

USING CORPORA TO INFORM TEACHING PRACTICE IN GERMAN STUDIES

Gwyndolen J. Ortner & Undine S. Weber
Rhodes University

Learner corpus research seeks to describe and thereby better understand learner acquisition, thus informing better teaching practice and creating an important bond between corpus linguistics and second/foreign language (L2) research. While much research exists for the study of L2 English, there is little research for the study of L2 German. This study explores the implementation of a corpus-based writing course in German studies at Rhodes University in South Africa with students at third-year level who were learning to write longer texts that present and defend an opinion. In order to evaluate the success of the writing course in teaching specific academic collocations, a diachronic collection of learner writing was undertaken. Written assignments were used to create a learner corpus, RUDaF (Rhodes University Deutsch als Fremdsprache).

The creation of a learner corpus of German before and during the writing course served as a reflective tool to aid in the evaluation and improvement of corpus-based teaching materials. This is exemplified through the students' use of three specific opinion-related everyday-academic German collocations before and after the writing course. The learner corpus revealed a marked improvement in the use of taught collocations before and during the writing course. These methods provide insights for other language educators who wish to implement corpus-based methods for improving writing in a second or foreign language, and for all educators wanting to assess improvements in student writing.

Keywords: academic writing in German as a foreign language; corpus linguistics; data-driven learning, formulaic language

INTRODUCTION

German studies at Rhodes University in South Africa consists not only of the teaching of grammar and vocabulary, but also literature, cultural studies and translation. Traditionally, due to the fact that German studies at Rhodes University mostly has *ab initio* (beginner) students, little focus has been placed on what would be deemed 'academic literacy' skills in German. This is because, in the first two years of academic study, students are still in the A1 and A2 stages of language acquisition according to the Common European Framework of Reference (CEFR), and a specific focus on formal academic writing skills has not always been possible due to time and curricular constraints. Studies that have introduced regular writing activities in the foreign language classroom have provided much qualitative evidence of an improvement in writing fluency and confidence (see Estes et al., 1998; Homstad & Thorson, 1996). However, such studies (e.g., Homstad & Thorson, 1996) often struggle to provide quantitative evidence to support their qualitative perceptions of learner improvement, due to methodological barriers. This study seeks to improve on this and show how a learner corpus may be used to provide quantitative evidence of specific areas of improvement in learner writing of everyday academic texts, with regard to the improved use of high-frequency

academic collocations (formulaic phrases). The study also shows how an academic writing course at the B1 level (CEFR), which draws on current corpus-based research on high frequency academic collocations in German, can provide effective results in improving learner academic writing at that level.

What is a corpus? Why are corpora relevant to language teaching?

A linguistic corpus is a collection of texts, principally selected, stored on a computer, which can be processed by accompanying computer software (McCarthy & O’Keeffe, 2010: 9; McEney & Hardie, 2012(a): 1; O’Keeffe, McCarthy & Carter, 2007: 7). Corpora allow linguists to rapidly deduce patterns in language, leading to better descriptions of language and many practical implications for all branches of linguistics and lexicography (McEney & Xiao, 2010: 364). Language teaching has also been increasingly influenced by insights gained from the investigation of corpora (Gabrielatos, 2005: 1; McEney & Xiao, 2010: 364), and many specialised types of language corpora have been developed to suit specific pedagogical needs. Some corpora have been developed to provide data on language usage which can be used for teaching purposes (Gabrielatos, 2005: 1). Another type of specialised corpus is the learner corpus, which is a principled collection of the writing (or transcribed speech) produced by language learners (Granger, 2003: 538). Learner corpora can be analysed on their own, or compared to mother-tongue corpora, a process which enables researchers to identify areas of over- and underuse of specific lexico-grammatical structures, and thereby create better teaching materials that address specific re-occurring learner errors (Granger, 2004: 127-128; Krummes & Ensslin, 2014: 192).

Context of the study

Even though much corpus-based research has been undertaken for learner English (the type of English produced by learners) (see Goossens & Granger, 2017 for a comprehensive online list of existing learner corpora around the world), there is considerably less research available for the study of learner German and other foreign languages worldwide (Jaworska et al., 2015: 501, Krummes & Ensslin, 2015: 110). This study seeks to provide insights into the creation of a learner corpus of German in South Africa, and its usefulness as a tool to reflect on learner writing. To do so, it assesses the effectiveness of, in this case, a corpus-based every-day academic writing course for 3rd year (B1¹) level German students at Rhodes University.

A writing course was instituted in 2015 to teach students typical everyday academic German words and phrases (collocations/formulaic sequences) based on German mother-tongue corpus evidence. The phrases taught were adapted from those identified by the WHiG project (‘What’s Hard in German’) (Krummes & Ensslin, 2015), based on a comparative analysis of student writing in the UK, contrasted with a first language corpus of mother-tongue high school level argumentative German writing, FALKO L1 (*Fehlerannotiertes Lernerkorpus/error-annotated learner corpus L1 – first language*) (Lüdeling, et al., 2008). This method of language teaching had never been undertaken at Rhodes University before, and as such had an effect on teaching practice in the section, fostering an awareness of the importance of the role of collocation in language teaching. Further postgraduate research is currently being undertaken in this area.

Participants in the course were tasked to write short assignments in German at regular intervals both before and during the writing course. The writing pieces were used to create a learner corpus, which we named RUDaF (Rhodes University *Deutsch als Fremdsprache*²).

The RUDaF learner corpus was then used as a tool to evaluate whether what was taught in the writing course was internalised and made use of by students in their own writing output by comparing writing before the course with writing during the course.

The overarching question for this research was: What evidence can a learner corpus provide to support qualitative notions of improvement in learner everyday academic writing as the result of a writing course? This study presents specific extracts from the learner corpus as evidence which to support an evaluation of the usefulness and effectiveness of the method of teaching everyday academic collocations in German. The study also presents elements of the didactic methodology employed for instituting a corpus-based writing course. This is done in order to provide other language educators with empirical evidence to support an argument for the use of corpus-based teaching materials in the language classroom, and an argument for the use of self-built learner corpora as an evaluative tool.

BACKGROUND AND LITERATURE

Collocations and their importance in language teaching

Foreign language learners in the university context (such as our students at Rhodes University) must make fairly rapid linguistic advancements in a short period of time, given the limited amount of contact hours allocated to language teaching. The goal of the foreign language learner is to achieve as close to mother-tongue-like competence as possible in three years. To do this, learners must master the linguistic patterns of the target language (Wray, 2000: 463). The reason for this is that words have a tendency to reappear with one another, in fixed sequences also known as collocations (Krummes & Ensslin, 2015: 112) or formulaic sequences (Conklin & Schmitt, 2012: 45; Wray, 2000: 465). These are the building blocks of language, and often constitute single choices for a mother-tongue speaker (Sinclair, 1987: 320) for whom words do not exist in isolation, but in psychological relation to other words (Hoey, 2005: 5). Collocations therefore play an important role in language use for mother-tongue speakers, as they are theorised to be stored as single units in the long-term memory, thus allowing for more space or processing capacity in the working memory (Conklin & Schmitt, 2012: 45). L2 research has shown that learners may rely on memorising formulaic sequences, later returning to these for grammatical analysis³ (Ellis, 1996: 91; Ellis, 2005: 211). Thus, learners need a good knowledge of formulaic sequences to assist in speaking and writing fluency, as well as a knowledge of specific grammatical rules to assist in complexity and accuracy of learner language production (Ellis, 2005: 211).

According to Wray (2002: ix), formulaic language is one of the greatest ‘stumbling blocks’ for second language learners to sounding like mother-tongue speakers, even for advanced learners. Extensive exposure to natural language use can help to make a language user aware of the patterns in language, and help learners to acquire a repertoire of formulaic expressions. However, in a foreign language learning environment exposure to the target language is limited to the classroom where, due to time and curricular constraints, most language lessons are focused on communicative functions, and teaching and practising grammatical rules. Advanced students (such as those in the third year, at B1 level) may thus have a good grasp of grammatical rules, but may rely on ‘overusing “safe”[,] but repetitive language strings’ (Krummes & Ensslin, 2015: 110), or rely heavily on their own creativity, creating grammatically sound strings that would not be used by mother-tongue speakers, as shown by Pawley and Syder (1983). This may be seen particularly in the case of academic writing.

Collocation in everyday academic language

‘Everyday academic language’, or as coined by Ehlich (1995) in German *alltägliche Wissenschaftssprache*, is a semantic category used specifically to describe, and thereby analyse, the use of everyday, ordinary German vocabulary for the purpose of academic writing (Ehlich, 1995; Skrandies, 2011: 99; Steinhof, 2007). The vocabulary which is typical of this semantic category is characteristic of both oral and written communication across most academic disciplines (see Ehlich, 1995; Steinhof, 2007). This entails an often-metaphorical use of vocabulary items (Fandrych et al., 2012), which is to say that the vocabulary has undergone a change in function (Skrandies, 2011: 100). Examples of this include noun-verb collocations such as *einer Frage nachgehen* (to pursue a question) or *ein Problem beleuchten* (to ‘illuminate’ a problem, or rather to gain a deeper understanding of a problem).

Language learning textbooks often do not address these types of formulaic phrases or collocations explicitly, even though there is current research on their importance for language learners (Krummes & Ensslin, 2015: 116). Thus, the creation of additional materials is necessary. This can be done by providing learners with concordance lines of a keyword in context (KWIC) from a corpus of mother-tongue writing to expose learners to real-life language data containing instances of this specialised language usage, normally referred to as data-driven learning (Johns & King, 1991: iii). In order to view collocations within a corpus, a researcher or teacher must make use of specialised software called ‘concordancers’.

When using concordance software, patterns in language become clear through a combination of frequency lists, the KWIC view and extracted word clusters. Frequency refers to the number of times a particular word ‘type’ occurs in a corpus of a total amount of ‘tokens’ (normally the total word count). For example, the most frequent ‘type’ that occurs in most English corpora is the definite article ‘the’ (Baker, 2006: 47). A frequency list can show the learner or teacher what words appear most often, and give an indication of what words to focus on. When a teacher or learner has identified a word of interest, they can then use the concordancer to reveal which words often co-occur with the chosen item, and in what set patterns (i.e., collocations) (Durrant, 2009: 158). Some concordancers allow one to extract word clusters, which can then be sifted through to find, in academic texts for example, clusters relevant to discourse structuring devices (Krummes & Ensslin, 2015: 116).

Collocational patterns revealed through concordancing software can highlight the formulaic sequences and patterns in language use with regard to a specific lexical item in the corpus, making these explicit for the teacher and language learner. Of course, if working directly from a corpus, both teachers and learners need to have some basic training in how to use the software, and have access to some written data stored electronically. Alternatively, teachers can create paper-based exercises for students or make use of corpus-based worksheets already developed by others.⁴

Learner corpora, a tool for assessment

Language input is thus highly important, but so are opportunities to practise language. The benefits to this are numerous, as learners are forced to pay attention to grammar in their own language production. This reinforces learners’ existing knowledge, which then leads to automation (Ellis, 2005: 218). Teachers can assemble their students’ writing outputs to spot trends in their writing and areas which may need attention. This is an everyday task, and there is ‘nothing new’ for the language teacher about collecting and evaluating learner output

(Granger, 2004: 123; Nesselhauf, 2004: 128). What makes a learner corpus of advantage to a teacher is the size, variability and automation (see Granger, 2004: 123-128). A teacher may gain subject-relevant insights through looking at even a small corpus of learner writing, and spot patterns that may otherwise be difficult and time-consuming to pinpoint, including what students are avoiding in their writing (Ragan, 1996: 214).

Thus, it is important that learner corpus data not simply be a collection of errors; a learner corpus must consist of ‘continuous stretches of discourse which contain both erroneous and correct use of the language’ (Granger, 2002: 9). Learner corpora can only be useful in describing specific types of learner language if they are collected with specific design criteria in mind (Granger et al., 2002: 9; Ragan, 1996: 214). The variables which are normally controlled for in learner corpus design include those pertaining to the learner (for example learning context, mother tongue, level of proficiency and command of other foreign languages), and those relating to the task setting or language situation (for example time limit, use of reference tools, exam and audience/interlocutor) (Flowerdew, 2008: 5; Granger et al., 2002: 9; Nesselhauf, 2004: 130). Where these are controlled for, a researcher can make claims about the language used within the design criteria, which can then be compared to other learner language collected with similar or different criteria (Nesselhauf, 2004: 130). The following method presents a learner corpus with two sub-corpora whose design criteria differ according to one aspect: writing before and writing during a writing course. This allowed the researchers to assess differences in the use of particular academic collocations.

METHODOLOGY

The following methodology expands on the creation of the learner corpus, RUDaF, as an assessment tool to show, based on quantitative evidence, whether student writing improved as a result of the writing course. The method followed in designing the writing course is also expanded upon. The writing course followed a corpus-based approach to teaching academic collocations. The teaching materials for the writing course were adapted from Krummes and Ensslin (2015).

Considerations in building the RUDaF learner corpus

The design of the RUDaF learner corpus was led by its purpose, which was to investigate the level of writing of learners of German at Rhodes University, and to compare the writing before and after the instituted writing course in order to provide evidence of exactly where and how writing had improved or not. The following criteria (following on from Granger, 1998: 7) were controlled for in the RUDaF learner corpus design and recorded:

Factors pertaining to the learner:

- Learning context: German as a foreign language in South Africa.
- Level: intermediate learners of German from Southern Africa, with a minimum CEFR level A2 and a maximum level B2 (Council of Europe, 2012). Two to four years prior exposure to German.⁵
- Age: early 20s.
- Sex: male and female.
- Mother tongues: English, Afrikaans and African languages (mainly Nguni languages).
- Region: Southern Africa.
- Other foreign and second languages: French; Afrikaans and African languages.

- Timeframe: 2013-2015.

Factors pertaining to the task setting:

- Typewritten texts of approximately 300 words each. No spoken data.
- Argumentative or opinion-based essay topics, and some creative writing pieces.
- Texts written at home – the use of dictionaries and web-based sources and resources was permitted. Students were required to reference sources, and submit writing tasks electronically via email.
- Files saved as UTF 8 text format.
- No annotation.⁶
- Diachronic data collection: text collection covered a period of time with multiple sets of data collected from the same learner over the course of a year, allowing for the comparison of writing before and after corpus-based instruction.

The raw data, submitted electronically, were cleaned and converted to text format, UTF 8.⁷ File names were allocated showing participant number, topic number and year, as in the following example: P01_T01_2015. Metadata such as this is an invaluable accompaniment to the researcher using corpus linguistic methodology, as this information becomes a kind of bibliography, so that the original text can easily be found and referred to within any study of the corpus as a whole (Granger, 1998: 12).

Corpus composition

As the corpus data collection took place diachronically, the learner corpus is divided into two parts in order to draw internal comparisons between participants. The first section of the corpus (RUDaF-Pre)⁸ was made up of the 95 pre-writing course text submissions (17 998 tokens) and included the 45 texts collected in 2013 for the pilot study, as well as 50 texts collected in 2014 before the writing course. The second part of the corpus (RUDaF-During) consisted of a total of 56 texts (15 989 tokens), of which 38 texts were collected in 2014 during the pilot writing course and 18 texts were collected during the 2015 corpus-based everyday academic German writing course.

Data submitted before the writing course beginning in the second semester of 2014 (RUDaF-Pre) were compared, using corpus tools, to data collected during the writing course (RUDaF-During). This was achieved by comparing the use of specific lexical items and academic collocations associated with everyday academic German (which were introduced during the writing course) in the two data sets. This aimed to show whether or not corpus linguistic methods had helped with the writing process, and offered an objective method of assessing the improvement of students' academic writing skills in German, in conjunction with qualitative questionnaires and interviews. While our learner corpus may be viewed as relatively small at a total size of 33 987 tokens and 5 500 types, it was sufficient for the purposes of examining the use of the high-frequency typical academic German words introduced during the writing course (see Granger, 1998: 9; Ragan, 1996). Given the relatively small class sizes in German studies at the third-year level, no control group was possible to validate findings. Our data can, however, be compared to the findings of the WHiG project with regard to the overuse and underuse of certain collocations in learner academic writing (Jaworska *et al.*, 2015: 511).

The writing course

Per Linguam 2018 34(2):69-83
<http://dx.doi.org/10.5785/34-2-789>

The writing course made use of the everyday-academic corpus-based exercises developed by the authors of the WHiG project. These resources are freely available online (see Krummes & Ensslin, 2015). These paper-based exercises present typical everyday academic German words and phrases. The handout provides five main keywords, followed by examples of their use in five different formulaic phrases/collocations, with three to five examples of usage (corpus-based) for each phrase to provide learners with enough context (Krummes & Ensslin, 2015: 118). In so doing, the WHiG authors provide students of German with idiomatic alternatives to raw and ineffective (if grammatical) translations when writing an academic assignment. Participants were encouraged to use these newly learnt collocations in the weekly/fortnightly writing assignments. The typical academic words and phrases were not introduced in other lectures, and so their use in learner writing may be seen as either a product of the writing course or a result of self-study. By examining the participants' use of these words before and during the writing course, we gained insight into how the writing course impacted upon their knowledge and ability to use these words, and their ability to write in a way closer to that of mother-tongue German writers.

The paper-based exercises were supplemented with an exploration of real-life mother-tongue German corpus data obtained from the FALKO L1 corpus, available under a Creative Commons license, and the freely available and searchable online parallel corpora, Linguee.de.⁹ The FALKO L1 corpus (developed by Lüdeling et al., 2008) was deemed appropriate for our context as it was perceived as neither so large that it would drown the learners in data, nor so small that the instances for which we were searching would not be present.¹⁰ Both the WHiG handouts and the FALKO L1 corpus were chosen for their high degree of relevance to the learning context, as they provide evidence of everyday academic German written by mother-tongue speakers of German. The web-based parallel corpus Linguee.de, while not only composed of academic texts, is an additional useful resource as it provides a general translation of a search item. As a class exercise, we also had the participants create their own small corpus of German writing in order to familiarise them with corpus software, as well as to provide an extra resource for learning. The free corpus linguistic software programme, AntConc ver. 3.4.4, developed by Anthony (2014), was used to perform all operations.

An important aspect of the writing course was that writing topics were issued each week and students were able to put the newly learnt collocations into practice. This enabled the students to make the move from language research to language production, an important aspect of language learning which is often underscored in corpus-based learning (see Ellis, 2005: 218; Oghigian & Chujo, 2010).

DISCUSSION

Giving an opinion, exploring the use of the words *Meinung*() [opinion (plus inflected ending)], *Erachten*() [in my mind (plus inflected ending)] and *Ansicht*() [view/opinion (plus inflected ending)]

The aim of the writing course was to enable students to write longer argumentative writing pieces in German, and to be able to express their opinions in a more academic way. The following discussion presents the students' use of the opinion-based typical academic German words *Meinung*(), *Erachten*() and *Ansicht*() in their writing pieces before and after the instituted corpus-based writing course. Research conducted in the UK found that British learners of German overused certain phrases in their writing as compared to mother-tongue

writers (Krummes & Ensslin, 2014). When trying to express an opinion, British learners of German overused the word *Meinung*() (as used in the dative phrases *meiner Meinung nach/ der Meinung sein, dass* [in my opinion/ to be of the opinion that]) in their essays. This trend was also found with our own students in the RUDaF-Pre corpus.

In the handout, the words *Erachten*() and *Ansicht*() are presented as alternative forms to use when presenting one's own opinion. *Erachten*() is used in the genitive form *meines Erachtens*¹¹ and the abbreviated form *m.E.* when one wants to express one's opinion in the main body of an essay (Krummes & Ensslin, 2014). *Ansicht*() is also used to express one's own opinion in the body of an essay, and may be used in a number of ways: for example, *meiner Ansicht nach/ der Ansicht sein, dass/ die Ansicht vertreten, dass* or *die Ansicht teilen, dass* [in my opinion/ to be of the opinion that or to share the opinion that].

Participants expressed their opinions in a limited way before the writing course

In RUDaF-Pre, participants made use of the word *Meinung*() on 29 occasions (1.61 times per 1 000) when trying to express their own opinion. The word *Meinung*() was used in various phrases, with the most prevalent collocation being those identified by Krummes and Ensslin (2014) as overused by British students, namely: *nach meiner Meinung/ meiner Meinung nach* (4 instances) followed by *ich bin der Meinung, dass* (3 instances). However, aside from these, participants used many other phrases with *Meinung* before the writing course, most of which were incorrect, illustrating a need for direct clarification of the correct collocations of *Meinung*(), as was performed during the writing course.

For example, Participant 2 (P2) made use of the incorrect collocation *ich bin *an der Meinung, dass* in three different writing pieces, inserting the preposition *an* unnecessarily. *Ich bin/er war *von der Meinung, dass* also made an appearance in the RUDaF-Pre (P3 and P7, T6_2013), which can be seen as a direct translation from the English phrase 'I am/he is of the opinion that'. Variations of the incorrect phrase *in *meine/r Meinung* were the most prevalent learner errors involving *Meinung* in RUDaF-Pre (four instances). The incorrect phrases *Meine [] Meinung *über* (P14_T11_2014) and *in *meine Meinung *auf* (T6_2013) were also used by participants as alternatives to *meiner Meinung nach*. P20 (T9_2014) made use of the phrase **streit für die Meinung, dass* (a failed translation of the English: to argue for the opinion that).

Examples from the corpus of both correct and incorrect uses, may be seen in the concordance lines of the KWIC sort below:

Concordance Hits 29		
Hit	KWIC	File
1	Meiner bescheidenen Meinung nach, kann platonische Freundschaft	P3_T3_2013.txt
2	was Wahrheit hat. Ich bin an der Meinung, dass die Unterhaltungsindustrie e	P2_T3_2013.txt
3	schwierig auf, aber ich bin an der Meinung, dass die Aussprache auch schwer	P2_T4_2013.txt
4	lichen Welt wohnen, bin an der Meinung, dass typischen Geschlechterrolle	P2_T7_2013.txt
5	uerger auszustellen. Ich bin der Meinung, dass die Miniserie ein neue Darst	P2_T8_2014.txt
6	Macht sollte haben. Ich bin der Meinung, dass der Staat limitierte Befugnis	P3_T6_2013.txt
7	tion ihn ermuterte. Ich bin der Meinung, dass wir in ander Luther in unsere	P19_T11_2014.b
8	ausgegeben hat. Er war von der Meinung, dass der amerikanischer Publikur	P3_T6_2013.txt
9	e zu verhindern. Ich bin von der Meinung, dass Spionage waere kontraprodu	P7_T6_2013.txt
10	en. Die Medien beeinflussen die Meinung des Volks und die fortgesetzte Bes	P2_T7_2013.txt
11	Nostalgie. Barney streit fuer die Meinung, dass die Nostalgie in Good Bye Le	P20_T9_2014.txt
12	in ihrer Freizeit. Daher wird die Meinungsfreiheit nicht zu viel beschaenkt.	P6_T5_2013.txt
13	zu machen. Ich denke auf diese Meinungen, wenn ich ueber Luther spreche	P18_T11_2014.b
14	it diesem Aufsatz meine eigene Meinung vorzubringen, wie es zu grammati	P15_T10_2014.b

Figure 1: Concordance of *Meinung*() in RUDaF-Pre, examples 1-14 of 29

The correct use of prepositions in conjunction with the word *Meinung*() thus appeared to pose a particular difficulty for our learners (as seen above), and this was addressed in the writing course.

The word *Ansicht*() was used three times in RUDaF-Pre (P1_T7_2013), as in the examples below. „Das ist meine allgemeine Ansicht und nicht unbedingt die Norm heute“; „aber seine Ansichten und Kommentare *wegen Judaismus war intolerant“; *„als eine Methode der folgenden denen gegensaeztliche Ansichten“. These usages of *Ansicht* are not incorrect, but they do not reflect the most frequent collocations of *Ansicht*() used by mother-tongue writers of German.

Concordance Hits 3		
Hit	KWIC	File
1	das Haus. Das ist meine allgemeine Ansicht und nicht unbedingt die Norm	P1_T7_2013.txt
2	er folgenden denen gegensaeztliche Ansichten und dann bestrafen sie genu	P7_T6_2013.txt
3	tionaere Sache gatan hat, aber seine Ansichten und Kommentare wegen Ju	P2_T11_2014.txt

Figure 2: Concordance of *Ansicht*() in RUDaF-Pre

The word *Erachten*(), a good alternative to *Meinung*, was not used at all before the writing course.

Participants expressed their opinions more correctly, and more diversely, during the writing course

During the writing course, participants were able to express their opinions more clearly by making use of more diverse typical academic German words, and making use of more mother-tongue-like, idiomatic collocations significantly more often. The first instance in which this can be seen is in the use of the word *Meinung*, which was used significantly less during the writing course (13 times (0.81 times per 1 000) as opposed to 29 times in RUDaF-Pre), and with a far smaller margin of error. The most frequent (and correct) collocational pattern of *Meinung* was *meiner Meinung nach/ nach meiner Meinung*.

While *Meinung* was significantly overused in RUDaF-Pre, *Erachten*() was not used at all and was presented as an alternative form during the writing course. The phrase *meines Erachtens*, which was absent in RUDaF-Pre, was used during the writing course a total of 11 times (0.69 times per thousand) by six of the participants (P12, P13, P15, P16, P18, P21) who used this phrase to express their opinion in their writing. The abbreviation *m.E.* was used twice by P13 in 2015 (T16, T18). What is more, this phrase was only ever used in its correct genitive form, as seen in Figure 3 below.

Concordance Hits 11			File
Hit	KWIC		
1	gelesen habe, dachte ich, dass meines Erachtens das wirkliche Ende auch so kreativ		P13_T18_2015.txt
2	enomen, aber auch nicht unbekannt. Meines Erachtens , In Suedafrika ist es haeufiger Krim		P12_T17_2015.txt
3	e leiden unter den Analphabetismus. Meines erachtens , ist es voellig untragbar. Fast ein		P12_T16_2015.txt
4	ist, um Deutsch zu lernen. Meines Erachtens ist Deutsch mehr strenger mit der		P18_T14_2014.txt
5	Meines Erachtens ist es von groesster Wichtigkeit, so		P21_T14_2014.txt
6	warzen Mensch geringwertig war, ist meines Erachtens schwierig . Nehmen wir als Beispiel		P13_T13_2014.txt
7	Meines Erachtens , sind die groessten Unterschiede z		P12_T15_2015.txt
8	teiligte Jungen. (Motakef:2008, 146) Meines Erachtens sind viele Auslaender nach Deutscl		P13_T16_2015.txt
9	von ihnen „haeufig den Tatort“. Meines Erachtens sind die Deutschen nicht von Tator		P21_T17_2015.txt
10	olfen mehr akademisch zu schreiben. Meines Erachtens werde ich fuer viele Zwecke meine		P15_T14_2014.txt
11	Aufbau mehr als die Englische. Meines Erachtens wird den deutsche Leser mehr von		P16_T14_2014.txt

Figure 3: Concordance of *Erachten*() in RUDaF-During

Similarly, the word *Ansicht*() was used significantly more often (on 17 occasions as opposed to 3) in RUDaF-During than in RUDaF-Pre. The collocations used were *meiner Ansicht nach* (8 of 17 instances or 47.05%), *Ich vertrete/ [subject] vertrat die Ansicht* [I am of the opinion/ [subject] is/was of the opinion] (5 of 17 instances or 29.41%), and *nach Ansicht [subject, but not a name]* (4 instances or 23.52%). This showed a greater diversity in the students' ability to express their opinions, and a good implementation of the opinion-related typical academic words taught using corpus-based methods in the writing course.

Concordance Hits 17		
Hit	KWIC	File
1	egler, 2013). Amrhein vertrat die Ansicht , dass Tatort so beliebt bei der	P15_T17_2015.txt
2	3). Andere Deutschen vertrat die Ansicht , dass die Deutschen krimibes	P15_T17_2015.txt
3	Ich vertrete die Ansicht , dass es sehr gut Deutsch zu	P18_T14_2014.txt
4	s.(Wolters:2005) Ich vertrete die Ansicht , dieses Problem bedeutet fun	P13_T16_2015.txt
5	Motakef (2008, 145) vertrat die Ansicht , Einwanderer werden automa	P13_T16_2015.txt
6	Meiner Ansicht nach , ist das Schreiben auf D	P12_T14_2014.txt
7	ders interessanterweise. Meiner Ansicht nach , sind die Endungen, die	P12_T18_2015.txt
8	ihn, besser zu verstehen. Meiner Ansicht nach schreiben erst auf Deut	P13_T14_2014.txt
9	n (Axinia:2008), koennte meiner Ansicht nach ein Grund sein, warum	P13_T15_2015.txt
10	endet das Buch zu rasch. Meiner Ansicht nach koennte das Ende ein bi	P13_T18_2015.txt
11	nal wie ein 12-Jaehriger. Meiner Ansicht nach haben die Hausaufgabe	P15_T14_2014.txt
12	hichte wirklich schwach. Meiner Ansicht nach , erstaut diese Geschicht	P15_T18_2015.txt
13	en den swachen Schluss. Meiner Ansicht nach habe ich und manche n	P15_T18_2015.txt
14	llan in Rautenbach, 2013). Nach Ansicht der Kritiker wurden Krimis di	P15_T17_2015.txt
15	erreich gesprochen wird ist nach Ansicht von einem Blogger (HF:2014	P13_T15_2015.txt
16	, aber wir treffen ihn nie. Nach Ansicht von Van Dine (1928) muss d	P13_T18_2015.txt
17	wollten es nicht zugeben. Nach Ansicht Ziegler (2015) koennen Anal	P13_T16_2015.txt

Figure 4: Concordances of *Ansicht*() in RUDaF-During

Summary of the results

A test of log-likelihood confirmed the significance of the typical everyday academic German opinion words used by participants before and after the writing course.¹² It was based on four simple figures: the frequency of a word in each of the two (sub)corpora, and the total frequency of each (sub)corpus (number of opportunities that the term could appear) (McEnery & Hardie, 2012b). The log-likelihood score must be ‘above 3.84 for the difference to be significant at the $p < 0.05$ level’ (McEnery & Hardie, 2012 (b)). The log-likelihood of the typical academic words discussed above were as follows: *Meinung*(), 4.51; *Erachten*(), 16.59; *Ansicht*(), 12.54.

These scores show that the typical everyday academic German words introduced during the writing course were used significantly more often during the writing course as compared to before the writing course, with the exception of *Meinung*(), which was used significantly less as a result of the writing course, which encouraged variation of opinion-giving collocations.

CONCLUSION

The creation of a learner corpus provided a useful tool for evaluation of the effect of the instituted writing course. An example of the process of learning new variants has been shown in the overuse (and incorrect use) of *Meinung*() and its collocations before the writing course, and the more correct use of *Meinung*() during the writing course, with the alternatives *Erachten*() and *Ansicht*() used equally often where they had not been used at all before the writing course. Creating a diachronic learner corpus as a tool for assessment can thus provide educators with empirical evidence to highlight exactly how and where learners are improving in their writing. This has implications for all language educators, as well as for academic writing centres, who often need to provide evidence of learner improvement.

Additionally, this study has shown that a corpus-based writing course can provide learners with useful tools for improving writing. Once students have learnt how to use corpus tools, they themselves can investigate other alternative words and collocations to use in their writing (for example by using Linguee.de or by creating a small corpus for personal use), thus creating long-term benefits for participants through understanding corpus concepts. This fostering of independence for writers may enhance writing experiences and encourage writers to take more responsibility for their own learning (see Yoon, 2008, who reports similar findings).

Thus, while this article has mainly concerned itself with quantitative analysis, there were qualitative findings as well regarding participants' perceptions of their own writing improvement, which the participants attributed directly to the collocational exercises. These findings might be demonstrated in a further publication. However, it is worth mentioning here already that contributing to a corpus for further use and to the hard-copy, in-house publication of 'Schreiben für RUDaF' ('Writing for RUDaF', 2015), with their names as contributors on the cover page, students felt a sense of achievement. We, the authors, like to believe that this integrative teaching measure successfully acknowledged students as 'co-creators of knowledge' (Bovill et al., 2011: 133).

ENDNOTES

¹ B1 is a standardised language level according to the CEFR framework, which stipulates certain communicative competencies and level of ability (Council of Europe, 2012).

² *Deutsch als Fremdsprache* is German as a foreign language, abbreviated *DaF* in our discipline.

³ For example, German studies students learn the greeting: *Wie geht es dir/Ihnen?* (Literally: How goes it to/with you? Meaning: How are you?) in their first week of language instruction, and only six months later, when they encounter the dative, do they retrospectively apply the new grammatical concept to this formulaic sequence.

⁴ See Krummes and Ensslin (2015: 116-120) for an example of how to design corpus-based worksheets for everyday academic vocabulary for German as a foreign language.

⁵ The contact hours which the German studies section at Rhodes University has with its students are as follows: first year, 130 hours (116 language teaching); second year, 143 hours (91 language teaching; excl. translation); third year, 146 hours (78 language teaching; excl. translation); fourth year, 156 hours (78 language teaching; excl. translation).

⁶ Given the high rate of morphological, lexical and syntactical errors, particularly in the first set of writing tasks, it was decided not to annotate the RUDaF learner corpus. This decision was based on insights provided by Van Rooy and Schäfer (2002), who found that the success rate of part of speech annotators decreases as the number of learner errors increases. As the RUDaF learner corpus is relatively small in size, it was possible to manually observe these errors in the analysis, which focused specifically on the use of words introduced in the writing course.

⁷ Data were taken in their original form, without the corrections made in track changes.

⁸ 'RUDaF-Pre' was so named as it consisted of writing collected before the writing course. 'RUDaF-During' was so named as it consisted of writing collected during the writing course.

⁹ Linguee.de is an online dictionary that consists of a large corpus of translated official documents that show users how the word(s) for which they are searching are used. However, most of the translations have not been vetted, and especially in documents of the European Union there are some very un-idiomatic translations, clearly undertaken by non-mother tongue speakers of the target language.

¹⁰ In future we would recommend making use of a larger corpus in order to see sufficient examples of specific collocations.

¹¹ Obviously, another possessive pronoun could have been used, like *unseres Erachtens* [in our opinion], but we focused on the first person singular in these writing assignments.

¹² The Lancaster University log-likelihood wizard (created by Paul Rayson) was used in this analysis as it allows one to perform tests for a significant difference in frequency (of a specific word or sentence) between two corpora, and is easily accessible online (see the Lancaster University website: <http://corpora.lancs.ac.uk/clmtp/2-stat.php>).

ACKNOWLEDGEMENTS

The authors wish to express their gratitude for the financial support of the Ada & Bertie Levenstein Bursary (Rhodes University) and thank Sylvia Jaworska (Queen Mary/ Reading University) for her invaluable input regarding corpus-based data collection and interpretation.

REFERENCES

- ANTHONY, L. 2014. *AntConc* (version 3.4.4) [Computer software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/> [Accessed: 15 November 2017].
- BAKER, P. 2006. *Using corpora in discourse analysis*. London/New York: Continuum.
- BOVILL, C & A COOK-SATHER & P FELTEN. 2011. Students as co-creators of teaching approaches, course design, and curricula: implications for academic developers. *International Journal for Academic Development*, 16(2):133-145.
- CONKLIN, K & N SCHMITT. 2012. The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, 32:45-61.
- COUNCIL OF EUROPE. 2012. *Common European Framework of Reference (CEFR) for languages: learning, teaching, assessment*. Available from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp [Accessed: 14 June 2013]
- DURRANT, P. 2009. Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3):157-169.
- EHLICH, K. 1995. Die Lehre der deutschen Wissenschaftssprache: Sprachliche Strukturen, didaktische Desiderate. [The teaching of German scientific language: speech structures and pedagogic requirements.] In HL Kretzenbacher & H Weinrich, (Eds), *Linguistik der Wissenschaftssprachen*. Berlin: de Gruyter. 325-351.
- ELLIS, N. 1996. Sequencing in SLA: phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18:91-126.
- ELLIS, R. 2005. Principles of instructed language learning. *System*, 33:209-224.
- FANDRYCH, C, C MEIBNER & A SLAVCHEVA. 2012. The GeWiss corpus: comparing spoken academic German, English and Polish. In T Schmidt & K Wörner (Eds), *Multilingual corpora and multilingual corpus analysis*. Amsterdam: John Benjamins. 319-337.
- FLOWERDEW, L. 2009. Applying corpus linguistics to pedagogy. *International Journal of Corpus Linguistics*, 14(3):393-417.
- FLOWERDEW, L. 2012. *Corpora and language education*. London: Palgrave Macmillan.
- GABRIELATOS, C. 2005. Corpora and language teaching: just a fling, or wedding bells? *TESL-EJ*, 8(4):1-35.
- GOOSSENS, D & S GRANGER. 2017. *Learner corpora around the world*. Available from <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html> [Accessed: 02 August 2018]
- GRANGER, S. 1998. The computer learner corpus: a versatile new source of data for SLA research. In S Granger (Ed). *Learner English on computer*. London: Addison Wesley. 3-18.
- GRANGER, S. 2002. A birds-eye view of learner corpus research. In S Granger, J Hung & S Petch-Tyson, (Eds), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins. 3-33.

- GRANGER, S. 2003. The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 37(3):538-546.
- GRANGER, S. 2004. Computer learner corpus research: current status and future prospects. *Language and Computers*, 52(1):124-145.
- HOMSTAD, T & H THORSON. 1996. *Using writing to-learn activities in the foreign language classroom*. Minneapolis, MN: University of Minnesota.
- JAWORSKA, S, C KRUMMES & A ENSSLIN (2015) Formulaic sequences in native and non-native argumentative writing in German. *International Journal of Corpus Linguistics*, 20(4):500-524.
- JOHNS, T & P KING. 1991. *Classroom concordancing*. Birmingham: University of Birmingham.
- KRUMMES, C & A ENSSLIN. 2014. What's Hard in German? (WHiG): a British learner corpus of German. *Corpora*, 9(2):191-205.
- KRUMMES, C & A ENSSLIN. 2015. Formulaic language and collocations in German essays: from corpus-driven data to corpus-based learning materials. *Language Learning*, 43(1):110-127.
- LÜDELING, A. et al. 2008. Das Lernerkorpus Falko. [The Falko learners' corpus.] *Deutsch als Fremdsprache*, 2:67-73.
- MASGORET, AM & RC GARDNER. 2003. Attitudes, motivation and second language learning: a meta-analysis of studies conducted by Gardner and associates. *Language Learning*, 53(1):123-163.
- MCCARTHY, M & A O'KEEFE. 2010. *The Routledge handbook of corpus linguistics*. London: Routledge.
- MCENERY, T & A HARDIE. 2012(a). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- MCENERY, T & A HARDIE. 2012(b). *Statistics in corpus linguistics*. Available from <http://corpora.lancs.ac.uk/clmtp/2-stat.php> [Accessed: 15 March 2015].
- MCENERY, T & R XIAO. 2010. What corpora can offer in language teaching and learning. In E Hinkel (Ed), *Handbook of research in second language teaching and learning* (volume 2). London: Routledge. 364-380.
- NESSSELHAUF, N. 2004. Learner corpora and their potential for language teaching. In JMcH Sinclair (Ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins. 125-156
- OGHIGIAN, K & K CHUJO, 2010. An effective way to use corpus exercises to learn grammar basics in English. *Language Education in Asia*, 1(1):200-214.
- O'KEEFE, A, M MCCARTHY & R CARTER. 2007. *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- ORTNER, GJ. 2013. A pilot study for the proposed masters dissertation, 2014. *Examining the formal writing skills of German Studies 3 students, using corpus linguistics, to uncover common lexical and morphosyntactic errors in student writing*. Unpublished honours thesis, Grahamstown, Rhodes University.
- PAWLEY, A & FH SYDER. 1983. Two puzzles for linguistic theory: native-like selection and native-like fluency. In JC Richards & RW Schmidt (Eds), *Language and communication*. London: Longman. 191-226.
- RAGAN, PH. 1996. Classroom use of a systemic functional small learner corpus. In M Ghadessy, A Henry & RL Roseberry (Eds), *Small corpus studies and ELT*. Amsterdam: John Benjamins. 207-236.

- SINCLAIR, JM. 1987. Collocation: a progress report. In: R Steele & T Threadgold (Eds), *Language topics: essays in honour of Michael Halliday*. Amsterdam: John Benjamins. 319-331.
- SKRANDIES, P. 2011. Everyday academic language in German historiography. *German as a Foreign Language*, 1:99-123.
- STEINHOF, T. 2007. *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. [Scientific text competence: oral language use and writing development in the scientific texts of students and experts.] Tübingen: Niemeyer.
- VAN ROOY, B & L SCHÄFER. 2002. The effect of learner errors on POS tag errors during Automatic POS tagging. *South African Linguistics and Applied Language Studies*, 20:325-335.
- WRAY, A 2000. Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* 21(4):463-489
- WRAY, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- YOON, H. 2008. More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2):31-48.

BIOGRAPHICAL NOTES

Gwyndolen Ortner is a PhD student in German Studies at Rhodes University. Her research concerns itself with teaching and learning of German as a Foreign Language and blended learning.

Undine Weber is Head of German Studies at Rhodes University. Her research interests are the scholarship of teaching and learning, German literature of the 20th century, cultural studies and translation.