile and integrated tool which can address professional translators' needs in specialized communication. Thus they propose that the design of *Inteliterm* should have a modular and open structure that promotes users' performance through better access and improved results in translation. Three different modules (term search, a corpus management tool and a terminology management tool) are described to show how effectively and dynamically user-friendliness and customization can be achieved. Furthermore, to prove the functionalities, the authors offer user evaluation and feedback by an empirical experiment. They also note further improvements according to the feedback, and new possibilities worth exploring.

Vázquez and Riveiro introduce the fundamental aspects of PORTLEX, a noun phrase multilingual online dictionary, which is built on valency grammar. One defining feature the authors highlight is that PORTLEX is the result of collaborative work, not only by a team of specialists, professionals and students, but also a semi-collaborative dictionary in which users can add content under supervision of the editing team. The authors also explain why it is collaborative in the light of different stages of development for a time span of 10 years. Moreover, great innovation is made by exploring how to represent noun valency in multilingual lexicography. They further exemplify the new concept of multilingual lexicography by other features such as cross-lingual and collaborative ongoing updating. Additionally, they offer inspiring discussions on how to satisfy different users' needs by developing a custom-made database and by designing adaptive search methods and visualization.

In chapter 6, the authors Durán-Muñoz and Pastor again discuss translators' needs, but this chapter highlights the advantages of the combination of corpora and e-resources by delving into the translators' needs in different translation phases, and offers a thorough overview of current corpus-based multilingual e-resources for translators. Under the proposal of "access to database through targeted search" (Fuertes-Olivera and Nielsen 2012), they examine translators' different needs in three different stages of translation and elucidate how the use of corpora can be important tools for translators for their source and pragmatic information affordances. However, surveys they cite show that few professional translators use corpus tools. To raise translators' awareness of using corpora, a broad overview of the different types of corpus management tools with in-depth analysis is presented, including free web-researchable online corpora, corpus-based web crawlers, web-based translation memories, and e-dictionaries with access to corpora. Moreover, the authors stress the importance of corpora training for translators, and the need to develop corpora of specialized domains to satisfy translators' specialized needs.

In the next chapter, Guinovart and Portela outline the development of a WordNet-based multilingual lexical ontology project for Galician-Galnet and

describe its applications in the fields of terminology acquisition and ontology learning. As Natural Language Processing plays an increasingly important role in this information age, revolutionary changes have taken place in the compiling and using of lexicography and terminology. The need for exploring the computational representation of lexical information that can be automatically processed by computer has become a pressing issue. The authors believe that the construction of a lexical ontology can be a practical and suitable solution for this problem, even for the development of a multilingual database like the Multilingual Central Repository (MCR). Through the project Galnet, the authors not only introduce the development and functions of a WordNet-based lexical ontology, but also demonstrate how the significant result and valuable information can be generated by Termonet in terminology learning under the guidance of a communicative theory of terminology. Moreover, in ontology learning, they prove how a new semantic categorization (epinonym) can be obtained automatically to explore the terminological implications between synsets, and how to build a web application to verify the Galician variants in specialized corpora by extracting all the variants of a category from other ontologies.

In the next chapter, Martínez and Varela present the steps taken to design and compile a multilingual terminological dictionary of language teaching and learning for a Spanish readership. They justify how the need for contextualized information from a local perspective becomes the actual starting-point of dictionary compilation. They also demonstrate how to make smart decisions before and during compiling from the angles of scope and content, contextualized information, targeted users, and the source used. Moreover, they explain how the content should be structured and presented in a user-friendly way. As the current edition is just a print one, they note the need to develop an online version in future. Thought-provoking ideas are offered for compiling a user-oriented terminological dictionary. However, the authors do not mention whether corpus tools or techniques are used. The compiling process would be more streamlined if corpus tools or techniques are employed in the data-selecting.

The last chapter by Taborek illustrates the conception and compilation of a multilingual LSP dictionary with the dictionary of football language *Wörterbuch der Fußballsprache* (WFS) as an example. He explains how the specific types of extra-lexicographical situation and the growing appeal for a multilingual dictionary of football, resulted in the motivation for compiling the multilingual football dictionary. He further showcases how to present the frame structure (front matter, main part and back matter), the macrostructure (lemma selection, base form and multi-word units), and the microstructure. In particular, he touches upon a number of thorny issues in the design of the microstructure. Questions such as how to present multi-word units, how to deal with ambiguity, and how to present grammatical properties in the example sentences are explained in detail. Readers may benefit a lot on how to compile a print multilingual dictionary. However, some readers may still wonder how to adapt

these techniques to the developing of a digital one. Therefore, it would be of more interest to readers if suggestions for developing of a similar online dictionary in the age of the Internet could be provided.

Overall, this book brings together a number of insightful researches on the multilingual electronic dictionary in this increasingly digital and multilingual world. Previously, few studies have touched on these particular topics. Teubert's volume (2007), adopting the bottom-up approach, looks into the compilation of corpus-based multilingual resources, whereas the volume edited by Boas (2009) describes the methods and applications of how to use frame semantics in multilingual lexicography. However, the studies in these collections specifically concentrate on computational lexicography with a strong focus on the construction of a multilingual FrameNet. With technological strides made in recent years, dictionary-making faces new challenges both theoretical and practical. Taking into consideration all these changes, this book offers illuminating ideas on how to address the current challenges both from a theoretical and practical perspective, giving examples of various dictionary projects and tools. However, users' experience is rarely mentioned in most contributions in this collection. As Hartmann (2001) underlines that dictionary use is one of the key perspectives in lexicography, the majority of researchers in this book do design their dictionaries with user situations and targeted users in mind, but few studies validate their claim with experimental data or answer the question to what extent users' needs and expectations are satisfied, except that conducted by Pastor and Durán-Muñoz. Of course, to some extent, the absence of verification of users' experience is understandable as tools in some of these digital multilingual projects are still in their testing phase.

To sum up, even if this volume still needs improvement with regard to the user's experience, it has already updated and contributed much to our understanding of multilingual lexicography in the digital age. The following points are worth special mentioning. Firstly, the new definition that dictionaries should be flexible and integrated information tools that are no longer to be sold as finished products but as ongoing services would turn the traditional static view of dictionary-making into a dynamic continuous process. This definition could better reflect the current dictionary-making reality in the increasingly digital age. Secondly, the view proposed by this book that lexicographical tools should deal not only with words, but also with terms, facts, and things could integrate the research on lexicography and terminology, also expanding the research scope of lexicography in the Internet age. Thirdly, with the ongoing digital revolution of lexicography in the emerging e-societies, the book highlights that the needs and reference skills of a new generation of users are worth special exploration. Lastly, the various types of multilingual tools and projects which are based on various linguistic theories and target different users can be presented as good examples to follow both for interested dictionary researchers and dictionary-makers.

Acknowledgements

This review was funded by Gansu Provincial Educational Science "13th Five-Year" Planning Program (GS[2019]GHB2056), Gansu Provincial High School Innovation Enhancement Program (2019A-068), Educational Reform Program of Lanzhou University of Finance and Economics (LJY201924), and Graduate Research and Innovation Program of Sichuan International Studies University (SISU2020YZ18).

References

- Boas, Hans C. (Ed.)**. 2009. *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin/New York: Mouton de Gruyter.
- Cenoz, J.** 2013. Defining Multilingualism. *Annual Review of Applied Linguistics* 33: 3-18.
- De Schryver, Gilles-Maurice**. 2009. State-of-the-Art Software to Support Intelligent Lexicography. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary and Lexicology*: 565-580. Beijing: Beijing Normal University.
- Fuertes-Olivera, Pedro A. and Sandro Nielsen**. 2012. Online Dictionaries for Assisting Translators of LSP Texts: The Accounting Dictionaries. *International Journal of Lexicography* 25(2): 191-215.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. London/New York: Routledge.
- Nielsen, Sandro**. 2018. LSP Lexicography and Typology of Specialized Dictionaries. Humbley, John, Gerhard Budin and Christer Laurén (Eds.). 2018. *Languages for Special Purposes: An International Handbook*: 71-95. Berlin/Boston: De Gruyter Mouton.
- Ruiz de Zarobe, Leyre and Yolanda Ruiz de Zarobe**. 2015. New Perspectives on Multilingualism and L2 Acquisition: An Introduction. *International Journal of Multilingualism* 12(4): 393-403.
- Teubert, Wolfgang (Ed.)**. 2007. *Text Corpora and Multilingual Lexicography*. Amsterdam: John Benjamins.

Hebing Xu and Wenfei Hu
Sichuan International Studies University
Chongqing
China
(163-xhb@163.com)

and

He Zhao
Lanzhou University of Finance and Economics
Lanzhou
China
(zhaohe.12@163.com)

Michael Klotz and Thomas Herbst. *English Dictionaries: A Linguistic Introduction*. 2016, 287 pp. ISBN 978 3 503 16751 7. Berlin: Erich Schmidt Verlag. Price: €19.95.

For a period spanning several centuries, lexicography has been seen, by professional and lay people alike, as a philological undertaking for which sprachgefühl alone will suffice. This is partly true, however, because it is not that linguistics is not helpful, but that modern linguistics appeared much later. Even after the genesis of Saussurean linguistics around the 1920s, few lexicographers realized that there is a substantial common ground between lexicography and linguistics. Preliminary drawing on linguistics was first seen in the developing English pedagogical sector of lexicography in the 1930s. Still, the linguistic influence on pedagogical lexicography at that early phase was not very significant. Actually, the influence was basically confined to such topics as controlled vocabulary and pedagogical grammar in applied linguistics. The scene, however, changed gradually but remarkably, and from the 1980s onwards lexicography in general has become growingly responsive to the advances in linguistics to the extent that many lexicographic innovations are realized through the creative application of cutting-edge linguistic theories (Wang 2004). Therefore, quite a number of introductions to lexicography are linguistics-laden (e.g. Al-Kasimi 1977, Jackson 2002, and Landau 2002). *English Dictionaries: A Linguistic Introduction* by Michael Klotz and Thomas Herbst is a recent addition to this fruitful association.

As suggested by the title, *English Dictionaries* deals mainly with English lexicography from a linguistic perspective. The authors are two German lexicographers, who had co-authored an introductory book on lexicography some ten years ago, written in Germany. However, it is not to be assumed that the present volume is a sheer translation of the previous work. The German original, as a matter of fact, underwent a complete revision in structure and content, with a focus on English dictionaries and references to up-to-date dictionaries (or new editions) and the latest literature.

Comprising 287 pages, the book proper has eight chapters, bookended by a preface and a terminological index plus a bibliography. Roughly, the eight chapters fulfil two major tasks: a general introduction and a thematic elaboration. Chapters 1–3 are devoted to general issues in lexicography. Unlike other writers of similar books, Klotz and Herbst deserve special credit for applying their own theory, namely 'the theory of lexicographic scope' (elaborated in Chapter 2). Integrated with the structural approach, the theory proposed in Chapter 2 serves as a thread running through the whole book in one way or another. In a sense, Chapters 1 and 2 can be seen as context chapters surrounding the core of the book. Subsequent to the introductory outline in Chapter 1 and the theoretical exposition in Chapter 2, Chapter 3 familiarizes the reader with the structural components of dictionaries. In Chapter 3 two traditional structural aspects, macrostructure and microstructure, receive greater attention and are treated comprehensively. Chapters 4–8 consider

detailed topics ranging from information categories and presentation modes to lexicographic sourcing, ending with a survey of English dictionaries that completes the book. English mono- and bilingual dictionaries are dealt with in Chapters 4 and 5 respectively, with well-conducted comparisons within some typological genres. E-lexicography is one of the vigorous driving forces that shape current lexicography and therefore it is presented individually in Chapter 6. Either for lexicographers or for lay users, the knowledge of exploring (re)sources for dictionary compilation can be more significant than imagined. Short as it is (running only 15 pages), Chapter 7 succinctly and innovatively surveys the language materials used in lexicography. The book ends with a sketchy account of various English dictionaries.

On the whole, the structural and thematic development of *English Dictionaries* does not possess too much innovation, but the reader can still find much lexicographic guidance. In response to the intention of being an introductory work, the pedagogical function is best demonstrated in its textual accessibility and logical clarity, as well as the informative and instructive bibliography which will direct the reader to further lexicographic publications. For readers who approach lexicography for the first time, the design is conducive to acquainting them with the new subject; on the other hand, readers will be in a better position to reflect on their linguistic knowledge from a lexicographic perspective. The discourse of the book unfolds in an objective but inviting tone. Wording like 'love at second sight' provides not only intertextual association, but is also revealing in interpreting the divided impression on dictionaries. The extensive use of dictionary extracts and cross references also facilitates the holistic networking of the subject.

To the present reviewer, the most outstanding merits that these two authors deserve are their fresh approaches to and insights into some lexicographic issues. For instance, thanks to their German background, they expertly transplant the Vermeerean Skopostheorie into lexicography with adapted parameters. Evidently, the authors strive to extend their horizons beyond structuralist linguistics in seeking theoretical support, as can also be seen in their adopting the Hallidayan construct of thematic progression in Chapter 3 to interpret the microstructural configuration of entries. Even within the 'orthodox' paradigm of linguistics, they manage to expose the reader to some neglected insights. A case in point is to be found in the use of 'addressing' for interpreting the aforementioned entry development as thematic progression. The notion of an 'addressing structure', which is often absent in other English literature on lexicography, thus figures prominently in the structural analysis of entries and sub-entries. The introduction of language materials is common in many similar books, but the authors present them in a novel way in this regard. In Chapter 7 the lexicographic data are classified into langue and parole data, with the latter subclassified into citations and corpora. This angle of presentation is conducive to the critical use of dictionaries, as well as the understanding of the very nature of dictionary generation.

Admittedly, there are still possibilities for the improvement of the book. It

would be advisable for the authors to draw more on other linguistic resources so as to better live up to the book's subtitle and to the recent developments in English lexicography. Except for the brief reference to the Hallidayan terms of theme vs. rheme and the Vermeerean Skopostheorie, the linguistic resources the authors turn to are basically confined to the structuralist school. The structural perspective alone, in the opinion of the present reviewer, would not suffice to bring descriptive and explanatory adequacy to the fast-developing discipline. For example, lexicographic application from the flourishing cognitive linguistics has brought about the 'cognitive turn' and exerted a great influence on lexicography. Another minor criticism is the neglect of comprehensiveness of English lexicography. Anglo-centeredness is always discernible in many narrations of English lexicography and *English Dictionaries* is not fully immune to this bias. Except for a few Germany-produced dictionaries, the survey of English dictionaries covers only those made in the two major varieties from the 'Inner Circle' (Kachru 1985), namely the UK and the USA. Over the past decades, English lexicography in the other two circles has flourished and thus constitutes a crucial share in the lexicography of English(es). British and/or American English dictionaries should not be the custodian of the Anglo-Saxon language and culture. In particular, the sector of learner's lexicography from non-Anglo-Saxon areas is in a position to shed some light on English learner's lexicography in general.

Overall, *English Dictionaries* deserves more credit than criticism. In the present-day times, English has grown into the de facto Esperanto in the world and the population of English learners has kept surging to outnumber English native-speakers. English lexicography occupies a significant standing in Anglistics. A close examination of the evolution of English dictionaries will bring the conclusion that linguistics has constituted the major catalyst for the progress of English lexicography. *English Dictionaries* is undoubtedly a stimulating publication worthy of its name. The present reviewer is convinced that *English Dictionaries* is a sine qua non for university students majoring in English and is also confident in recommending it to the wider community of English lexicographers and linguists.

Acknowledgements

This review is supported in part by the Social Science Foundation of the Fujian Province of China (Grant No. FJ2015A029) and the Fundamental Research Funds for the Central Universities of China (Grant No. 20720150008).

References

- Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
Jackson, H. 2002. *Lexicography: An Introduction*. London/New York: Routledge.

- Kachru, B.B.** 1985. Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. Quirk, R. and H.G. Widdowson (Eds.). 1985. *English in the World: Teaching and Learning the Language and Literatures*: 11-30. Cambridge: Cambridge University Press.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- Wang, F.F.** 2004. *Current Linguistics and Lexicographic Innovation*. Shanghai: Shanghai Lexicographical Press.

Kaihuai Du
College of Foreign Languages and Cultures
Xiamen University
Xiamen
P.R. China
(kindhawaii@163.com; kindhawaii@xmu.edu.cn)

Marie-Claude L'Homme. *Lexical Semantics for Terminology: An Introduction*. 2020, xxi + 264 pp. ISBN: 9789027204677 (Hardback). Terminology and Lexicography Research and Practice. Volume 20. Amsterdam/Philadelphia: John Benjamins Publishing Company. Price € 99.00.

When it comes to terminology, many studies have been undertaken from the perspective of lexical semantics and a fairly large body of work has been published, including Frawley (1988), Binon et al. (2000), Aldestein and Cabré (2002), Schmidt (2009), Pimentel (2013), Bernier-Colborne (2016), Ghazzawi (2016), the DiCoEnviro (2018), and the DiCoInfo (2018), etc. (pp. 52-53). L'Homme has constantly been working on the investigation of terminology and specialized lexicography from the perspective of lexicology and has contributed a series of works, including L'Homme (2002, 2003, 2004, 2007, 2012, 2015, 2018), L'Homme and Bernier-Colborne (2012) and L'Homme et al. (2018). As the latest publication of its kind, *Lexical Semantics for Terminology: An Introduction*, explores the correlation between lexical semantics and terminology, especially how principles and methodologies from lexico-semantic frameworks can help to make terms understandable and describe them in terminological resources. In addition, how lexical analysis complements perspectives primarily focused on knowledge is illuminated with a wide coverage of such topics as term identification, meaning, polysemy, relations between terms, and also equivalence. In short, this book well complements other textbooks on terminology that do not focus on lexical semantics *per se* and its distinctive focus makes it especially appropriate for its main target readership, i.e. those who are interested in words and work with specialized terms, e.g. terminologists, translators, lexicographers, and corpus linguists.

This book consists of nine chapters. Chapter 1 presents a panoramic view of the whole book. First, the author explains the reason why lexical semantics is applied to terminology by presenting the similarities between terminology and lexicology or lexical semantics, i.e. both disciplines deal with "the nature of words, word content (i.e. the nature of meaning), the relationship between word content and our construal of reality, and relations between words and word meanings" (p. 1). Subsequently, in the way of answering questions, L'Homme demonstrates how the principles, methods and tools developed by lexical semantics can help understand terms, how lexical analysis can be integrated in most steps of terminology work and how it complements perspectives entirely focused on knowledge. The answers to these questions naturally form different chapters (Chapter 4 to Chapter 9), followed by the concrete applications of lexical semantics in terminology, lexico-semantic analysis of terms, together with the intended readership and introduction to illustrative examples.

Chapter 2 sketches the discipline terminology and its basic principles, particularly the knowledge (or conceptual) paradigm in which terms are often understood. The author presents a definition of terminology and introduces

some keywords, i.e. *domains, structure, concepts, conceptual structures, designations and terms* and a distinctive feature of terminology, application, which terminology is deeply rooted in and is strongly evidenced by the examples from different fields the author employs in the book. "The knowledge paradigm", "Storing and accessing concepts and terms" and "The knowledge paradigm from a linguistic point of view" constitute the rest of this chapter. Firstly developed by Wüster (1979), the General Theory of Terminology (GTT), offers a knowledge-driven approach to understanding terminology. It considers "terms as labels for concepts and as units emptied of most of their linguistic properties" (p. 17), and of course receives increasing criticisms. However, a perspective guided by the lexical content of specialized texts (lexicon-driven approach) regards terms as full linguistic units which may undergo variation and ambiguity like general lexical units. The two different ways to handle terms are necessary to fully understand the relationship between linguistic units and specialized subject fields, though a challenge is posed here for terminologists on how to strike "a balance between an ideal for unambiguous communication and the fact that language changes, and accommodates vagueness as well as ambiguity" (p. 18).

Chapter 3 compares the way in which the knowledge-driven and the lexicon-driven perspectives consider terms, and draws the conclusion that the former is more suited for applications of knowledge modeling while the latter is more adequate for the explanation of the behavior of terms in open-running texts. L'Homme further highlights the important role of corpora in terminological analysis, since they are often used by terminologists to acquire knowledge, acquaint themselves with terms, make subtle meaning distinctions and establish relations between terms. In spite of their limitations like most frameworks, such frameworks as Explanatory Combinatorial Lexicology and Frame Semantics used to analyze and describe terms are also introduced in this chapter.

In Chapter 4, the identification of terms from a short text on *endangered species* leads us to the belief that the notion of 'term' is relative to the delimitation of a special subject field and to the application, and different experts produce diverging lists when asked to identify terms in texts. Then four criteria based on lexical semantics, including the relationship with a field of knowledge, the nature of arguments, morphological and semantic relatedness and paradigmatic relations are presented to aid terminologists' selection of terms in specialized texts. Additionally, the author argues that terms are open-class units and can belong to one of the four following parts of speech: nouns, verbs, adjectives, and adverbs, which is quite insightful and different from the overwhelming preference for nouns in most specialized dictionaries and other terminological resources.

Chapter 5 further explores the distinctions between knowledge-based and lexicon-based approaches to terminology since they have a direct impact on how terms are defined, related to each other and described in terminological resources. The two different perspectives on linguistic content also uncover

underlying reasons why different explanations are often offered for similar phenomena. The knowledge-based method is based on the feasibility to differentiate concepts and to define them according to a finite list of features, whereas the lexicon-based method addresses the issue by polysemy. This chapter also contains five criteria for handling polysemy in specialized texts, which include substitution with a synonym, differential opposition, differential morphological derivation, differential paradigmatic relations and compatible and differential co-occurrence. The five criteria used to make meaning distinctions are a perfect illustration of the relational approach to which the meaning of linguistic units (here lexical units) is delimited according to the way they interact with others.

Since the knowledge-based perspective and the lexicon-based perspective advocate two different philosophies of defining terms, Chapter 6, on the basis of discussions in the previous units, presents fundamentally how to differentiate between types of lexical units and terms, i.e. predicative units and non-predicative units and how to represent predicative units used in terminology and other frameworks. Since terms, according to the author, can be non-predicative, predicative or quasi-predicative, arguments as obligatory participants are required to fully capture the meaning of terms and two criteria are suggested to describe the predicative or quasi-predicative: (1) "arguments are necessary to account for the meaning of the unit; (2) arguments are expressible in texts" (p. 142). Different systems that capture generalizations, including variables, semantic roles, semantic classes as well as frame elements are employed to represent the argument structure of terms, while different methods showing the interaction between arguments and predicative units in sentences are used to represent the link between the argument structure and the syntactic realization of arguments.

"Understanding and representing relations between concepts or terms are fundamental aspects of terminological analysis" (p. 145) in both knowledge-driven and lexicon-driven approaches. Treating relations as central in terminology, Chapter 7 profiles and distinguishes among different kinds of conceptual relations and terminological relations. The former include taxonomic relations, partitive relations, synonymic relations, oppositional relations and other conceptual relations, while the latter mainly include paradigmatic (vertical) and syntagmatic (horizontal) relations.

Following the description and application to small sets of concepts and terms, Chapter 8 focuses on larger and more complex networks of relations, structures and exhibits how different conceptual and terminological structures are created and exploited. Based on concepts, as opposed to meanings in lexical analysis, conceptual structures account for the organization of knowledge and can be represented graphically through tree representations or graphs. Terminological structures usually account for various kinds of relations between terms and the paradigmatic and syntagmatic relations are often encoded by lexical functions of terms and are represented and linked in semantic frames.

In addition to analysis of relations in a single language in Chapters 7 and 8, Chapter 9 concludes the book by examining relations across languages, more specifically relations of equivalence, and various challenges posed when adopting knowledge-driven and lexicon-driven approaches. These are further elaborated in the succeeding three parts, i.e., conceptual equivalence, terminological equivalence and problems when establishing equivalence (including non-equivalence, partial equivalence, structural divergences as well as equivalence in running texts). Whichever approach it is, exact equivalence is always the aim to achieve since it is what the terminological resources (dictionaries, term banks, thesauri, etc.) record. Equivalence can also be established between terms in running texts despite the fact that difficulties will arise when automated methods are used.

As an introduction to the application of lexical semantics in terminology, this book exhibits a number of merits, as discussed below.

Firstly, as a collection, reiteration and development of the author's long-standing views to understand terminology from the perspective of lexical semantics, this book is different from, and thus supplementary to, the traditional knowledge-based method. The proposal of a lexicon-driven approach is a big leap forward in the manifestation and interpretation of terms since it takes a completely reversed path, from lexicon to knowledge rather than the traditional knowledge-to-lexicon mode (L'Homme 2018). The weaknesses and strengths together with suitable circumstances and illustrative examples present a detailed and panoramic view of both approaches and the lexicon-based approach is thus highly underscored since it is more compatible with current corpus-based practices (L'Homme 2004). Secondly, in accordance with the lexico-semantic analysis, examples used in the book mainly come from active and dynamic corpora instead of term banks and specialized dictionaries, which further reinforces the indispensable role of corpora in terminological research. It is also stressed that corpus evidences should be supported and complemented by other sources of information. Thirdly, the lexicon-driven method lays particular emphasis on the explanation and analysis of senses in corpora, thus it draws our attention to different parts of speech of terminologies and makes it possible to draw subtle distinctions through the meticulous analysis of verbs, adjectives, and adverbs related to noun terms which take up the overwhelming majority in specialized dictionaries and term banks. The shift of focus from noun-dominant terms to terms with different parts of speech suggests that the inconsistencies caused by the omission or overlooking of verbs, adjectives and adverbs can be addressed with a descriptive method (L'Homme 2003). Fourthly, the clues provided by lexical semantics as to how humans encode meaning in their minds give us the opportunity to establish hypotheses on the storage and processing of lexical units, terms and their meanings by human beings. This indicates much potential in future research in the justification or falsification of these hypotheses combined with experimental methods. The last highlight of this book lies in its reader-friendliness owing to its elaborate design

and arrangement of contents. The introduction to basics of terminology (Chapter 2), the copious demonstrative examples from different fields and the diagrammatic and graphic presentations break the barrier for its intended readers who do not necessarily have a background in terminology. In different chapters, the introductory part serves as a link between the relevant contents in the previous chapters and what follows. The summary provides a concise and precise synopsis of the main topics covered, while the annotated list of references offers useful resources for further reading. All these are rather considerate and helpful for the readers.

In addition to the strengths discussed above, this book could even have been better if the following aspects were more carefully considered. To begin with, it could be stated more clearly that parts of speech do not equate word classes, although the author has already noted their differences when talking about limitations of frameworks derived from lexical semantics (Chapter 3). Lehmann (2013: 143) notes that *part of speech* is a calque on the Latin *pars orationis*, which is a calque on the Greek *méros lógou*, all of which mean literally 'part of speech' or 'part of sentence'. Word classes instead, are defined as lexeme classes in structural linguistics. A lexeme is an abstraction corresponding to a class of word-forms and, therefore, a component of the system rather than of the text. Consequently, lexeme classes, too, are essentially components of the language system. Thus, a word class in the sense of 'lexeme class' is not actually a 'part of speech' (or of the sentence). Wang (2014: 346) and Wang and Huang (2017: 88-89) hold that word class categorization takes place at two levels — *parole* and *langue*: 'part of speech', a term for word tokens in syntax at *parole*, refers to the speaker's propositional speech act functions including reference, predication and modification; 'word class', a term for word types or lexemes in the lexicon of a communal language (*langue*), refers only to the conventionalized propositional speech act functions. Word tokens are often found in specific contexts and can assist researchers as corpus concordances while word types are usually listed in dictionaries or term banks as lexemes by dictionary compilers. The two terms applying to different contexts indicate two different levels of word class categorization and should be distinguished from each other. In addition, it is argued by the author that the knowledge-based and lexicon-based approaches are supplementary to each other and a balance is supposed to be maintained when terms are represented and explained (L'Homme 2018). However, it appears that this viewpoint mainly ushers in "What", a new and supplementary lexicon-based method, whereas the issue as to "How" to maintain the balance has not been sufficiently addressed. What specific rules to follow? What yardsticks to use to assess whether a balance is kept or not? Seemingly no definite answers are given in this book. What is more, the author employs plenty of examples from a wide range of fields, most of which, however, come from natural sciences, such as computing, anatomy, environment, zoology, biology, chemistry and geology. The observations drawn may be more repre-

sentative and persuasive if more examples were taken from social sciences as well.

To sum up: As an introductory monograph on the analysis of terms with the help of philosophies and methods borrowed from lexical semantics, this book provides a different perspective and a new solution to terminological representation and interpretation. Differing from the traditional knowledge-driven approach which depends on a prior delimitation of concepts and advocates a prescriptive way of terminology representation and interpretation, the lexicon-driven approach introduced in this book treats terms as lexical units delimited syntagmatically and paradigmatically and defines their meanings according to their relations to other units. This relational approach, which indicates a more descriptive way, is more compatible with current corpus-based practices, provides useful criteria for selecting terms and making fine-grained semantic distinctions, and supplements well the traditional knowledge-based method in term analysis. This book tightly bonds terminology and lexical semantics and will elicit more interesting future studies and contribute to the development of lexicon-oriented terminology studies; as such, it is strongly recommended.

Acknowledgements

This work was funded by the Excellent Youth Program of the Educational Department of Hunan Province (Grant No. 18B375), and the Graduate Research and Innovation Program of Sichuan International Studies University (Grant No. SISU2020YZ13).

References

- Aldestein, A. and M.T. Cabré.** 2002. The Specificity of Units with Specialized Meaning: Polysemy as Explanatory Factor. *D.E.L.T.A.* 18: 1-25.
- Bernier-Colborne, G.** 2016. *Aide à l'identification de relations lexicales au moyen de la sémantique distributionnelle et son application à un corpus bilingue du domaine de l'environnement.* PhD thesis, University of Montreal, Montreal.
- Binon, J., S. Verlinde, J. Van Dyck and A. Bertels.** 2000. *Dictionnaire d'apprentissage du français des affaires.* Paris: Didier.
- DiCoEnviro.** 2018. *Dictionnaire fondamental de l'environnement.* Accessed 4 April 2018: <http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search.cgi>
- DiCoInfo.** 2018. *Dictionnaire fondamental de l'informatique et de l'Internet.* Accessed 4 April 2018: <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>
- Frawley, W.** 1988. New Forms of Specialized Dictionaries. *International Journal of Lexicography* 1(3): 189-213. Available at: <https://doi.org/10.1093/ijl/1.3.189>

- Ghazzawi, N.** 2016. *Du terme prédictif au cadre sémantique: méthodologie de compilation d'une ressource terminologique pour des termes prédictifs arabes en informatique*. PhD thesis, University of Montreal, Montreal.
- L'Homme, M.-C.** 2002. Fonctions lexicales pour représenter les relations sémantiques entre termes. *Traitement automatique de la langue (TAL)* 43(1): 19-41.
- L'Homme, M.-C.** 2003. Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives: A Model for Specialized Lexicography. *International Journal of Lexicography* 16(4): 403-422. Available at: <https://doi.org/10.1093/ijl/16.4.403>
- L'Homme, M.-C.** 2004. A Lexico-semantic Approach to the Structuring of Terminology. *Proceedings of CompuTerm 2004. 3rd International Workshop on Computational Terminology*: 7-13. University of Geneva, Geneva, Switzerland: COLING.
- L'Homme, M.-C.** 2007. Using Explanatory and Combinatorial Lexicology to Describe Terms. Wanner, L. (Ed.). 2007. *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel'cuk*: 163-198. Amsterdam/Philadelphia: John Benjamins. Available at: <https://doi.org/10.1075/slcs.84.06hom>
- L'Homme, M.-C.** 2012. Using ECL (Explanatory Combinatorial Lexicology) to Discover the Lexical Structure of Specialized Subject Fields. Apresjan, J. et al. (Eds.). 2012. *Words, Meanings and Other Interesting Things. A Festschrift in Honour of the 80th Anniversary of Professor Igor Alexandrovic Mel'cuk*: 378-390. Moscow: RCK.
- L'Homme, M.-C.** 2015. Predicative Lexical Units in Terminology. Gala, N., R. Rapp and G. Bel-Enguix (Eds.). 2015. *Recent Advances in Language Production, Cognition and the Lexicon*: 75-93. Berlin: Springer.
- L'Homme, M.-C.** 2018. Maintaining the Balance between Knowledge and the Lexicon in Terminology: A Methodology based on Frame Semantics. Peters, P., J.G. Yongwei and J. Ding (Eds.). 2018. *Medical Lexicography and Terminology. Special Issue of Lexicography. Journal of Asialex* 4(1): 3-21.
- L'Homme, M.-C. and G. Bernier-Colborne.** 2012. Terms as Labels for Concepts, Terms as Lexical Units: A Comparative Analysis in Ontologies and Specialized Dictionaries. *Applied Ontology* 7(4): 387-400.
- L'Homme, M.-C., B. Robichaud and N. Prével.** 2018. Browsing the Terminological Structure of a Specialized Domain: A Method Based on Lexical Functions and their Classification. Calzolari, N. et al. (Eds.). 2018. *Proceedings of LREC 2018, Eleventh International Conference on Language Resources and Evaluation, May 7-12, 2018, Miyazaki, Japan*: 3079-3086. Paris: European Language Resources Association. Accessed 18 April 2019: <http://www.lrec-conf.org/proceedings/lrec2018/summaries/840.html>
- Lehmann, Christian.** 2013. The Nature of Parts of Speech. *STUF – Language Typology and Universals* 66(2): 141-177. Available at: <https://doi.org/10.1524/stuf.2013.0008>
- Pimentel, J.** 2013. Methodological Bases for Assigning Terminological Equivalents. A Contribution. *Terminology* 19(2): 237-257. Available at: <https://doi.org/10.1075/term.19.2.04pim>
- Schmidt, T.** 2009. The Kicktionary — A Multilingual Lexical Resource of Football Language. Boas, H.C. (Ed.). 2009. *Multilingual FrameNets in Computational Lexicography. Methods and Applications*: 101-134. Berlin/NewYork: Mouton de Gruyter.
- Wang, R.Q.** 2014. The Two-Level Word Class Categorization in Analytic Languages. *Proceedings of the 36th Annual Conference of the German Linguistic Society, 5-7 March 2014, Philipps University Marburg, Germany*: 345-347.

Wang, R.Q. and C.N. Huang. 2017. Heterosemy of Self-reference Lexemes in Modern Chinese From the Perspective of the Two-level Word Class Categorization Theory'. *Foreign Language and Literature* 33(1): 87-96.

Wüster, E. 1979. *Introduction to the General Theory of Terminology and Terminological Lexicography*. Vienna: Springer.

Lingyun Lü
Sichuan International Studies University
Chongqing
China
and
Hengyang Normal University
Hunan
China
(dillonlv@163.com)

and
Renqiang Wang
Sichuan International Studies University
Chongqing
China
(richardwang71@163.com)

Van Wyk, E. (Compiler). *The Talking Dictionary*. 2019, 288 pp. ISBN: 978-1-920217-74-7. Pretoria: Briza Publications. Price R1250.00. (Additional language sets: R280.00 each.)

The Talking Dictionary hit the market last year in October. It appeared to me that the product was well received by those attending the launch, but is this excitement warranted? In this review, I will explain how the product works, examine its merit and also its usefulness as far as practical application is concerned. I will also speculate as to the possibilities of this product within the context of a classroom. For the purposes of this review, English will be considered the source language and the target language would be any language available from the publisher. Much of what I'm writing about this publication in use is pure speculation, as there is no way to know how it will be received in a classroom situation or even in a home situation at this time.

How does it work?

The Talking Dictionary (hereinafter TTD) includes over 5000 entries and 2000 illustrations which each can be listened to by making use of a call finder, which will be discussed later. The dictionary is organised thematically and none of the entries are organised in alphabetical order. The other aspect that makes this dictionary unique is the fact that there is a section with common phrases and sentences, which you could listen to and read in the target language.

TTD is rather unique and I believe a first on the South African market (the publisher claiming it is one of a kind). While the product is specifically aimed at children (therefore making it a pedagogical dictionary), it may be used for other purposes too, such as basic language acquisition, though this statement is purely speculative.

This dictionary comes with a device called a call finder. The call finder recognises a special pattern printed into the word or illustration, which then enables the reader to find the correct entry in its memory. Similar products have been produced by the publisher of this book, specifically to play bird sounds in field guides.

The device can be loaded with as many language packages as the user intends to purchase. English comes standard with the purchase of the dictionary (call finder included). Additional language packages can be purchased, which include languages like Zulu, Xhosa, Pedi, Tswana and more. A Mandarin language package is in the pipeline, according to the publisher.

One could quite easily plug in a set of headphones into the call finder, making it practically quiet for when a child wishes to use the dictionary on their own, or plug in a set of speakers to play the word to an entire class.

Not fully electronic and not fully print

TTD is a strange mix of digital and print. It is neither fully a print dictionary nor an e-dictionary. Early e-dictionaries were simply digitised versions of print dictionaries on floppy disc or CD ROM (Bothma 2011: 72). Since then, many print dictionaries, such as the HAT, VAW and PWN include CDs in the print dictionaries, but many digital dictionaries are only available online (OEDOnline, Wiktionary and Urban Dictionary springing to mind). It would be completely pointless, in my opinion, to have print editions of either Wiktionary or Urban Dictionary, as the users of these products are digital natives. The user perspective therefore, I believe, is "it's online and convenient" because it's fully digital like OEDOnline or many of the dictionaries published by Pharos. These dictionaries are also much more suited to an online environment and the act of browsing than a print dictionary would ever hope to be.

TTD is an odd mix between the two. Yes, you need internet access to load the information onto the call finder, but that is where it ends. You are not bound by internet or at the mercy of bad signal or power cuts, as the call finder works with batteries and stores the entries in its memory. It is meant for children that are essentially digital natives, which may seem bizarre at first glance. One could argue that there are places in South Africa where the 4IR has not yet taken place, never mind the 2nd or 3rd. The media runs stories at the beginning of every school year to show how great the digital divide in the country is and how bad some of the conditions in some schools are, especially in Limpopo and the Eastern Cape. It seems bizarre to try and implement a system such as TTD in this context.

In a way TTD is a happy medium between the two extremes. In a way it gets younger children used to technology and more sophisticated devices, but on the other also exposes them to books, which is still the norm in this country, provided they get delivered.

Print dictionaries are therefore still the standard in South African schools, provided they have access to any kind of lexicographic work. A print dictionary is not dependent on electricity, internet or any device. It is a self-contained unit that children still need to know how to use, should they one day wish to browse a digital dictionary such as the OEDOnline. The OEDOnline has a section where you can browse and uses dictionary typology to do so with a menu listing all the possible words starting with the given n-gram entered into the search bar.

Dictionary skills both for print and digital dictionaries are still useful to have. It may also prove that this product fosters both in time, but that will have to be seen in future.

It clearly resembles picture books very much in the style of "my first words" or "my first dictionary". These publications usually have one word per page in big print, usually on cardboard pages. Slightly more advanced material

may have more than one illustration per page, but similarly have a rather large print and only one translation equivalent.

It is however a slightly more advanced dictionary, like Collins My First Dictionary (Youtube 2011). This publication however is structured in alphabetical order, even though it does have colourful illustrations and large print. It is also specifically aimed at the age group 5 to 7, as stated in the video. TTD's intended audience is a bit more fuzzy.

TTD has very small print in comparison to the aforementioned publications (both the picture books and the First Dictionary) because it is clearly intended for older readers (bearing in mind older readers range from about 9+).

Another feature of the dictionary are the sentences and sentence fragments in the final section and the paragraphs introducing the various themes/sections. These can be played over the call finder too, and in theory, one could use the sentence fragments to build sentences with words from other sections.

Because the nature of this dictionary is a bit difficult to determine, it makes it difficult to render a clear opinion. On the one hand, it's great for kids learning new words. On the other, how long will it stay useful before more advanced material is needed? How long until the call finder is replaced with a smart device where you need internet access all the time to work this system, again leaving those who would benefit most from the product in the cold? This situation is definitely one to keep an eye on, as it may cause a bit of a stir.

It is also the only dictionary I've ever come across with an index and every additional language package also has an index. It makes it easier to find a specific word, but it's still not as easy as a traditional dictionary.

Is it any good?

Using the continuum of the ideal dictionary and ideal user as set out by Gouws and Prinsloo (2005: 42) as a guideline, we can rate this dictionary as having fair to good lexicographic merit (taking into account its purpose and target user) and of course the user would be a user with little to no dictionary skills, since the target user of this dictionary is children. One could only hope that use of this dictionary will eventually lead to the user using more appropriate dictionaries as they age, such as the *Oxford Afrikaans-English School Dictionary* (OAESD) or the *Pharos Junior Bilingual Dictionary* (PJBD).

Getting young children interested in dictionaries and dictionary use is of vital importance. TTD can of course be a stepping-stone toward more advanced dictionaries which will match the needs of the user.

The jump from the usage of TTD to, for example, the OAESD is rather vast, as TTD is not organised as a traditional dictionary is. It should however not be an enormous problem, because the organisation of traditional dictionaries already make it more user friendly by giving the lemmas in alphabetical

order as opposed to category. Of course, the fun of listening to the word as you read it might make a traditional dictionary a bit more boring at first.

Another big difference between the two publications is the fact that the OAESD includes a comment on grammar and comment on semantics with the translation equivalent, and in many cases also gives examples of usage, where TTD only gives the translation equivalent.

It's a pity that the product (as any book in South Africa) is quite expensive. At the time of the launch, the basic package (TTD, the call finder and 1 target language package including a book and download) was R1250. It therefore may not reach those who would benefit the most from it, namely second or third language learners in rural or underprivileged schools. It's fine to argue that you could have a TTD per school or even per class, but would it become yet another thing that you could steal from a school and sell, like computers, TVs or smart boards? Only time will tell.

A pedagogical dictionary, no matter how novel or advanced, is no substitute for a teacher or tutor, who will point out the finer points of usage and grammar where any pedagogical dictionary may fail.

The visual dictionary

A newer genre of dictionary (at least, one I recently discovered) is the visual dictionary. It seems that TTD fits better into this category (still as a medium between traditional and electronic).

It is stated in TTD that it's based on *Le Nouveau Dictionnaire Visuel Junior* (NDVJ) by QA International publishing (Van Wyk 2019: 2). This publication is very similar to TTD in many ways, even sharing some of the illustrations, as the publisher bought the copyright to many of the images. One only needs to look at the preview of NDVJ on the website to see many similarities. On the preview of the dictionary (QAI 2020), it is clearly stated that NDVJ is intended for readers 9 and up, which leads me to believe that the target audience for TTD is around 9+ as previously stated.

From what I can gather from the preview, these two publications follow the exact same approach to categorisation and "entries". NDVJ is a bidirectional dictionary, meaning the English translation equivalents are printed in the dictionary itself. It is not a multi-book system with a call finder and looks to be more user friendly.

It seems QAI was at the forefront in providing more visual stimulus to their readers than other publishers.

There are also similar publications online, but those that I could find using Google seem all to be affiliated with QAI.

The "entries" are built around a theme, and possibly even a single image, where multiple words may be present.

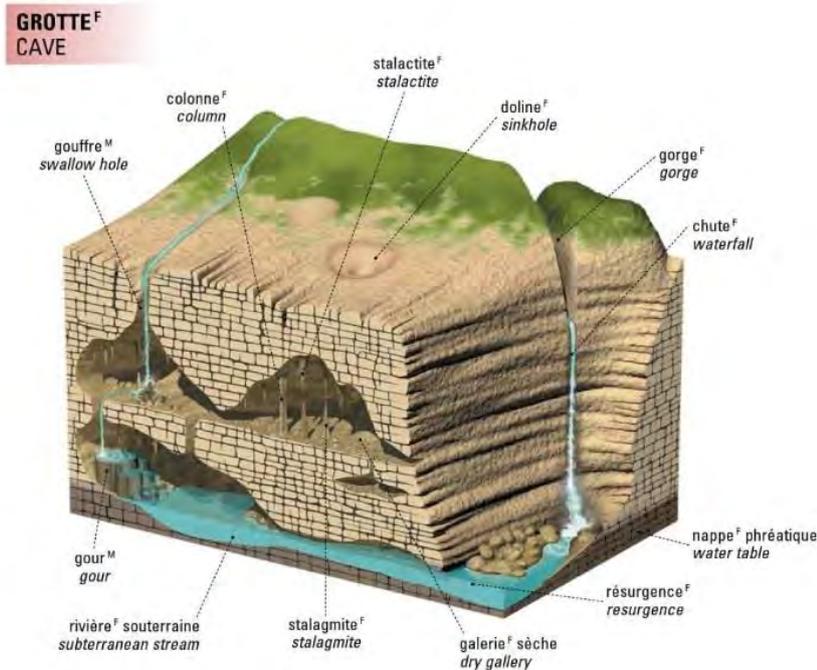


Figure 1: An example of an illustration from NDJV (2003: 30)

It is rather simple to find the word from the image. I am hesitant to use the word "entry", as I will discuss in the next section.

A review of the 2004 edition the NDVJ (as reviews seem hard to find about this publication) was very positive that not only children would benefit from this book (Rufiange 2004: 69). In the review it is indicated that the NDVJ is intended for those 7 and up, which contrasts with the 9+ given in the newer catalogue.

Be that as it may, it gives credence to my thought that it might benefit readers of all ages and levels who may be learning a specific language found in the system. A study using an online visual dictionary seems to support my hypothesis (Osmani 2014: 711).

Research into this format might prove both interesting and fruitful.

The content

The "entries"

Most of the words found in the dictionary are nouns, leading me to believe that

it is somewhat unbalanced. There are basic pronouns and prepositions and very few verbs and adjectives. As far as language acquisition goes, it's very good to expose learners (in this case children) to as many nouns, verbs, adjectives and adverbs as is possible. I do believe that there could be a greater focus on adjectives and basic verbs, so to aid the acquisition of new vocabulary. It may prove difficult to illustrate adverbs in the context of this dictionary where a word is given with little or no context.

The problem however is that some of the nouns found in the dictionary may be a bit too advanced for younger readers — such as specific botanical or mathematical terminologies that would only be of use later in their school career. Many of these terminologies will also be of little or no consequence in a second language or foreign language classroom with young readers. It may however interest older readers who are familiar with those terms or are learning them.

On the other hand, there are terminologies both basic (e.g. parts of the body) and more advanced (e.g. space travel), which might prove useful for school projects and the like in terms of building vocabulary. I must however mention that it is not within the scope of this review to comment on any of the African languages or the Mandarin language packages, as I lack the knowledge to do so. I would think it prudent that linguists and lexicographers specialising in those languages examine the quality of these translation equivalents, but more about that later.

While it is not impossible to illustrate demonstratives in a dictionary, doing so may prove difficult from English to Afrikaans and vice versa solely via an audio recording of the word "this" or "that". "This" and "these" are perfectly acceptable translation equivalents for the Afrikaans words "hierdie" and "dié", which may lead to confusion, should there be little or no guidance given to the user. It may well have been wise to omit demonstratives from the publication. It does however illustrate how problematic it can be where more than one translation equivalent exists and where only one word is given. An example of this problem follows in the next section.

I believe that the pronouns could show more than just the subject form thereof. The possessive forms and object forms of the pronouns could quite easily be given in a table, such as the colours are, but illustrating it may prove difficult, given that it is a printed dictionary.

While a good job was done with prepositions and words denoting position, it could be improved upon in the next edition by adding concepts such as "near", "back", "next to", "far" and other words.

I have purposely not referred to the words in this dictionary as entries, because very often it's just a word accompanying an illustration. No other grammatical or semantic information is given, therefore making the use of the word "entry" questionable.

The Afrikaans language package

If we set aside the fact that it's difficult to find specific words owing to the structure of the dictionary and the fact that some of these words may be a bit too advanced for the intended readership, the Afrikaans language package performs rather well. The translation equivalents given are accurate and there are not enormous issues with the package itself. Yes, the adjectives are all given in predicative form, but that is a minor issue which can be solved by a teacher, tutor or grammar. You could in theory build basic Afrikaans sentences by combining the sentences given in the last section with many of the words in the dictionary.

The pronunciation is standard and formal. The words, sentences and paragraphs are all expertly read at a pace that younger users can follow. The problem of course will come in where a native speaker uses this system and there are dialectic differences or some such, which are possible in Afrikaans. It may be an even bigger problem in the African languages or Mandarin.

The system is also a bit clunky because you essentially need two books to do what one bidirectional dictionary can do, but this will be true of all the language packages.

But what about the African and Mandarin language packages?

The greater problem staring this review in the face, of course, is the fact that I can only review the Afrikaans language package. Afrikaans and Mandarin (which again is only in the planning stage to the best of my knowledge) would be the only two languages that have sufficient lexicographic and digital resources as well as scientific vocabulary for advanced themes, like botanical or paleontological terms (which are both included). A discussion with one of the voice artists who did recordings for the dictionary, leads me to believe that terminologies for the African languages may have been created for this publication. I however lack sufficient knowledge to even make an educated guess as to the accuracy of any of the translation equivalents given in the publication in an African language. I would like to encourage a similar review that focuses on the African languages and the quality of the translation equivalents, given the concerns I have expressed here.

One translation equivalent per word could prove a tricky business when other options are available, such as formal and informal greetings or even adjectives that may need to be declined for concord. In isiXhosa *molweni* is the greeting which is used for those who are older than you, higher in stature or as a plural, where *molo* is more informal/familiar. The different options are not given, which might also account for the massive amount of nouns in the publication, which do not necessarily need more than one translation equivalent. A similar problem exists in the Afrikaans language package, where most, if not

all, adjectives are given in the predicative form, which may confuse learners who encounter the attributive forms in a text. There are words where two possible equivalents are given, but those are few.

Morpho-syntactic issues like these would need to be addressed in language packages where languages, like the Nguni or Sotho sa Leboa languages, are morphologically complex. Again, I am encouraging others to review these language packages as I am no expert on African languages or Mandarin.

Some thoughts

It is rather difficult to review anything unique. After all, it is not usual to have the sense of hearing engaged to this extent in a print dictionary. One usually needs computer access for that to happen or know how to read IPA transcriptions. It is also quite unusual to have movement of such a nature involved with dictionaries too, which of course also engage kinaesthetic learners more easily.

But given the aim of the dictionary and the price range thereof, it's a nice to have which would be completely out of reach to those who would benefit most from this product: those who do not have access to pronunciation by native or near native speakers in a second language or foreign language classroom.

The big influence this publication may have, is that it may encourage younger dictionary users (from about grade RR age 4 to about grade 3 age 10) to engage with more suitable material later on (grade 4 upwards), such as the OAEDS or even the various bilingual dictionaries offered to that market, such as the PJBD.

To put it simply, TTD is a "nice to have". The quality of the dictionary is commendable and the idea that is being promoted here is quite good. From a purely lexicographic stance, the initiative could be highly praised, provided the African language and Mandarin packages perform as well as the Afrikaans package. However, lexicography is a user driven enterprise and TTD should be evaluated with that in mind.

When I contacted the publisher in January 2020, sales figures did not look good. That may very well be due to the fact that the product was launched last October. Very few children, if any in my opinion, would like a dictionary for Christmas, and money is usually tight in January when new school supplies need to be purchased. A traditional dictionary would then probably be the dictionary of choice to many parents, given the fact that definitions and not just translation equivalents are given in many of these publications (such as OAEDS or Pharos).

The use of the call finder could also be a hindrance in the classroom, even when used with earphones. And once again, those who probably would benefit the most from the product, would be left in the cold because of price.

Conclusion

I would recommend further inquiry into the potential of this publication. It might be prudent to see how children react to the product and then write another review based on the target market's reception thereof, again working with the frame of user driven lexicography. It is therefore with much turmoil I have to conclude this review with "let's see what happens". It is a concept which is new so we have no other option but to wait and see how it will be received by lexicographers and users alike.

Bibliography

Dictionaries

Collins My First Dictionary. See **Other literature**.

(HAT) **Odendal, F.F. and R.H. Gouws (Eds.)**. 2005, 2011. *Handwoordeboek van die Afrikaanse Taal*. Pinelands: Pearson Education South Africa.

Louw, P. (Ed.). 2007. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. Cape Town: Oxford University Press Southern Africa.

(NDVJ) **Corbeil, J.-C. and A. Archambault**. 2003. *Le Nouveau Dictionnaire Visuel Junior*. Montreal: Québec-Amérique International.

OEDonline as accessed on 11 March 2020 <https://www.oed.com/> Oxford University Press.

(PJBD) **Pharos Dictionaries**. 2018. *Pharos Junior Bilingual Dictionary/Junior Tweetalige Skoolwoordeboek*. Cape Town: Pharos.

(PWN) **Weijnen, A.A. et al.** 1955, 2007. *Prisma Woordenboek Nederlands*. Utrecht: Het Spectrum.

Urban Dictionary as accessed 12 March 2020 <https://www.urbandictionary.com/>.

Van Wyk, E. (Compiler). 2011. *The Talking Dictionary*. Pretoria: Briza Publications.

(VAW) **Labuschagne, F.J. and L.C. Eksteen**. 1993, 2007. *Verklarende Afrikaanse Woordeboek*. Pretoria: Van Schaik. Later editions: Cape Town: Pharos

Wiktionary as accessed on 11 March 2020 https://en.wiktionary.org/wiki/Wiktionary:Main_Page.

Other literature

Bothma, T.J.D. 2011. Filtering and Adapting Data and Information in an Online Environment in Response to User Needs. Fuertes-Olivera. P.A. and H. Bergenholtz. 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: Sun PReSS.

Osmani, R. 2014. The Effectiveness of Using Technology (Online Visual Dictionary) in Teaching and Acquiring Vocabulary. A Study Conducted at Seeu. *The Journal of Teaching English for Specific and Academic Purposes* 2(4): 707-713.

QA International Home page 2020 as accessed on 16 March 2020 <http://www.qa-international.com/> and Catalogue Preview Page <http://www.qa-international.com/docs/CatVisuelJunior.pdf> as accessed on 16 March 2020.

- Rufiange, C.** 2004. Review of *Le nouveau dictionnaire visuel junior français–anglais*. *Lurelu* 27(2): 68-73.
Youtube. 2011. *Collins My First Dictionary — Buying Dictionaries for 5-7 Year Olds* as accessed on 25 Feb 2020 <https://www.youtube.com/watch?v=1rvkkbBvMsE>.

Maria Hermione le Roux
Department of Afrikaans and Theory of Literature
Unisa
Pretoria
South Africa
(lrouxmh@unisa.ac.za)

Publikasieaankondigings / Publication Announcements

John Considine (Editor). *The Cambridge World History of Lexicography*. 2019, xii + 961 pp. ISBN 978-1-107-17886-1. Cambridge: Cambridge University Press. Price: \$155.00. (Review in this issue.)

L.G. de Stadler en A. de Stadler. *Tesourus van Afrikaans*. Uitgebreide, hersiene tweede uitgawe. 2020, xxxii + 992 pp. ISBN 978-1-868-90213-2 (hardeband). Kaapstad: Pharos Woordeboeke. Prys: R450.

María José Domínguez Vázquez, Mónica Mirazo Balsa and Carlos Valcárcel Riveiro (Editors). *Studies on Multilingual Lexicography*. 2019, vi + 240 pp. ISBN 978-3-11-060467-2, ISSN 0175-9264. Lexicographica. Series Maior 157. Berlin/Boston: Walter de Gruyter. Price: €99.95. (Review in this issue.)

Michael Klotz and Thomas Herbst. *English Dictionaries: A Linguistic Introduction*. 2016, 287 pp. ISBN 978 3 503 16751 7. Berlin: Erich Schmidt Verlag. Price: €19.95. (Review in this issue.)

Marie-Claude L'Homme. *Lexical Semantics for Terminology: An Introduction*. 2020, xxi + 264 pp. ISBN: 9789027204677 (Hardback). Terminology and Lexicography Research and Practice. Volume 20. Amsterdam/Philadelphia: John Benjamins Publishing Company. Price € 99.00. (Review in this issue.)

Gaoretelelwe Benjamin Phuti (Chief Editor). *Oxford Bilingual School Dictionary: Setswana and English*. First SA edition. 2020, 624 pp. ISBN 978 0 19 042271 4 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R192.95.

Van Wyk, E. (Compiler). *The Talking Dictionary*. 2019, 288 pp. ISBN: 978-1-920217-74-7. Pretoria: Briza Publications. Price R1250.00. (Additional language sets: R280.00 each.) (Resensie in hierdie nommer.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://lexikos.journals.ac.za/>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, terminologie, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder een van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evalueerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index (ISI)*.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare

Lexikos is sedert volume 28 slegs elektronies beskikbaar op <http://lexikos.journals.ac.za>. Geen oordrukke of eksemplare is dus beskikbaar nie.

7. Uitnodiging en redaksionele adres

Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en verkieslik in elektroniese formaat aan die volgende adres te stuur: lexikos@sun.ac.za, of Die Redakteur: LEXIKOS, Buro van die WAT, Postbus 245, 7599 STELLENBOSCH, Republiek van Suid-Afrika.

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie 8 000 woorde oorskry nie.

Elke artikel moet voorsien wees van 'n opsomming van ongeveer 200 woorde en ongeveer 10 sleutelwoorde in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde in Engels. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n vertaling van die oorspronklike titel bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting. Buiten in spesiale gevalle moet verwysings na *Lexikos*-artikels tot twee of drie per artikel beperk word. Uitsonderings moet met die redakteur van *Lexikos* uitgeklaar word. Dit word gedoen om die status van *Lexikos* in verskeie internasionale indekse te behou.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website: <http://lexikos.journals.ac.za/>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.

(4) **Projects:** Discussions of lexicographical projects.

(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.

(6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index (ISI)*.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordboek van die Afrikaanse Taal. Authors are free, however, to use their

material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies

Lexikos is only available electronically on <http://lexikos.journals.ac.za> from volume 28 onward. No offprints or copies are available.

7. Invitation and editorial address

All interested authors are invited to submit contributions, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH, Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Format and length of articles

Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately 10 **keywords** in the language in which it is written, as well as **in English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

2. Graphics

Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos*, namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details. Self-references to *Lexikos* should be limited to two or three per article, except in exceptional circumstances. Exceptions should be cleared with the editor of *Lexikos*. This is done to preserve the status of *Lexikos* in various international indices.

4. Notes/footnotes/endnotes

Notes must be numbered consecutively by superscript numbers and grouped together at the end of the manuscript under the heading **Endnotes**.