

## CORPUS-DRIVEN ESP PEDAGOGY: A PRELIMINARY CASE STUDY

Maria Rudneva

Foreign languages department, RUDN University, Russia  
E-mail: rudneva-ma@rudn.ru

**Abstract.** *The study explores application of a freely available EcoLexicon online corpus in specific language classes at Environmental studies faculty of RUDN University, Russia. Although corpus tools are traditionally used by linguists, I intend to demonstrate how they can be offered to non-linguist students to facilitate better understanding of inherent semantics, depart from generic surface understanding of professional texts and delve deep into structures and meanings, enhancing overall linguistic competence of learners and instructors alike. One of the main results of the study is that implementation of corpus-based tools clearly demonstrates unreliability of automated translated tools and fosters departure of novice professionals from assisted translation towards independent exploration of semantic nuances of target specialized languages. Despite numerous benefits, corpus-based tools can present certain challenges as well. These deficiencies, however, may be remedied by careful design of teaching materials and assignments.*

**Key words:** *data driven learning, specialized corpora, ESP*

### 1. INTRODUCTION

The paper investigates potentials and limitations of direct application of a specialized corpus in ESP classes. Since 1960's corpora have been used to inform pedagogic content of grammars, textbooks and dictionaries alike. Their value for pedagogic material design has been widely recognized (Römer, 2011). This approach is known as data driven learning (DDL) (Johns & King, 1991) and has been applied in a number of contexts for second and foreign language acquisition in a few educational institutions. Content-wise DDL can be divided into two major categories featuring indirect and direct application of corpus tools (Vyatkina & Boulton 2017). Indirect application involves utilizing corpus tools for material development by second language instructors, in this case learners are not involved in hands-on experiences with corpus query outputs. Direct application, however, involves corpus searches by learners and can target various aspects of lexicogrammatical contextualization of linguistic units. According to meta-analysis conducted by Boulton and Cobb (2017) DDL proved to be an efficient approach for increasing linguistic awareness of learners and instructors alike. Besides, numerous perception studies suggest that learners are on the overall of a positive attitude towards DDL. DDL activities allow departure from traditional deductive learning, promoting learner autonomy as well as developing problem-solving and analytical skills. Using a

---

Submitted August 30<sup>th</sup>, 2020, accepted for publication October 27<sup>th</sup>, 2020

specialized corpus as a reference tool allows novice professionals to achieve better understanding of L2 terminology and eventually overcome terminological asymmetries between source and target language. Contextualization of near synonyms fosters better understanding of semantic and functional nuances and provides important information which can not be elicited from bilingual and monolingual dictionaries. Delving deep in professional contexts, understanding inherent semantics also allows for patching lexical gaps which inevitably emerge when translating specialized texts. Co-creation of new professional terminology is one of the key aspects novice professionals face on a daily basis, especially in rapidly developing technological contexts. In this sense it seems very promising and important to provide them with lexicographic reference tools that would allow for departure from calques as well as for resolving the issue of false cognates.

Besides apparent benefits that DDL approach has to offer, there are a few limitations that hinder its wider dissemination. The limitations have been discussed in literature from the instructors' and learners' perspective. Past research (Römer, 2009) demonstrates language teachers' reluctance to implement corpus tools in the classroom for various reasons, such as lack of necessary computing skills, as well as unavailability of user-friendly and free resources. According to Callies (Callies et al. 2016) to implement DDL approach language instructors need at least three kinds of competencies. The broader one would be technical skills, high level of proficiency in corpus query tools. Secondly, the instructors need to be able to formulate queries and interpret results, thus corpus linguistics proficiency is necessary. Thirdly, pedagogical skills are necessary to incorporate data retrieved from a corpus in teaching materials. Apart from professional competences DDL activities require significant preparation time and yet available materials are few (Vyatkina & Boulton 2017).

On the learners' side there are several obstacles as well. First of all, all available corpora were designed by linguists for linguists, therefore user interface might seem complex and unfriendly. Apart from that, the output of concordances might be overwhelming for inexperienced users, it takes time and effort to induce linguistic information, which might be discouraging.

Despite abovementioned limitations there has been a shift from using corpora as a learning aid towards their use as a reference resource, especially in language for specific and academic purposes (Vyatkina & Boulton 2017). ESP textbooks are usually overseen by publishers due to limited target audience and questionable potential profitability (Bennett, 2010). Therefore, ESP instructors have to draw on their own resources to design pedagogic materials, thus specialized corpora are of invaluable assistance.

Having taken into consideration the abovementioned potentials and limitations, the author intends to explore the benefits of utilizing a freely available specialized corpus of environmental texts for ESP classes, as well as to what extent we can push non-linguist students in their quest for semantic and functional features of the target language, and what their perceptions of the hands-on DDL experience are.

## 2. METHODS AND OBJECTIVES

The research focuses on investigation of learners' perception of using EcoLexicon specialized corpus in ESP classes. The students were offered training, guided and then independent practice with a freely available corpus of environmental texts. At the end of the semester a de-identified questionnaire on the DDL experience was administered to the students.

## 2.1. Participants

A hands-on corpus experience was offered to three groups of students (N=39) aged 19-21 at RUDN university. All students are Environmental sciences majors and also minor in specialized translation. Their English proficiency level is B2-C1 CEFR, according to the results of Cambridge English exams, which all students take at the university exam center. All students had at least two years of prior training in ESP by traditional textbooks.

## 2.2. Instruments

The students were offered preliminary training on a freely available corpus of contemporary environmental texts EcoLexicon (León-Araúz et al 2018). EcoLexicon is the newest and most expensive terminological knowledge base on the environment, it is widely used by language and ecology experts. The corpus is available in the Sketch Engine query system, which has user-friendly interface and offers numerous modes of data representation.

## 2.3. Procedure

The students had no prior knowledge of corpus tools, or any formal training in lexicographic research, therefore the training consisted of several stages as presented in Figure 1.



Fig. 1 Corpus training

At each stage the feedback was collected and the course was aligned to achieve better understanding and higher degree of engagement and motivation. The following query functions were introduced:

- normalized frequencies as the most revealing means to ensure proper linguistic choices. The issues of L1 interference, calques, as well as false cognates were consistently addressed and supported by quantitative corpus data;
- wildcards and frequent chunks – invaluable source of authentic contextualization to develop written and oral fluency;
- word sketch difference function – a complex yet powerful tool to elicit semantic and functional differences between near synonyms.

## 3. DISCUSSION

The first objective of corpus training was to achieve proper understanding of query system functionality. The participants were offered an orientation session with brief introduction of corpus methodology followed by demonstration of key features of Sketch Engine tool. After the class the participants were asked to rate their perception of the experience, the results are presented in Figure 2. As indicated in Figure 2, 13% of the

participants (N=5) considered corpus tools impossible to comprehend, 18 % (N=7) found the query system extremely complex, 41 % (N=16) acknowledged it was quite complex and only 28% (N=11) admitted the corpus tool was understandable. Since 69% (N=27) rated their experience as fairly positive. It was decided to extend the training session with three weeks of guided practice to ensure better understanding of the corpus tool functionality. On week 4 the participants were requested to rate their experience again, and this time the rating yielded a more positive result. Only 5% of the students (N=2) considered the corpus tool impossible to comprehend, 13% (N=5) still found it quite complex. However, the majority of the participants found the challenge agreeable with 33% (N=13) rating the tool as quite complex and almost a half, 49% (N=19) confirming corpus tool functionality was understandable.

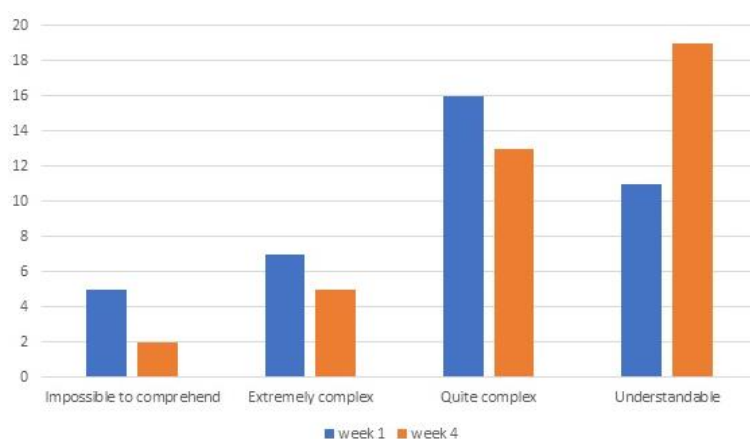


Fig. 2 Complexity

From week 5 onwards, the participants were offered weekly take-home corpus-based assignments which they were supposed to perform independently with a follow-up discussion in class. The students who still struggled or were reluctant to use corpus tools were still participating in discussions, using the data that their peers managed to induce; therefore, the activity was overall beneficial for the entire group.

The participants were offered to explore several basic functions of the query tool.

### *Normalized frequencies*

The frequencies function ensures proper linguistic choices, tackling false cognates and calques as well as L1 interference. Table 1 illustrates the case of assessing variant of translation:

Table 1 Normalized Frequencies

Экологические проблемы	Ecological problems	Environmental problems	Environmental issues
	1.08/mln	12.16/mln	9.75/mln

Due to L1 interference the students tend to translate collocation «экологические проблемы» as “ecological problems”. However, quick frequency check demonstrates, that this choice is the least authentic (1.08 occurrences per million words), whereas “environmental problems” and “environmental issues” would be more appropriate in the given context.

### *Wild cards and frequent chunks*

Wild cards and frequent chunks functions allow for the creation of authentic multi-word expressions, which is of critical importance for developing written and oral fluency. Informing academic writing with corpora data is not new, however, corpus-informed ESP writing seems under-researched. The students were assigned to revise their ESP writing tasks searching for multi-word expressions in the corpus in order to enrich their vocabulary, as well as increase authenticity. The excerpts below represent initial text produced by the student and its revised version.

#### *Initial text*

*According to our recent research of air pollution, the amount such of outdoor pollution as smog in Cherepovets region is alarming. There is a big necessity of technological modernization in coal-burning factories to prevent from stronger pollution. Also the indoor pollution is caused by bad ventilation. Different pollutants like formaldehyde and other chemicals are exposed inside cheap homes. Another cause of indoor pollution is tobacco smoke that appears in buildings where people use cigarettes. These two types of pollution, outdoor and indoor, affect people's health and ecological situation of the world.*

#### *Revised text*

*According to our recent research of anthropogenic air pollution, the amount of such types of **atmospheric** pollution as **industrial and photochemical** smog is alarming. There is **ultimate demand** for technological **advances** in coal-burning factories to prevent **severe** pollution. Indoor pollution, however, is caused by **poor** ventilation. The **negative effects of such range from exposure to various pollutants, such as formaldehyde, to tobacco smoke**. These two types of pollution, outdoor and indoor, affect people's health and **global** ecological balance.*

The examples above demonstrate that consulting a specialized corpus allows to significantly increase the level of sophistication of the text, as well as its authenticity, which, in its turn, dramatically improves overall quality of written production.

### *Word sketch difference*

Word sketch difference function is a powerful tool for disambiguation of key terminological concepts. Contextualization of lexical bundles allows for inducing linguistic information which cannot be found in traditional monolingual and bilingual dictionaries. This is specifically valuable for improving the quality of specialized translation, eliciting inherent semantic information as informing the content of ESP pedagogic materials.

To illustrate the features of word sketch difference function I would like to analyze semantic and functional differences of near-synonyms *pollution* and *contamination*, which are both offered by bilingual dictionaries as a translation of the key environmental concept *загрязнение*.

According to the results of the query *pollution* is used more frequently than *contamination*, with normalized frequencies of 198.84 per million vs 47.94 per million. Apart from that, there are a few functional differences.

Figure 3 indicates collocations of *pollution* and *contamination* as object.

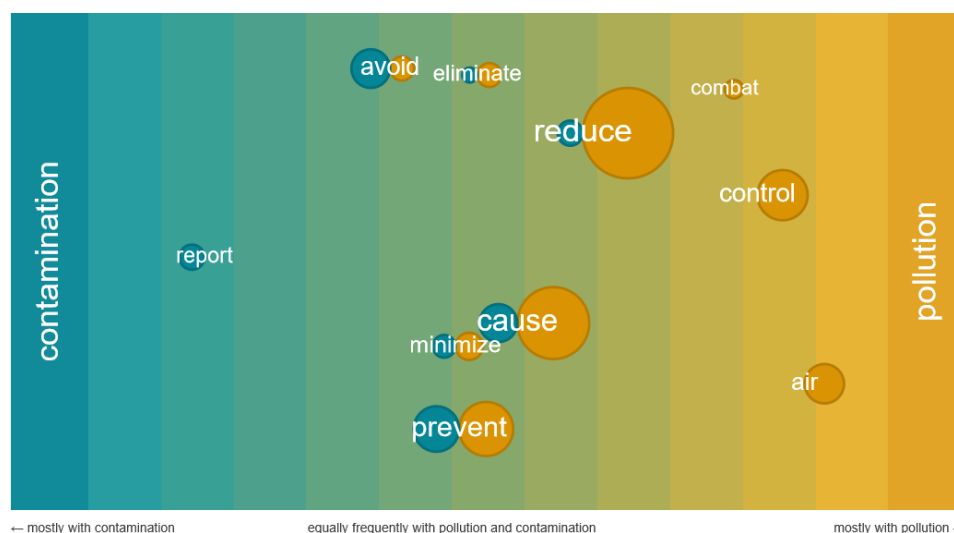


Fig.3 Verbs with pollution/contamination as object

As we can see, the verb *report* mostly collocates with *contamination*, whereas *combat* and *control* are mostly used with *pollution*. The verbs *cause* and *reduce* gravitate towards collocating with *pollution*, even though there are a few occurrences of *reduce contamination* and *cause contamination* in the corpus. Verbs *prevent*, *minimize*, *eliminate* demonstrate equal degree of co-occurrence with both terms, however, the verb *avoid* gravitates towards *contamination*.

Figure 4 represents modifiers of *pollution* and *contamination*.

The query indicated that modifier *air* is linked through a high frequency number of occurrences with *pollution*, but no statistically significant link has been established for *air contamination* collocation. The search returned high frequency collocations with the term *pollution* when speaking about big-scale location, i.e. *transboundary*, *air*, *indoor pollution*. There is collocation differentiation depending on the modifying agent, as demonstrated, i.e. we normally collocate *contamination* with *fecal*, *metal*, *radioactive* and *nitrate*, whereas *oil* usually co-occurs with *pollution*.

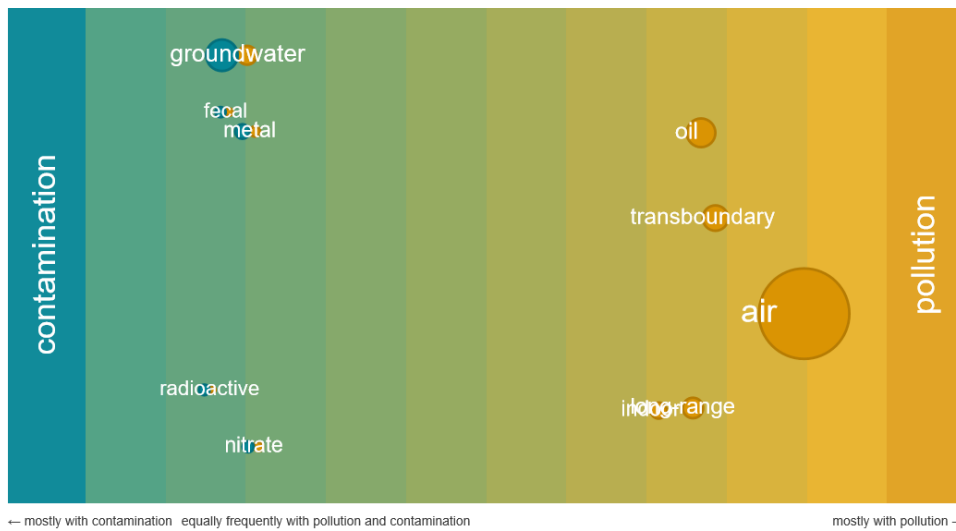


Fig. 4 Modifiers of pollution/contamination

This query output is a valuable source of information on the syntagmatic behavior of key terminological units.

#### 4. RESULTS

An important part of the experiment was the evaluation of learners' perception of corpus-based experience. The author conducted a de-identified open answer survey in which the participants were asked to highlight their overall attitude towards using the corpus tool as well as to comment on key challenges that they faced. Figure 5 demonstrates the results of the open answer survey.

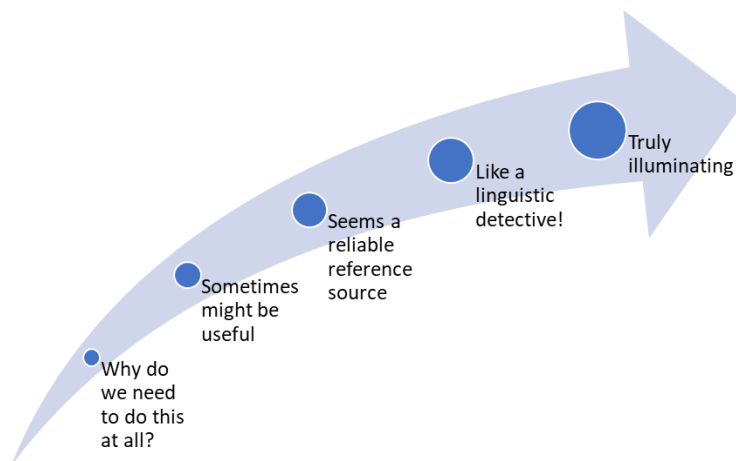


Fig. 5 Learners' perception

According to the results of the survey 85% of learners (N=33) considered corpus tools useful for terminology disambiguation, as well as the development of oral and written fluency. Some of them were more optimistic and compared the experience to a linguistic detective investigation, and also called it illuminating. Others responded in a moderate manner acknowledging overall benefits of corpus tools as reliable reference sources that can be useful sometimes. The remaining 15% of learners (N=6) were less enthusiastic about the corpus experience and questioned rationale behind data-driven activities.

The students were also asked to provide feedback on the most challenging aspects of data-driven learning. Their elaborations can be divided into three major categories. The first issue was overall complexity of authentic professional contexts which require high degree of L2 proficiency. This issue can be addressed by careful tailoring of offerings for lower level students. Secondly, as Figure 5 demonstrates, non-linguist students are generally less motivated by the notion of linguistic discoveries, therefore corpus tools should be introduced thoughtfully. Thirdly, as Figure 2 demonstrates, many students consider corpus tools complex, which can be partly remedied by resorting to basic query functions only as well as sufficient instructional intervention.

## 5. CONCLUSION

Overall, the learners demonstrated positive attitude towards hands-on corpus-based experience. Corpus tools have immense potential for providing precise, accurate and non-ambiguous data on specific terminology in professional contexts. Increasing availability of specialized corpora holds great promise of new advances for ESP learners, shifting the pedagogic focus from prescribed vocabulary lists to inductive learning and learner autonomy.

## REFERENCES

- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.
- Callies, M., Kreyer, R., Schaub, S., & Güldenring, B. (2016). Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. *Angewandte Linguistik in Schule und Hochschule*, 391-415.
- Johns, T., & King, P. (1991). Classroom Concordancing: English Language Research Journal, 4. *University of Birmingham: Centre for English Language Studies*, 1-31
- León-Araúz, P., Martín, A. S., & Reimerink, A. (2018). The EcoLexicon English corpus as an open corpus in Sketch Engine. *arXiv preprint arXiv:1807.05797*.
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? Aijmer, K.(Ed.) *Corpora and Language Teaching* (pp. 83-98.)
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art—and beyond. *Corpus technology and language pedagogy: New resources, new tools, new methods*, 5-24.)
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual review of applied linguistics*, 31, 205-225
- Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning and Technology*, 21(3), 1-8