Corpus-Based Teaching of Specific and Academic Vocabulary for ESL Students

Elena G. Ivanova
National Research Nuclear University MEPhI

Nina A. Skitina
State University of Education

Veronika P. Shabanova
State University of Education

Nataliia V. Poliakova
State University of Education

The current paper presents the results of applying the corpus-based approach to teaching academic vocabulary as well as the vocabulary of English for specific purposes for ESL students. The approach relies on the bedrock principle of data-driven learning – referring to the language directly, refraining from mediated resources (coursebooks, dictionaries, and teachers), thus increasing language awareness by building a glossary tailored to students' academic needs. The glossary encompasses not only lexico-grammatical meanings of the word but their functioning as well. The empirical data of the research proves that using corpora enhances conversational, reading, or writing skills.

Keywords: corpus, corpus-based teaching, teaching a foreign language, corpus-based exercises, ESP teaching

INTRODUCTION

The research focuses on employing corpora to enhance academic and ESP vocabulary. The use of corpora in foreign language teaching and learning has been an object of close attention in modern language education since 1980s. Several articles prove the urgency of research: since 2019 Google Scholar web search engine has provided 17,100 hits when searching for 'corpus-based teaching of vocabulary' in English but only 1,220 while searching for articles on the same field in Russian. In this study, the authors aimed to draw more attention to the potential capacity of corpus linguistics. According to the results of the literature review, various ways of using corpora have already been considered: mastering the use of collocations and prepositions (Bennett, 2010), tracing the use of discourse markers (Beeching, 2014) or numerous exercises to enhance language awareness (Shi, 2017). However, those methods are still not widely used in Russia.

Moreover, a relatively small number of articles describing corpus-based pedagogy for teaching and learning languages focus on ESP and academic vocabulary. The survey results among peer teachers, lecturers and senior students show that more than two-thirds of respondents are unaware of advances of corpus-based linguistics and corpora integration in ESL classrooms. Some scholars indicated the lack of pedagogical experiments and the necessity of building specific corpora (Gorina, 2014; Sysoyev, 2010; Youssef, 2020). The time has come for corpus-based teaching to become part and parcel of the curricula of trainee teachers and refresher courses for foreign language teachers, as the given method is problem-based and makes the process of knowledge acquisition active and more personified.

"Do-it-yourself" corpus, which is rarely, if ever, used while teaching English for specific purposes, has become an integral part of the educational process in experimental groups. With a wide range of exercises designed by an educator, it has proved that it is much easier for students to take and memorize the stances they use while expressing their positions or participating in academic discussions. Another significant aspect is the opportunity to provide students with up-to-date study materials, since the existing text-books do not introduce the quick changes taking place in the language, while with the "do-it-yourself" corpus an educator stays current and teaches topical issues. It should be noted that students obtain the skill of unlimited mastering their vocabulary. Thus, the concept of life-long learning is promoted.

The practical aspect of the study is highly significant: there are different lists of common English words, but scarcely any specialized vocabulary words. Using corpora makes updating the former and compiling the latter vocabulary easy. The advantages of devising a list of specialized vocabulary words with the help of corpora are thrown into sharp relief for an ESL teacher as it glosses over the need for being a well-rounded specialist.

LITERATURE REVIEW

Corpus Linguistics for English Teachers
The corpus-based approach to teaching English dates back to 1994 when the conference on Teaching and Language Corpora (TALC) took place at Lancaster University, followed by the publication of the proceedings in 1997 "Teaching and Language Corpora". Geoffrey Leech then noted the importance of language corpora in teaching, which was in line with Computer-Assisted Language Learning principles: automatic searching, sorting, and scoring; promoting a learner-centered approach; open-ended supply of language data; and enabling the learning process to be tailored (Friginal, 2018). Since then, Corpus linguistics (CL) has been continuously influencing the methodology of teaching foreign languages. However, the number of publications on the subject in question in Russia is scarce (Kogan et al., 2020).

CL yields more benefits while teaching grammar and vocabulary, reading and writing, not being strong in speaking and listening. Frequencies, contexts, and visualized data make the exposure to the target language more salient, thus helping learners to comprehend the semantics and syntactic aspects of lexical units. A corpus-based analysis of stance markers (hedges and boosters) in an academic context positively influences argumentative and explanatory writings of aspiring researchers. Corpus-based tasks enable students to comprehend the language forms better and get rid of misconceptions about the target language brought about by the interference of the native language. Nationally specific mistakes of Russian-speaking students are: to influence on sth/sb; explain me; depend from sth/sb; there are two variants, etc., to name just a few, and can be eradicated by using a concordance: the more often a linguistic unit is encountered by a student in different contexts, the better it is remembered. This approach to learning is psychologically sound. F. Barlett, back in 1932, introduced the concept of a scheme into cognitive psychology (Bartlett, 1995). The scheme is a memory component formed under the influence of the environment, contributing to the organization of incoming information. Consequently, new linguistic observations fall on the preexisting schemes, making such errors inevitable. However, corpus-based activities facilitate the defossilisation of such mistakes. Exposure to authentic materials in a foreign language lesson is of paramount importance. However, the main limitation in the use of authentic texts lies in the fact that learners are not always able to deduce the meaning of new words. Concordance in language learning can help the situation. Besides, with corpus-managers' help, the teacher can adapt a specialized text to a significantly smaller version of the initial text by removing difficulties in advance: cutting out low-frequency lexical units, or replacing them with frequent

synonyms. This issue can be easily managed with the help of the following educational corpus managers: Compleat Lexical Tutor (https://www.lextutor.ca) or Word and Phrase (https://www.wordandphrase.info). The founder of Compleat Lexical Tutor, Tom Cobb, puts forward a new reading model – resource assisted reading of extensive authentic documents, or R-READ for short. The essence of his approach lies in the idea that students will better understand authentic texts with the help of multifunctional linguistic consultants connected with linguistic corpora. The corpus-managers above allow users to study concordance, synonyms, word forms and frequency of any word in a text by clicking on it.

Our related challenge is determining how far Russian lecturers of foreign languages and their students are aware of the corpora and their tools. While university students of languages are familiar with corpora in conducting linguistic research, corpus resources are not sufficiently employed in language education. A recent update on the issue of corpus literacy conducted by Marcus Callies in 2019 showed that "there is a younger generation of language teachers who have been trained in the use of corpora for research purposes in their university studies, but this seems not to have made a significant impact on their teaching practice" (Callies, 2019, p. 252). The same trend is observed in Russia. To study the degree of corpus awareness, an online survey was conducted among the senior and graduate students of future foreign language teachers of the Faculty of Romance and Germanic Languages of Moscow Region State University. They were asked to fill in a Google form answering three closed-ended questions: "Have you heard of CL?", "Can CL tools be applied to teaching foreign languages?", "Have you used corpora for designing teaching materials or completing language assignments?" The results of the survey were estimated by the statistical analysis method: firstly, the total number of respondents was identified; then the total number within each data point was counted; and then the total number of observations divided by the number of observations within each data point. Thirty-two graduate teachers took part in the survey. The results show that more than two-thirds of respondents are unaware of advances in corpus-based linguistics and corpora integration in EFL classrooms.

CL – Applications to ESP and Academic Vocabulary

The corpus-based approach to teaching English for Special Purposes (ESP) can prove reliable and efficient as long as students are motivated and eager to learn the peculiarities of the specialized language. In 1986 Tim Johns, the initiator of Data-Driven Learning (DDL), admitted that this methodology was designed for a particular type of student: "adult: well-motivated: a sophisticated learner with particular needs (fairly closely specifiable in terms of target texts) in a particular learning/teaching situation (in which a great deal of emphasis is placed on developing students' learning strategies and on their responsibility for their learning)" (Johns, 1986, p. 161).

I.S.P. Nation defines specialized vocabulary as lexical units that are "recognizably specific to a particular topic, field or discipline" (Nation, 2001, p. 198). Language for special or specific purposes (LSP) is traditionally considered an integral part of the general language system. LSP can be viewed from linguistic and didactic perspectives. Linguists have studied its specific terminology, texts of different types, and registers. Nowadays, the focus is more on psychological and sociological dimensions, thus expanding the term to professional discourse or professional communication (Gunnarsson, 1997). LSP texts accumulate special (professional) knowledge. In the second half of the twentieth century, English took the lead in international communication, especially in science, technology, and trade.

Within ESP, several branches of study have emerged over the years, the best known and most frequently researched of these being English for Academic Purposes (EAP), English for Science and Technology (EST), and, to a lesser extent, English for Occupational Purposes (EOP) (Trace et al., 2015). The debate over the specifications of EAP is a long-standing one. Some scholars advocate the idea of the necessity of teaching students in a general way for their academic studies, while others argue that a subject-specific approach is more feasible, thus extending it to English for Specific Academic Purposes (ESAP). "ESAP emphasizes higher order skills, student development, and authentic text and features while working within specific epistemological traditions associated with different disciplines" (Enongene, 2013, p. 59). This aspect is a key factor in ESAP course design and content.

Mastering specialized vocabulary is important for students' overall academic literacy. One way to increase students' motivation is to expose them to the language relevant to their field. Still, there is not only a lack of teaching material tailored to a specialized field of students' expertise, but language teachers are not proficient in the area in which their students are majoring. That is why reading authentic materials with the help of corpus-managers and

corpora is essential for enhancing students' knowledge of general academic and specific vocabulary. Teachers' role in such circumstances is that of a facilitator: to provide students with academic professional materials and make them revise and review those materials to consolidate their knowledge. In this context, the corpus-based approach to ESP and EAP involves students compiling their corpora of specialized vocabulary and texts, creating a learner-centered atmosphere.

METHODOLOGY

The given study has followed the experimental method to prove the necessity of applying the corpus-based approach to developing academic competence of students on a par with enhancing their critical thinking abilities and autonomy. The study included a placement test, a pretest, and a posttest, which all the learners had to take.

The experiment was conducted during 2020-2021. The experimental base of the research was National Research Nuclear University MEPhI. Its participants were students in the master's program, which explains the choice of ESP and academic vocabulary program. There is little doubt that the participants must obtain more knowledge in the field of specialized and content-specific vocabulary: becoming part of academic discourse within the chosen field falls within the scope of the professional standard.

The study sample consists of 30 students. A group of 15 students was assigned as the control one where the traditional approach for teaching vocabulary was used. The other group of 15 students was assigned as the experimental one. They were learning through a corpus-based method. Both groups were expected to do reading and speaking assignments for 2 hours a week during their classes and had some writing and reading tasks. Both groups had to discover a certain amount of specific and academic words while reading the articles they chose.

The experiment started with a placement test all the learners had to take. It was a ready-to-use one that met our requirements to define the students' language level. The placement test consisted of 65 questions and checked grammar and general vocabulary. The results showed that the levels varied from Pre-Intermediate to Upper-Intermediate. (See the figures in "Results" section.) Nearly equal results are accounted for the standards all bachelor's degree students had to meet.

Furthermore, there was a pre-test to evaluate the knowledge of specialized and academic vocabulary. It is known that academic and discipline-specific vocabularies range from those particular to some specific field of knowledge to those that share their meanings and usage with words in other fields. Hence, the administered pre-tests were aimed at reading comprehension. The texts were authentic articles in which less-frequently used words were substituted with more frequently used ones. It was done with the help of the Compleat Lexical Tutor Corpus Manager. The test was based on a pass/no pass grading system. The mark "passed" was given in case of completing more than 30% of the tasks.

At the beginning of the course, the experimental group students were taught how to use corpora and corpus tools (managers). The most often used corpora were COCA (https://www.english-corpora.org/coca) and BNC (https://www.english-corpora.org/bnc/). Here are the main benefits of using corpora:

- We can find a word and see its context.
  For example, the word *"economy"* (some of the contextual usage from COCA):

...sustained 2.2 million jobs and stimulated the economy 3.5%...He oversaw the creation of more jobs in 2010...
...fast-tracked increased fuel economy standards for vehicles...
...Also, I don't believe those who receive disability monies is affected by the economy, so not relevant to trying to pin bad numbers on the president.

- We can see the most frequently used collocations (+different parts of speech) and, if necessary, read the original text, for example:

market economy, fuel economy, stimulate economy etc.

- There is an opportunity to see the word clusters (loose, medium or tight).

Example 1: *economy in, global economy etc.*

- If the meaning of the word is not clear, there is a video explanation.
- Concordance cases can be viewed in different ways according to the purpose.

Example 2: *This was coupled **with the changing international economies and a shift** in demand…*

*… Perhaps, **the economy and the financial rules** have changed…*
*Consider the absence of this **taxable income on the economy as** an example…*

- Searching for synonyms is also part of the corpus functions (also with more specific/general meaning), which is extremely practical.

Example 3: *economy – frugality – parsimony – efficiency …*

- It is possible to trace the usage of phraseological units in functional styles (spoken, fiction, academic, magazine, newspaper).

One of the corpus managers used in the experiment is the Compleat Lexical Tutor. In the lead-in session, the experimental group students were introduced to most functions of the aforementioned website: List Learn, Group Lex, and Concord Writer. List Learn allows learners to not only choose a correct definition of the word while reading the articles but also see how native speakers use this word in different discourses. Moreover, the 'save' function enables students to compile their vocabulary lists with definitions and concordances. The idea was to make the students' reading easier and more effective with the help of corpora. Of interest are some other functions suitable for teachers and learners, which this site comprises: vocab tests arranged according to the levels (based on the number of keywords a student knows), Keyword extractor etc.

Another useful corpus tool that is worth mentioning is AntConc. Before reading an article, every student had to upload the text into the corpus manager, extract the keywords, and find their definitions in the corpus. Later, they added these words/collocations into their glossaries. All the students were assigned to gloss a certain amount of academic and specific words/collocations from the articles they read. The groupmates also used the extracted keywords to predict the content of the articles and discuss them, using different stances. It is possible to exchange the keywords and use them in the discussions.

To implement a more individual route towards determining students' personal development strategies and bring cooperation into a classroom, a corpus of all the articles the students had read was created at the end of the term. It proved to be especially useful, as it enabled the students to see the chunks, not separate words, and come up with the vocabulary used while discoursing on the subject of their specialty.

It should be noted that creating specialized vocabulary word lists makes it possible for students and teachers alike to be equipped with the terminology or most frequently used words in any sphere. Of no less importance is the fact that such familiarizing does not take a lot of time.

A concordance section from our corpus was used in quizzes. The most frequent collocations were brought to the students' attention through the Group Lex interactive vocabulary-learning component of Compleat Lexical Tutor. Regarding the post-test, the following aspects were considered: the least frequent words were eliminated, and the exercises and a test to check the knowledge of ESP and academic vocabulary were created.

RESULTS

The results of the experiment proved that the students engaged in corpus-based learning retain the ESP vocabulary much better and their academic reading, writing and speaking skills improve significantly.

The round-off results are as follows. For the experiment group (with the corpus-based approach):

1. **Placement test**
   - Pre-Intermediate – 12% (2 students)
   - Intermediate – 58% (9 students)
   - Upper Intermediate – 30% (4 students)
2. **Pre-test (special vocabulary)**
   - Passed – 73% (11 students)
   - Failed – 27% (4 students)
3. **Posttest**
   - Pre-Intermediate – 7% (1 student)
   - Intermediate – 50% (8 students)
   - Upper Intermediate – 43% (6 students)

For the group with the traditional approach:

1. **Placement test**
   - Pre-Intermediate – 20% (3 students)
   - Intermediate – 60% (9 students)
   - Upper Intermediate – 20% (3 students)
2. **Pre-test (special vocabulary)**
   - Passed – 67% (10 students)
   - Failed – 33% (5 students)
3. **Posttest**
   - Pre-Intermediate – 13% (2 students)
   - Intermediate – 67% (10 students)
   - Upper Intermediate – 20% (3 students)

Thus, the difference in the results can be considered statistically significant. Depending on the language level, it varies from 5 to 13% in the experimental group and only up to 7% in the group with the traditional teaching approach. This proves that vocabulary is better learned and the language level can be enhanced through authentic contexts.

It should be noted that creating and using the "do-it-yourself" corpus allowed us to provide the students with up-to-date study material. It is a matter of great importance since the existing textbooks do not introduce the quick changes in the language. Word lists, collocations and concordance cases can be updated quickly with the help of the corpus created by the students to show the changes in the linguistic input. Thus, an educator has an opportunity to stay current and teach topical issues.

Not the least important factor stating the advantage of applying the corpus-based method is its ability to ease the teachers' work. Compiling and using the exercises, quizzes or tests is not difficult, while the results are promising. Moreover, corpus-based teaching suits the modern age and saves time, thus addressing the major challenges for any educator.

Another significant aspect is the shift of the paradigm of studying, which means promoting the concept of life-long learning. The abovementioned concept includes "creating the knowledge" via "mastering the knowledge" instead of merely "acquiring the knowledge," which meets educational standards of higher education and UNESCO guidelines for teachers (UNESCO, 2019). In this way, students obtain the skill of mastering their vocabulary.

CONCLUSION

The use of the corpus-based approach has proved to be of help in enhancing the students' language awareness and cultural education. The compiled lists of specialized vocabulary words can be easily used for further training. All the exercises used in the experiment may be implemented in further training courses to boost ESP/academic vocabulary and concordance skills. We also consider it reasonable to include the work with corpora into the curricula for master degree programs. The corpus-based teaching method should

become part of the syllabus of trainee teachers and refresher courses for experienced foreign language educators, at least in the form of a frontal talk.

REFERENCES

Bartlett, F.C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511759185

Beeching, K. (2014). Corpora in language teaching and learning. *Recherches en Didactique des Langues et des Cultures*, *11*(1), 1–19. https://doi.org/10.4000/rdlc.1672

Bennett, G.R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan, the USA: University of Michigan Press ELT. https://doi.org/10.3998/mpub.371534

Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. *Learner Corpora and Language Teaching. Studies in Corpus Linguistics*, *92*, 245–263. https://doi.org/10.1075/slcs.201.12cal

Enongene, E.E. (2013). English as a foreign language at the university of Yaonde 1: Attitudes and pedagogic practices. *English Language Teaching*, *6*(3), 57–71. http://dx.doi.org/10.5539/elt.v6n3p57

Friginal, E. (2018). *Corpus linguistics for English teachers*. Routledge: New York. https://doi.org/10.4324/9781315649054

Gorina, O.G. (2014). *The use of corpus linguistics technology for the development of lexical skills by regional studies students in the process of a professionally-oriented communication in English* [PhD thesis, St. Petersburg: Saint-Petersburg State University].

Gunnarsson, B.L. (1997). Language for Special Purposes. Encyclopedia of Language and Education. In G.R. Tucker, & D. Gorson (Eds.), *Encyclopedia of Language and Education* (Vol. 4, pp. 105–117). Dordrecht: Springer. https://doi.org/10.1007/978-94-011-4419-3_11

Johns, T. (1986). Micro-Concord: A language learner's research tool. *System*, *14*(2), 151–162.

Kogan, M., Zakharov, V., Popova N., & Almazova N. (2020). The impact of corpus linguistics on language teaching in Russia's educational context: Systematic literature review. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. HCII 2020. Lecture Notes in Computer Science* (Vol. 12205, pp. 339–355). Cham: Springer. https://doi.org/10.1007/978-3-030-50513-4_26

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139524759

Shi, G. (2017). Pedagogic processing of corpora and its effects on ESL learners' medium-term language awareness enhancement. *International Journal of English Linguistics*, *4*(4), 63–73. http://doi.org/10.5539/ijel.v4n4p63

Sysoyev, P.V. (2010). Linguistic Corpus in Teaching Foreign Languages. *Foreign Languages at School*, *5*, 12–21.

Trace, J., Hudson, T., Brown, J.D. (2015). An overview of Language for specific purposes. In J. Trace, T. Hudson, & J.D. Brown (Eds.), *Developing Courses in Languages for Specific Purposes* (pp. 1–22). Hawaii: National Foreign Language Resource Centre.

UNESCO. (2019). *ICT Competency Framework for Teachers*. Retrieved from https://iite.unesco.org/wp-content/uploads/2019/05/ICT-CFT-Version-3-Russian-1.pdf

Youssef, A.F. (2020). The effectiveness of corpus-based approach on vocabulary learning gains and retention in Saudi tertiary EFL context. *Journal of English Language Teaching and Applied Linguistics*, *2*(3), 1–15. http://doi.org/10.32996/jeltal.2020.2.3.1