Corpus-Based Teaching of German Compound Nouns and Lexical Bundles for Improving Academic Writing Skills

Enseignement sur corpus des noms composés et des ensembles lexicaux pour améliorer les compétences en rédaction académique en allemand

Marina Kogan, Anna Yaroshevich and Olga Ni
Corpus-Based Teaching of German Compound Nouns and Lexical Bundles for Improving Academic Writing Skills

Enseignement sur corpus des noms composés et des ensembles lexicaux pour améliorer les compétences en rédaction académique en allemand

Marina Kogan, Anna Yaroshevich and Olga Ni

Acknowledgements

We would like to acknowledge the DAAD lecturer Rayk Olhöft for his invaluable help at the preliminary stage of the experiment, our German colleagues from LUH Dr Sigrun Schroth-Wiechert and Anna Tilmans, and Dr Vitor Zakharov from St Petersburg State University for their support during this research and discussion of its results.

The authors also express their gratitude to the guest editors for comments on an earlier version of this paper.

1. Introduction

Foreign language teaching in Master’s degree engineering programs at Peter the Great St Petersburg Polytechnic University (SPbPU) reflects the main trends in language learning for professional purposes in universities of Russia. In Bachelor’s degree engineering programs, students learn foreign languages only over two years, which does not provide sufficient opportunities for language learning. Master’s degree engineering programs require a foreign language course of two academic hours a week within one semester. Within this four-month period, students are expected to master both general language and language in their professional domains. With so few teaching hours, it is essential that both class work and independent learning be optimized. Consequently, the educational process should be structured so that students learn to use components of
English is the most common foreign language chosen by students at SPbPU and selecting German as the second foreign language, as well as English as the first foreign language, is very rare (Ni, 2011). German classes in SPbPU’s Master’s degree programs include students from different majors, which greatly complicates language teaching for professional purposes. Such a situation is caused by the fact that the number of students who study German at school and/or during the Bachelor’s degree program declines annually. This is in line with global trends. In 2017-2018 the number of SPbPU Master’s degree students who were studying German accounted for approximately 1% of students who learn English. Their German language course focuses on developing the skills necessary to understand professionally-oriented texts, professional verbal communication and academic writing. The teaching material is comprised of study packs, course books (Basova, 2013; Albul, Vasil’eva & Stratonova, 2008) and content of the Deutsche Welle website.²

Deutsche Welle (DW) is Germany’s public international broadcaster, which includes radio service in 30 languages, satellite television service and a news site in seven languages, including Russian. The DW news site contains regularly updated articles on a variety of topics. Articles about achievements in science and technology can be found in such sections as Deutschland, Wissen und Umwelt, Wirtschaft, and Projekt Zukunft global Ideas. Engineering students at SPbPU are mainly interested in the latest news in mechanical engineering, the motor industry, and nuclear engineering available on the DW website. These popular science materials contain technical details and professional vocabulary. The class work includes usage and revision of professional vocabulary and special collocations, information exchange on research issues, discussions and PowerPoint presentations. Mastering pre-fabricated structures is essential, as such structures are common in both spoken and written communication. Indeed, the major finding of corpus linguistic research over the past decades is that “language is highly patterned” (Römer, 2009, p. 141).

In our opinion, corpus linguistics methodology could be used to enhance the present curriculum, which is based on the communicative approach. To this end, the following section reviews the literature on DDL usage for writing skill development in the course of Language for Specific Purposes.

2. Literature Review

Recent decades have seen an increasing interest in using corpora in language learning and teaching. This approach is referred to as “data-driven learning” (DDL), the term first introduced by Tim Johns³ in 1990. Language teachers and learners today can access many free corpora on line, including very large general corpora, genre- and domain-specific ones, as well as parallel corpora, comparable corpora, and learner corpora. Numerous publications on DDL and the use of corpus related resources, including the recent surge of special journal issues on the topic (Godwin-Jones, 2017) and the first meta-analysis in this domain conducted by Boulton and Cobb (2017), could be taken as evidence that the predicted revolutionary change in teaching methodology and overwhelming usage of corpora in language teaching has finally started. However, our experience leads us to believe that Boulton’s conclusion is still true, especially for languages other than English:
“Despite the considerable research interest and multiplicity of resources available, public awareness is low: corpus consultation remains rare even in university and research environments and it has had virtually no impact on ‘ordinary’ learning practices elsewhere.” (Boulton, 2010a, pp. 18–9) Furthermore, researchers point out the lack of available research on using specialized corpora other than English (Boulton, 2010b; Yoon, 2016).

In Boulton’s comprehensive survey (2010c) of 93 papers, only 11 of these deal with a language other than English, with 6 devoted to the German language. None looked at the Russian language. Boulton and Cobb (2017) conducted a meta-analysis of data-driven-learning studies that was limited by the necessity of containing pertinent secondary data, English was the main target language of most of the 64 studies and only two studies concerned German (Godwin-Jones, 2017).

Boulton and Cobb’s meta-analysis (2017), as well as their other comprehensive paper (Cobb & Boulton, 2015), showed that written language was clearly the dominant focus of corpus research. This might be due to the fact that the development of modern computer technologies and electronic communication has caused written language proficiency skills to be seen as a central aspect of specialists’ professional expertise. Meta-analysis also reveals that accessible online corpora, such as the COCA and the BNC, are the most frequently used objects of study (Godwin-Jones, 2017, p. 19).

In terms of the effectiveness of DDL approaches for developing writing skills and error correction, research data are somewhat mixed. Students’ enthusiasm for corpus work often depends both on their level of English and the extent of training and support available. For example, in Gaskell and Cobb’s study (2004) pre-intermediate learners of English integrated specific language points well, but error types did not significantly improve as a result of instruction that used concordancing. Improvement was found in only three error types of the ten most frequent errors they had categorized. Gaskell and Cobb developed a special tool available on their web site Compleat Lexical Tutor, which they called Corpus corrector. This tool is meant to help users correct the typical mistakes in writing. In other experimental settings, native English graduate students majoring in French studies (O’Sullivan & Chambers, 2006) and Chinese students from a balanced mix of humanities, medical and science backgrounds (Crosthwaite, 2017) successfully corrected mistakes in their written works (in French and English correspondingly) through corpus consultation. However, they had higher levels of the foreign language competence and were familiar with the DDL fundamentals.

Learners’ interest in specialized corpora is largely dependent on whether relevant the corpus is perceived by learners’ as being relevant to their needs. Using a corpus as a reference source for academic English writing may be ineffective and demotivating if it does not contain examples of language use in students’ specific technological/scientific areas (Chang, 2014; Charles, 2014). An example of good practice which takes this factor into account is Chang (2014), where Korean IT and engineering students were encouraged to compile their own corpus, named Michelangelo, through student selection of papers and articles from journals in their fields. Chang reported that students appreciated the access to this additional local corpus as a complement to the Corpus of Contemporary American English (COCA) which they used, because the COCA does not have a sufficient number of technical articles. On the other hand, they complained about the lack of necessary examples in the specialized corpus of a small size, which later stimulated them to expand the corpus and refer to its updated version.
In a similar vein, researchers at SPbPU have developed a number of activities exploiting a DDL approach, based on learners’ personal corpora of research articles relevant to their scientific interests. The activities depend on the functionalities of the concordancer program available. For example, karTatekA developed at General Linguistics Department of Philological Faculty of St Petersburg State University, allows the user to access different lists (frequency list, inverted list, word-length list), to create lexical and grammatical homonyms, to segment words, to search by word element and morpheme, and finally the possibility to unite all word family members onto a single card and then to gain access to all contexts from that card (Almazova & Kogan, 2013). The main obstacle hindering karTatekA’s wider use is that it involves a time-consuming procedure of preparing the original text, usually in pdf format, into the required txt (plain text) format in order to build concordance with a sentence-length minimal context.

That research targeted work of post-graduate students who had used a Text-based concordance tool from the Compleat Lexical Tutor website to build concordances of their corpora and to identify unknown words from the wordlists they generated. Then, they had to try to memorize them using the variety of website tools, for example Multi-concordance, Text-based range, List_learn, and VocabProfile. Words from each participant’s “100-unknown-word list”, based on their own corpora, were then randomly selected and used to test vocabulary acquisition. The results show that the activities helped the post-graduate students of SPbPU to expand and consolidate field-specific vocabulary recognition (cf. Almazova & Kogan, 2014).

3. Availability of German Specialized Corpora

To our knowledge, there are no appropriate corpora resources, which would allow engineering students to benefit from using a DDL approach in developing writing skills while writing their Master’s degree or PhD theses in German. In her research on learning verb-preposition collocations by both advanced learners of German and beginners, Vyatkina used the Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart (DWDS⁵) corpus: a large, freely and publicly available corpus of contemporary German (Vyatkina, 2016a, 2016b). Jaworska refers to the international project Gesprochene Wissenschaftssprache Kontrastiv (GeWiss⁶) “Spoken Academic Language in Contrast”, which was developed for contrastive studies of spoken academic discourse across three languages: English, German and Polish. The GeWiss corpus complements resources such as the Michigan Corpus of Spoken Academic English (MICASE⁷) and the British Academic Spoken English Corpus (BASE⁸) (Jaworska, 2015, p. 185).

Schroth Wiechert from Leibniz University of Hannover (LUH⁹) decided to bridge this gap through the development of a specialized corpus as a resource for learners of academic and technical German. She concluded that the number of such corpora for learners of academic German is very small, are not easily accessible and the examples they contain are of low relevance to Master’s degree students of a specific engineering field (e.g. Turbomachinery and Fluid Dynamics or Civil Engineering). She began to concretize her idea as a part of the Strategic Partnership Program between Peter the Great St Petersburg Polytechnic University and the Leibniz University of Hannover.

Since 2014, LUH and SPbPU have been collaborating on the development of the Deutsch, English and Russkii (DEaR) Corpus of German, English and Russian languages for
engineering. The target audience is students and lecturers. The DEaR corpus will be an annotated on-line corpus with different search capabilities composed primarily from electronically published PhD and Master's degree dissertations written by engineering Master’s and Post-graduate students, native speakers of either German, English or Russian. The acquisition, preparation and annotation of technical texts for the DEaR Corpus is under way (Gärtner, Schroth-Wiechert & Kogan, 2015; Kogan, Gärtner & Schroth-Wiechert, 2016). When the development has been finished and the legal copyright aspects adjusted, the DEaR corpus is planned to be available online. Currently, the German part has been developed and named Kod.ING (Korpus der Ingenieurwissenschaften).

As part of the Kod.ING corpus, we have focused on the analysis of over 200 PhD theses in the fields of Electrical Engineering (EL), Civil Engineering (BG) and Mechanical Engineering (M) with total of seven million tokens in order to determine its relevance for developing academic writing skills in Russian university engineering students. Our analysis conducted with the freeware-license AntConc software\(^\text{10}\) (Anthony, 2006) pursued the goal of determining if the Kod.ING corpus is relevant for Russian students who learn German in order to correct their typical errors. This goal was achieved during the preliminary stage of the teaching experiment.

4. Teaching experiment: preliminary stage

4.1. Selecting language items

Following Boulton (2010b), Gaskell and Cobb (2004), we expected to detect language problems from learners’ own writings. This was done at the preliminary stage of the experiment, at the end of the spring term of the 2016–2017 academic year. We invited volunteers from the Tandem project (Stratonova, 2016) to write a short essay of up to 200 words answering the question: “If I had a chance to study at a German University, which subjects would I select and why?” Five students responded. Their essays were carefully checked and discussed with a native German speaker. Despite many mistakes with prepositions, articles, word choice, grammar mistakes related to verbs and link words, word order in simple and complex sentences, patterns could not be identified. This may be due to the small size of the student essay corpus.

Therefore, two problematic areas, compound nouns and lexical bundles, were chosen for the teaching experiment. Russian learners of German used the genitive phrases common for Russian scientific discourse instead of compound nouns common for German scientific discourse, possibly, as a result of interlingual interference. They also misused, underused and omitted academic lexical bundles.

4.1.1. German compound nouns

German compound nouns can consist of two or more morphemes. These nouns can be described as a system comprised of an attributing part and a main (attributed) part, where the first morpheme describes the subsequent one. German compound words can be formed from any part of speech and follow one of two patterns: I) a one-word compound noun and II) two-/multi-word collocations, as in the examples given below.
For category II, it is important to stress that a sequence of words constitutes a single unit at the semantic level and is often considered fixed according to dictionary entries. Research has shown that compound words are essential to scientific discourse (cf. Ickler, 1997), as numerous terms are compound nouns comprised of generally used words, for example der Weißfisch (white fish as a species). Moreover, generally used compound words can acquire a new connotation or a new meaning in certain types of discourse. Some compound nouns can be translated into English or French quite accurately, which is explained by similar syntactic structures of the expressions (e.g. absolute constructions in German, no declension endings) or hyphenated spelling. This facilitates visual reception of morphemes.

Translating such nouns into Russian can be difficult because of the incompatibility of declension forms in German and Russian. Moreover, the Russian language lacks constructions similar to absolute English constructions. The problem when translating German compound nouns is determined by a wide range of options and combinations of original morphemes, which is impossible in the Russian language. An excellent example is the pair Radiowecker (radio alarm) and Weckerradio (radio which has different functions and among the main ones is the alarm function). However, the relationship between components of compound nouns and their sequence in similar Russian constructions are different. For instance, das Koordinaten system is translated into Russian as a phrase where the genitive case is used: system of coordinates. These discrepancies can lead to students’ difficulties in using and translating the vocabulary of scientific discourse, perhaps due to the phenomenon of interlanguage interference which is widely studied (see, e.g., Kostina, Hackett-Jones & Bagramova, 2017) and also mentioned in studies in different domains of language teaching (Almazova, Kostina & Khalyapina, 2016; Almazova, Rogovaya & Gavrilova, 2018).

4.1.2. Lexical bundles

Formulaic language has been intensively studied in Second Language Acquisition research as can be seen from Wood’s monograph (2015). While referring to papers focusing on academic bundles in Spanish and Korean, Hyland affirms that the vast majority of research looks at academic bundles in English (Hyland, 2012, pp. 150–1). The analysis he conducted shows not only that bundles are central to the creation of academic discourse, but that they occur and behave in dissimilar ways in different disciplinary environments (Hyland, 2008). For ESP/EAP course designers this means that they have to take into account their students’ specific target context, when they design their teaching. For example, Valipour and Nassaji (2013) produced the Chemistry Academic Word List (CAWL) based on the analysis of a corpus of 1,185 chemistry research articles. They
found that 27.85% of the frequent words in their CAWL corpus had not been listed in the widely used AWL compiled by Coxhead (2000). An Academic Formulas List (Simpson-Vlach & Ellis, 2010) and an Engineering Academic Formulas List (Fox & Tigchelaar, 2015) have become available recently. Though integrating corpus-based techniques into the teaching of academic German vocabulary is “still in its infancy” (Jaworska, 2015, p. 188), academic German teachers and researchers have also produced lists of academic lexis, including Schroth-Wiechert (2011) and Graefen (2009). Moreover, a comprehensive academic writing textbook exists which focuses on the most frequently used academic vocabulary and word collocations including idiomatic and metaphorical expressions (Graefen & Moll, 2011).

In our study, we decided to identify the “most frequent recurrent sequences of words” in the Kod.ING corpus, following Biber, Conrad and Cortes’s statement that frequency searches can help us select the “basic linguistic constructs with important functions for the construction of discourse” (2004, p. 398). All selected lexical bundles meet the criteria of size (Min – Max: 2 – 4), frequency (≥100), and range of presence in all three field subcorpora (BG, EL, M). Of the three categories of lexical bundles (referential expressions, discourse organizing expressions and stance expressions) proposed by Biber, Conrad and Cortes (2004), we found Discourse organizing expressions to be the most frequent in the Kod.ING corpus.

For the experiment, we also selected the most frequent recurrent compound nouns formed from the three most frequent bases: System, Technik, and Maschine. All of them meet the following criteria: Frequency ≥70; Range – in all field subcorpora (BG, EL, M). We also included one three-word collocation (kartesisches Koordinaten system), which did not meet the frequency criteria, but is often incorrectly used in students’ writings. We queried the German-Russian parallel subcorpus of the Russian National Corpus (RNC11) for the selected lexical bundles and compound nouns from the Kod.ING corpus (Appendix A), but none of the compound nouns and all but two lexical bundles were found. This reflects the lack of professional terminology, including technical compound nouns, in the General reference corpora, to which the Russian National Corpus belongs.

4.2. Hands-on and/or hands-off training?

Using corpora in a teaching intervention necessarily raises the question of how to train students to query corpora, using hands-on training (the student directly manipulating the corpus) and/or hands-off training (the teacher preparing the corpus data). As Vyatkina (2016b) mentions in her in-depth analysis of the available empirical research on hands-on and hands-off DDL, there is a lack of studies that compare the outcomes of hands-on and hands-off DDL interventions. She concludes that “hands-on and hands-off DDL were equally effective” (p. 170) and recommends trying both types of DDL instruction.

The restrictions imposed on the direct usage of the Kod.ING corpus outside LUH exclude the hands-on option. However, we argue that awareness not only of a new method, but also of a newly available resource is very important for students’ future, independent learning of foreign languages. For this reason, we decided to familiarise students with the simplest queries in the parallel subcorpus of the RNC (Russian National Corpus).

The RNC is the largest, national scale reference corpus in the Russian language. Available on-line, free of charge, it is a linguistic resource, which can be used without registration.
Its planned size is 200 million word tokens taken from spoken genres, fiction, and written media (including academic and non-academic texts) in Russian from the mid-18th century to the present. The Russian National Corpus currently uses four types of annotation: metatextual, morphological, accentual and semantic; the introduction of syntactic annotation is planned for the near future. The system of annotation is constantly being improved, which allows for quite complex syntactic and morphological queries.

The RNC has 11 subcorpora of different types including a set of bidirectional parallel text subcorpora. In the latter subcorpora, Russian is complemented by its translation into a different language, and vice versa. The units of the original and the translated texts (usually, a unit is a sentence) are matched through a “leveling” procedure. At the time of publishing, nine bidirectional parallel text corpora are available including English/Russian, German/Russian and French/Russian parallel corpora.

In response to their query, a user receives a list of results resembling the one in Figure 1, which is a query for unter Berücksichtigung. One challenge for users is to find the Russian equivalent of the German highlighted expression, as the leveled sentences do not contain any graphic prompts.

Figure 1. – The results of an unter Berücksichtigung query in the German/Russian parallel corpus of the RNC.

The predominant issue related to using a DDL approach in language instruction remains the amount of necessary preliminary training (see Boulton, 2009). Boulton provides very limited instruction about corpus use, from “a short theoretical background followed by demonstrations of particular functions” (Boulton, 2012, p. 35) to student practice “without training” focused on their ability to derive useful information from impromptu concordances (Boulton, 2009, p. 40). On the other hand, Boulton (2012) admits that,
“students clearly would have liked further preparation in corpus use [...], especially ‘demonstrations’” (p. 36). Though it is obvious that students would benefit from more extensive preliminary training to become confident with a new tool or resource, we could only insert brief interventions with a DDL approach into regular German language classes.

5. Design

5.1. Research questions

The following research questions were explored:

1. Can Russian university students of engineering improve their academic writing skills following corpus-based instruction and the use of corpus-based teaching materials?
2. Will they be able to retain the gains in a mid-term perspective?
3. Do very short DDL interventions within a regular context of teaching/learning German develop participants’ interest in DDL?

5.2. Participants and Instructional Context

The participants of the current study were 14 Russian students of German enrolled in a compulsory Master’s degree course of German for Specific Purposes (GSP) at SPbPU. However, the study reports only on the 11 participants (six males, five females) who attended all DDL sessions (including pre- and post- and delayed tests) and submitted their homework exercises. Participants were aged from 21 to 23, with a mean of 22 years. All the participants were a group of only native speakers of Russian, with two students having also studied English. Students had different engineering majors, such as power plant engineering, electrical engineering, mechanical engineering, metallurgy and material science, technosphere safety. Most students had studied German for seven years at school prior to entering university, then the first two years of a Bachelor’s degree programs with a two-year break until their first year in the Master’s degree program. A start-of-year test showed that their proficiency level was low, equivalent to A2 and A2+, with only one student approaching a B1 level according to the CEFR.

Ninety-minute GSP classes took place once a week. Classes followed the uniform syllabus, so it was only possible to do short DDL interventions at the beginning of the first four classes. The hands-off method of corpus-based instruction was adopted, with hands-off exercises based on teacher-prepared worksheets used during regular classes and as homework. Nine compound nouns and eight lexical bundles were selected for the study at the preliminary stage of the experiment. Hands-on activities were based on the searches of Russian equivalents of the selected lexical bundles in the bidirectional parallel German/Russian subcorpus in the RNC.

DDL sessions ended in the middle of the term. At the end of the term we interviewed the group’s teacher to understand her opinion about the impact of the DDL intervention sessions on the learning process and students’ ability to use the compound nouns and linking expressions beyond the experiment.
5.3. Methodology: data collection

The data collection timeline over five sessions is presented in Table 1.

Table 1. – Data collection timeline.

<table>
<thead>
<tr>
<th>Oct. 9 (S1)</th>
<th>Oct. 16 (S2)</th>
<th>Oct. 23 (S3)</th>
<th>Oct. 30 (S4)</th>
<th>Nov. 20 (S5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test,</td>
<td>Hands-off activities,</td>
<td>Hands-on activities,</td>
<td>Post-test,</td>
<td>Delayed</td>
</tr>
<tr>
<td>Instructions</td>
<td>Hands-on instructions</td>
<td>Hands-off activities</td>
<td>Questionnaire</td>
<td>post-test</td>
</tr>
</tbody>
</table>

During the first DDL session (S1), participants took a 5-minute pre-test and received a set of instructions on the set of hands-off activities. They also were instructed about search word functions of the parallel German/Russian subcorpus of the RNC (S2) and participated in hands-on / hands-off corpus skill-building activities (S2-S3). The entire set of activities was split into in-class practice and homework that included a number of hands-on corpus skill-training tasks, as well as the self-instruction writing practice. The training was followed one week later by a 5-minute post-test, as well as a questionnaire concerning the corpus-based experience (S4). The delayed post-test was conducted 3 weeks later (S5).

The pre-test aimed at focusing the students’ attention on the gaps in their knowledge of the key words and collocations. The pre-test, as well as both the immediate and delayed post-tests, contained a list of seven target items (compound words and lexical bundles) to be translated from Russian into German. The words and collocations were scrambled so that the tests did not replicate each other. For each correct answer, the student received one point, with a maximum of seven points per test.

All test sheets contained seven compound words and lexical bundles to be translated from Russian into German. A series of worksheets included exercises on translation and concordance lines analysis (Appendix B, Figures B1, B2, B3), matching, filling-in exercises (Appendix B, Figure B4), leading students to the final creation of their own sentences with the key lexical bundles and compound words (Appendix B, Figure B5), and the use of the target vocabulary in free writing. Figure B6 shows an example of hands-on corpus skill-training task. The completed worksheets were collected on a regular basis during DDL intervention sessions; at the beginning of every session, the feedback on the previous homework task was provided.

6. Results

Three types of results will be presented here: pre-test, post-test, and delayed post-test results analysis, students’ responses to the questionnaire on corpus work perception, and interview data.
Overall, the quality of the homework tasks completed after the pre-test was satisfactory. However, the exercises focusing on lexical bundles were completed more successfully than those focusing on the compounds. None of the participants finished all of the tasks correctly. The mean success rate of the home task activities was about 60%. The student reports on the target-item searches in the parallel German-Russian subcorpus of the RNC proved that they coped with the challenges of independent hands-on corpus work as all students completed the task correctly.

In contrast, the results of the post-tests were not as high as expected. The 11 participants of the post-test scored from one to four points each. Overall, there is a correlation between learners’ homework task success and test scores. Table 2 contains information about the target items included in each test and the mean success rate in each of the two post-tests, in other words, the correctly translated items per test normalized to the number of subjects.

Table 2. – Test target items and the mean success rate for both post-tests.

<table>
<thead>
<tr>
<th>Target compound nouns and lexical bundles tested</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Delayed post-test</th>
<th>Post-test progression</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>das Koordinatensystem</em></td>
<td>0 (0%)</td>
<td>9 (81.8%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>in erster Linie</em></td>
<td></td>
<td>9 (81.8%)</td>
<td>7 (63.3%)</td>
<td>-2 (-18.1%)</td>
</tr>
<tr>
<td><em>das Gleichungssystem</em></td>
<td>0 (0%)</td>
<td>7 (63.3%)</td>
<td>9 (81.8%)</td>
<td>2 (18.1%)</td>
</tr>
<tr>
<td><em>im Gegensatz zu</em></td>
<td></td>
<td></td>
<td>3 (27.3%)</td>
<td></td>
</tr>
<tr>
<td><em>in Hinblick auf</em></td>
<td></td>
<td>3 (27.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>in der vorliegenden Arbeit</em></td>
<td>0 (0%)</td>
<td>3 (27.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>die Umformtechnik</em></td>
<td></td>
<td>2 (18.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>es wird deutlich dass</em></td>
<td></td>
<td>2 (18.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>die Werkzeugmaschine</em></td>
<td>0 (0%)</td>
<td>4 (36.2%)</td>
<td>4 (36.2%)</td>
<td></td>
</tr>
<tr>
<td><em>die Umformmaschine</em></td>
<td></td>
<td></td>
<td>1 (9%)</td>
<td></td>
</tr>
<tr>
<td><em>wird davon ausgegangen dass</em></td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>das kartesisches Koordinatensystem</em></td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>in Abhängigkeit von</em></td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>die Fertigungstechnik</em></td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The pre-test results show that none of the 11 participants of the experiment had previous knowledge of the targeted lexis. Intriguingly, the results for the post-test and delayed post-test show that success with the term *die Werkzeugmaschine* improved beyond the end of instruction. The results for the delayed post-test show that only three target items can be regarded as “learned firmly” by most participants: *das Koordinatensystem, in erster Linie, das Gleichungssystem*.

Table 3 summarizes the results concerning the mean number of words translated correctly and the portion of the correctly translated items to the total number of the items of both post-tests.

Table 3. – Results of the immediate and delayed post-tests.

<table>
<thead>
<tr>
<th></th>
<th>Mean number of correctly translated items</th>
<th>Percentage of correct word translation (7-item test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate post-test</td>
<td>2.36</td>
<td>33.8%</td>
</tr>
<tr>
<td>Delayed post-test</td>
<td>3</td>
<td>42.9%</td>
</tr>
</tbody>
</table>

The final homework task which students were asked to do between the immediate and delayed post-tests was to write sentences of their own using eight compound nouns and eight lexical bundles from the DDL interventions. We received only four completed papers. Therefore, just after the delayed test we repeated our call for making up “your own sentences”, stressing that they did not have to invent sentences using all the selected vocabulary items. They could do the task using just the most familiar and relevant terms. However, no one responded.

In total, students submitted 32 sentences with lexical bundles and 31 sentences with compound words, as one student completed only seven sentences with compound words. Table 4 presents the error types for this exercise.

Table 4. – Analysis of error types in sentence writing homework.

<table>
<thead>
<tr>
<th>Student errors</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical bundles</td>
<td></td>
</tr>
<tr>
<td>Grammar</td>
<td>10</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>6</td>
</tr>
<tr>
<td>Vocabulary, unrelated to the focus lexis</td>
<td>4</td>
</tr>
<tr>
<td>Compounds</td>
<td></td>
</tr>
<tr>
<td>Grammar</td>
<td>11</td>
</tr>
</tbody>
</table>
The results lead us to conclude that, while the subjects are good at receptive activities with the target vocabulary, they are much less confident in their writing.

6.2. Perception of corpus work

In order to understand students’ perception of working with the Kod.ING and the RNC corpora, they filled in a post-experiment receptivity questionnaire in Russian, partially drawing upon Boulton (2010b) and Vyatkina (2016b). In the seven closed questions, the students were asked to rate their satisfaction regarding RNC activities, as well as their satisfaction with the Kod.ING corpus activities and plans for continued use, by indicating agreement with statements on a five-point Likert scale scored from 1 (completely disagree) to 5 (completely agree). The “not sure” option was also available. Question 1 checked that the participants had used the RNC and question 7 asked them to confirm or disprove their willingness to study specific software enabling them to work with the Kod.ING corpus. Question 8 was open-ended and related to what the participants particularly liked or disliked in the RNC activities. Eleven students completed the questionnaire in the classroom immediately after the post-test. Due to the small sample, a purely quantitative analysis of the data is not statistically valid. However, the results are presented in Table 5 as they provide a general overview of the students’ perceptions.

Table 5. – Summary of students’ responses to the questionnaire (translated from Russian).

<table>
<thead>
<tr>
<th>Disagree (no) / agree (yes) questions</th>
<th>Disagree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Did you use the RNC to complete the homework?</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Q7. Are you ready to study specific software to work with the Kod.ING corpus?</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likert scale questions (1-completely disagree to 5-completely agree)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2. It was easy for me to work with the RNC.</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Q3. I liked to work with the RNC.</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4. I found it useful to work with the RNC.</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5. I faced a number of difficulties while working with the RNC.</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Q6. I plan to use the RNC to study German in future.</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On the whole, the students responded positively to working with the RNC. All the students participated in out-of-class RNC activities (Q1) and most of the students admitted they liked working with the corpus (Q3). Most of them indicated that they found
it easy to work with the RNC (Q2). Still, more than half the students pointed out that they faced a number of problems during their work with it (Q5). Most students agreed that the work done was useful (Q4) but in Q6 the results indicate that most were not sure if they would use the new skills in their future language learning. Nevertheless, more than half the students were willing to receive further instruction to obtain more corpus work skills (Q7).

The open-ended question 8 was only answered by five participants, one of whom expressed uncertainty about the subject mentioned. The other four students highlighted the usefulness of the RNC’s visual representation of the language data, such as highlighting of the search expression. The ability to see a full sentence translated was also appreciated. The participants also mentioned their curiosity about doing the homework with the help of a new tool and the ability to investigate use of the target German lexis in context. An unfavourable opinion was expressed towards the lack of highlighting in the Russian translation of the sought-for expressions. Participants also noticed that corpus work was quite time-consuming and the amount of instruction was insufficient.

6.3. Interview with the teacher

The teacher was interviewed at the end of the autumn term after the participants of the experiment had passed their final exam in German for specific purposes. Only one question was asked: “Was the impact of DDL interventions noticeable after the experiment ended and if so, could you discuss the evidence for this?” She replied affirmatively, adding that the DDL interventions helped students to grasp how compound nouns work in German. The theme on compound nouns in their regular syllabus was used as a follow-up to consolidate what students had already learned in the DDL interventions. As a result, the teacher felt she had saved time because she did not have to introduce the topic. She noticed that students were more confident doing tasks on compound nouns from their regular textbook, even though the compounds in the textbook were different from those in the experiment (e.g. Funktionstuechtigkeit, Taktstrasse). Also, she observed that students mostly used compounds correctly in a compulsory task during the term (writing a summary of a text relevant to their major), and also included them appropriately in their final oral exam.

7. Discussion and conclusions

This study demonstrates that, for a short-term teaching impact on a group of students learning German for specific purposes, the chosen teaching approach was effective. Despite lacking any knowledge of the focus lexis at the pre-test stage, all the learners improved their ability to remember and use the compound words and phrases. We would argue that the progress in the target vocabulary acquisition was made due to the willingness of learners to do the class exercises and homework.

The unexpectedly low post-test results might be explained by the participant’s low-level of German language mastery or by the insufficient length of interventions. We think that longer and more regular interventions incorporated into the non-DDL syllabus class will lead to better mastery of compound nouns and lexical bundles.
The interview with the teacher shed further light on our delayed post-test results. The improvement between two post-tests in our experiment (Table 3) contrasts with Vyatkina’s findings, as she found that “all outcomes increased on the immediate post-test and decreased on the delayed post-test, although not to the level of the pre-test” (2016b, p. 166). The improvement in our study could be attributed to the follow-up activities on the compound nouns that the teacher did within the regular syllabus.

One major discrepancy requires comment, that between a rather large number of mistakes in target vocabulary in sentence-writing task (Table 4) and the teacher’s comment that students used compound nouns quite confidently and correctly in the following term’s writing and oral tasks. This might be due to the different nature and focus of the writing exercises. It is probably easier and more natural for students to use compounds from the text relevant to their scientific interests while retelling or summarizing than to compose their own sentences.

It is also worth mentioning learners’ interest in doing hands-on tasks with RNC and their intention to apply the acquired knowledge. The learners’ interest in hands-on DDL activities leads us to conclude that direct DDL tasks should be included into even short paper-based DDL interventions, even in non-DDL syllabi. There is a risk, however, that paper based activities, e.g. based on the analysis of concordancing lines, though new and unusual for learners, may not provide as much of a rewarding feeling of discovery for learners, who are challenged to understand “how it works” when doing hands-on corpus work. For many engineering students, it is important to start by attempting to understand the technical aspects as they acquire the related language skills. This is especially important in a situation like ours where direct access to the target corpus (Kod.ING) is not possible.

In line with other studies (Cobb & Boulton, 2015; Vyatkina, 2016b), our research confirms that exercises such as sentence-writing are more difficult for low-level learners than other exercises and short DDL intervention sessions may not be enough for some of them to modify their writing. The low return rate of sentence writing worksheets may also be influenced by the low number of sentences relevant to students’ specific majors in the teaching materials. This must be taken into account in further studies.

The main limitations of our study concern the relatively small number of participants and the short length of DDL interventions. It would be difficult to devote more teaching time to these as the standard, non-DDL syllabus is already very full.

To conclude, while further research could examine the limits of a DDL approach, it could also help to diversify teaching materials by including more examples from specialized corpora that are relevant to students’ professional interests.

BIBLIOGRAPHY


3). Giessen, Germany. Available at <http://www.uni-giessen.de/faculties/f05/engl/ling/talc/home/programme/abstracts> (consulted 1 December 2017).


APPENDIXES

APPENDIX A

List of target compounds with English translations:

<table>
<thead>
<tr>
<th>German Expression</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>das Gleichungssystem</td>
<td>(system of equations)</td>
</tr>
<tr>
<td>die Konstruktionstechnik</td>
<td>design</td>
</tr>
<tr>
<td>das Koordinatensystem</td>
<td>coordinate system</td>
</tr>
<tr>
<td>die Fertigungstechnik</td>
<td>manufacturing system (machinery)</td>
</tr>
<tr>
<td>die Umformtechnik</td>
<td>forming operation process</td>
</tr>
<tr>
<td>das kartesisches Koordinatensystem</td>
<td>Cartesian coordinate system</td>
</tr>
<tr>
<td>die Umformmaschine</td>
<td>automatic forging machine</td>
</tr>
<tr>
<td>das Werkzeugsystem</td>
<td>tooling system</td>
</tr>
<tr>
<td>die Werkzeugmaschine</td>
<td>metal-working machine</td>
</tr>
</tbody>
</table>

List of target lexical bundles with English translations:

<table>
<thead>
<tr>
<th>German Expression</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>in Abhängigkeit von</td>
<td>depending (on); subjected (to)</td>
</tr>
<tr>
<td>es wird deutlich dass</td>
<td>it is clear that</td>
</tr>
<tr>
<td>unter Berücksichtigung + gen</td>
<td>in view (of)</td>
</tr>
<tr>
<td>wird davon ausgegangen dass</td>
<td>it follows / results from this that</td>
</tr>
<tr>
<td>im Gegensatz zu</td>
<td>contrary (to); as opposed (to)</td>
</tr>
<tr>
<td>in Hinblick auf</td>
<td>in view (of); taking into consideration / account</td>
</tr>
<tr>
<td>in der vorliegenden Arbeit</td>
<td>in the present paper / work</td>
</tr>
<tr>
<td>in erster Linie</td>
<td>in the first place</td>
</tr>
</tbody>
</table>

APPENDIX B

Worksheets for home task and classroom work (all directions are translated from Russian)
Figure B1.

Translate the sentences from German into Russian. Make them true for you.

5.1 Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Fertigungstechnik und Werkzeugmaschinen (IFW) der Leibniz Universität Hannover.

Die vorliegende Arbeit entstand während meiner Tätigkeit am Institut für Maschinen elemente, Konstruktions technik und Tribologie (IMKT) der Universität Hannover.
Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Technische Verbrennung der Universität Hannover.
Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Mechanik der Universität Hannover.

5.2 Durch integrierte Sensorik sowie präzise Fertigungstechnik sind enge magnetische Nutzflächen und somit relative große Tragkräfte erreichbar.

5.3 Diese Fertigungstechnik ist das gängigste Verfahren der Massivumformung und vom Mengenaufkommen auch das absolut bedeutendste.

Figure B2.

Translate the sentences from German into Russian. Notice the place of the bold expression in a sentence.

5.1 Im Gegensatz zu Experimenten mit einzelnen Blasen stellt er in Anwesenheit mehrerer Blasen nur die erste Oszillation des Druckes fest, wobei spätere Schwingungen stark gedämpft sind.

5.2 Da die Kosten für Diamantformrollen, im Gegensatz zu Profillöfen, unabhängig von der Schleiffrequenzbreite sind, bietet die Formrolle bei breiten Schleiffrequenzbreiten aufgrund der geringeren Kapitalbindung für das Werkzeug deutliche Kostenvorteile.

5.3 Eine auf diese Weise definierte beliebige Erregung kann auch genutzt werden, um nicht rotierende beschleunigte Scheiben zu untersuchen, die zwar eine zyklische Struktur aufweisen, im Gegensatz zu den obigen Annahmen aber aufgrund einer belastigen Anregung mehrere anregbare Schwingungsformen verschiedener Knotendickenmesser aufweisen.


7.1 In Anbetracht der zu erwartenden Variationen von wenigen Millimetern soll eine Kalibrierung in der vorliegenden Arbeit unterbleiben.

7.2 In der vorliegenden Arbeit werden die bisherigen Erkenntnisse bezüglich der Kapazität und der Qualität des Verkehrsablaufs in planmäßigen Knotenpunkten überprüft und weiterentwickelt.

7.3 Aufbauend auf dem Konzept der schnell exploriierenden Suchräume wird in der vorliegenden Arbeit ein Behauptungsalgorithmus erstellt, der nichtholonom und dynamische Zwangsbedingungen mit Reaktivität vereint.

Figure B3.

*Translate from Russian into German*

1. система уравнений (system of equations)

2. система координат (coordinate system)

3. технологическое оборудование (machinery)

4. декартова система координат (Cartesian coordinate system)

5. в зависимости от depending (on)

6. отсюда следует, что it follows

7. в настоящей (длинной) работе (in the present paper)

Figure B4.

Make up compound noun out of the given nouns, then fill in the gaps, using the words you have formed.

<table>
<thead>
<tr>
<th>das Werkzeug</th>
<th>+</th>
<th>das System</th>
</tr>
</thead>
<tbody>
<tr>
<td>die Koordinate</td>
<td></td>
<td>die Technik</td>
</tr>
<tr>
<td>das Umformen</td>
<td></td>
<td>die Maschine</td>
</tr>
<tr>
<td>die Fertigung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>die Konstruktion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>die Gleichung</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Die Umformung der ersten Stufe erfolgt in einem für das Prutzionsfahren typischen ________, welches während der Umformung geschlossen ist.
3. Die verwendete ________ muss jedoch einen ausreichenden Presseneinbau- raum zur Verfügung stellen, der sowohl das komplexe Umformwerkzeug als auch weiter Umformstufen zur Vor- und Zwischenformung aufnehmen kann.
4. Dieses dreidimensionale ________ ist durch seine Aussicht unbestritten verbunden und vollzieht somit eine Drehung um seine Z-Achse ...

Lidil, 58 | 2018
Figure B5.

Translate the given collocations and make up your own sentences using them.

1. Gleichungssystem der Form
2. Werkzeugsystem zum Präzisionsschmieden
3. Fertigungstechnik und Werkzeugmaschinen
4. Koordinatensystem mit den
5. ein lineares Gleichungssystem
6. Umformtechnik und Umformmaschinen

Figure B6.

<table>
<thead>
<tr>
<th>Expression / Examples of usage</th>
<th>The number of examples</th>
<th>Variants of translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>in Abhängigkeit von</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im Gegensatz zu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unter Berücksichtigung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in erster Linie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in Hinblick auf</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTES

3. Tim Johns (1936–2009), professor at Birmingham University, UK, the author of an often-quoted comment that, “Research is too important to be left to the researchers”.
5. DWDS: <www.dwds.de>.
7. MICASE: <https://quod.lib.umich.edu/cgi/c/corpus/corpus?...cc=micase>.
ABSTRACTS

This study highlights the problem of the lack of German specialized corpora for German for specific purposes (GSP) courses for engineering students and describes a project aiming at the development of such a corpus, the Kod.ING corpus. The authors show the relevance of the Kod.ING corpus in meeting the needs of Master’s degree engineering students at St Petersburg Polytechnic University who are studying lower-level German. At the preliminary stage of the pedagogical experiment, nine compound nouns and eight lexical bundles were selected from the Kod.ING corpus. These were taught to students through hands-on and hands-off data-driven learning (DDL) activities. The immediate and delayed post-tests proved the effectiveness of short DDL interventions in terms of acquisition of target vocabulary. The follow-up survey revealed students’ particular interest in hands-on activities with the Russian National Corpus (RNC). In conclusion, further research and pedagogical applications are suggested.

INDEX

Mots-clés: apprentissage basé sur les données (ABD), corpus spécialisés, allemand à des fins spécifiques (AFS), noms composés allemands, blocs lexicaux, compétences en rédaction académique, corpus national de la langue russe (CNR)

Keywords: data-driven learning (DDL), specialized corpora, German for Specific purposes (GSP), German compound nouns, lexical bundles, academic writing skills, Russian National Corpus (RNC)
AUTHORS

MARINA KOGAN
Peter the Great St Petersburg Polytechnic University, Department of Linguistics and Cross-Cultural Communication, St Petersburg, Russia

ANNA YAROSHEVICH
Peter the Great St Petersburg Polytechnic University, Department of Linguistics and Cross-Cultural Communication, St Petersburg, Russia

OLGA NI
Peter the Great St Petersburg Polytechnic University, Department of Linguistics and Cross-Cultural Communication, St Petersburg, Russia