

Application of a corpus-based approach in teaching English for specific purposes to Master's degree students of engineering and technical majors

Stepan A. Boyko, Elena A. Koltsova & Valentina A. Spiridonova

Saint Petersburg Mining University
Saint Petersburg, Russia

ABSTRACT: The article is concerned with the practical application of a corpus-based approach as a supplement to the traditional approach in teaching English for specific purposes in a non-linguistic university. The developed algorithm of corpus-based work, as part of assignments, is represented by four successive stages in accordance with the formed corpus-based analysis model. The authors present and discuss a pedagogical experiment conducted in classes with Master's degree students of engineering and technical majors at a Russian university, including the assessment criteria and scoring scale used in the experiment, as well as its results. Based on the results from the control and the experimental group, the experimental group demonstrated a higher effectiveness in technical translation and a higher increase in the professionally relevant basic vocabulary for corresponding majors. Thus, it can be concluded that the proposed corpus-based approach would better prepare engineering and technical students for language-related challenges in their profession.

Keywords: Corpus-based approach, ESP course, translation, professional terminology

INTRODUCTION

All over the world the widespread development of high information technology and the increasing requirements to educational standards result in the revision and reinterpretation of both existing approaches and their content. In the course of modernisation of the Russian system of higher education special attention is also paid to the processes of computerisation. Contributions to the development of new forms and approaches in the organisation of educational technologies are reflected in current scientific research publications [1-4]. In the methodology of teaching foreign languages an important role is allotted to computer technology, introduced to improve clarity, mobility and effectiveness of the learning process [5-7]. According to the Russian Federal State Education Standards for Higher Professional Education, the graduates of Master's degree programmes in majors must be able to use modern information and communication technologies, as well as global information resources, including those in a foreign language, to carry out professional interaction and research activities.

As part of this trend, new approaches to teaching English for specific purposes (ESP), such as corpus-based learning, are being introduced into the educational process. In recent years, aggregations of written texts (corpora) and verbal communications have found successful application in the process of teaching a foreign language, allowing educators to achieve a significant pedagogical effect [8-10]. Thus, being focused on applied language learning, by forming a search query among a significant amount of continuously added authentic language data, corpora provide not only quantitative characteristics, but also forms of real-life speech usage of the required language element. In this article, the authors analyse the use of a corpus-based approach as a supplement to the traditional approach in teaching English for specific purposes in a non-linguistic university. This approach is based on database searches for relevant language information that may be used in the learning process in order to expand the vocabulary of students, including active professional vocabulary, and to develop teaching exercises based on them.

THEORETICAL BACKGROUND

In linguistic education, the use of information and communication technologies is reflected in the use of corpus-based linguistics resources. Acting as information resources containing authentic language data, corpora are used in the process of learning a foreign language. Thus, many Russian and foreign scientists studied the theory and practice of corpus linguistics in order to form lexical skills [11-15]. In the modern foreign language teaching methodology, the inductive approach is used through exercises aimed at identifying the linguistic form of the statement. The use of such exercises is common in the learning based on the analysis of language databases and computer-based learning [16][17].

The special feature of this learning method consists in the use of authentic corpus linguistic material by learners when conducting independent research in the corpus and drawing conclusions based on the researched material. The main tools for such research are frequency indicators and concordances for a given student's search query in the corpus [18-20]. In this way, the acquisition of new knowledge about the language takes place in the classroom through discussion and testing of assumptions. As part of this approach to learning, the role of the foreign language instructor in the learning process shifts toward the advisor to the student-researcher. However, the aforementioned approaches are not yet widely used in teaching English for specific purposes in higher technical educational institutions in Russia. There are some highly specialised corpora being developed and packaged within the framework of university projects, but these resources often remain local and are not freely available on the Internet. At the same time, the wide application of corpus-based technology potential in the framework of training a Master's degree student majoring in engineering opens up opportunities for packaging the educational content with modern information technologies, which corresponds to the current tasks of the higher education system.

When forming the content of a foreign language course for specific purposes, the basic vocabulary is key and subject to direct selection, since the expected quality of foreign language communication, within the professional activity of a learner, directly depends on its compliance with the actual units of professional language communication [21-23]. An important criterion providing for the objectivity of using a corpus of texts in language teaching is the representativeness of its volume, ensuring a natural contextual form of the language phenomena under study. Bilingual dictionaries often do not contain sufficient examples of the use of professional terms in different contexts. In this regard, a careful selection of vocabulary is necessary, relying not only on specialised dictionaries but also on relevant corpora. Thus, by searching the corpus and processing the obtained natural language data, learners are able to study the quantitative and qualitative indicators of the basic vocabulary units in their majors.

In this regard, a trial hypothesis was formulated, which consists in the fact that the use of a corpus-based approach in teaching English for specific purposes promotes the formation of professional abilities and skills necessary for Master's degree students majoring in engineering to implement effective communication in the framework of professional discourse. The objective of learning is to develop the skill of using a corpus as a source of linguistic information, the skill of identifying the frequency and concordance of professionally significant linguistic units, as well as the ability to apply them in professional activities. Thus, the purpose of learning was determined, which consisted in teaching the identification of keywords, collocations, and the formation of professionally significant basic vocabulary.

MATERIAL AND METHODS

In the course of the experimental teaching at Saint Petersburg Mining University, Russia, the effectiveness of teaching English for specific purposes backed by the corpus was tested. Within the framework of experimental teaching linguo-didactic materials were formed on the basis of freely available corpora: Hong Kong Engineering Corpus (HKEC), British National Corpus (BNC) and Corpus of Contemporary American English (COCA). Quantitative and qualitative research methods were used and implemented in English language classes for specific purposes in the experimental teaching. The control group included 25 students (17 males and 8 females), and the experimental group also had 25 persons (14 males and 11 females). All participants were Master's degree students, and the experiment took place during two academic semesters in the 2020-2021 academic year. The teaching was conducted within the framework of the curriculum for the first year of Master's degree programmes: *Automation of Technological Processes and Production* and *Oil and Gas Engineering*.

The experimental teaching approach included a preparatory stage, an educational experiment, introductory and final testing, statistical processing and analysis of the results. At the preparatory stage, an introductory questionnaire to identify the needs of learners, as well as exercises and tasks backed by the corpus were developed to conduct the experiment. Master's degree students of the control group took part in the teaching experiment. Their classes were conducted by a traditional method. In addition to the traditional method, during the classes with students of the experimental group tasks were used which made it possible to form search queries in linguistic corpora.

As a part of the practical teaching method, Master's degree students in the experimental group, were familiarised at the introductory session with the principles of work with the corpus. Then, in the process of translation, they were offered to independently find in the original texts of the specialised English-language literature words and phrases related to the category of special terms and to identify the features of their use. The developed algorithm of work with the corpora, within the framework of performing tasks, is represented by four consecutive stages in accordance with the designed model of corpus-based analysis (see Figure 1 below).

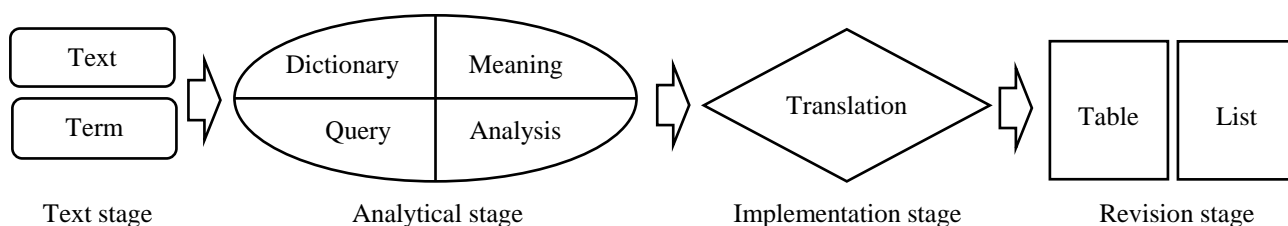


Figure 1: Model of corpus-based analysis of a technical text.

The model is represented by several consecutive stages:

Text stage:

1. Get an understanding of the original text;
2. Identify the term related to the subject area.

Analytical stage:

1. After referring to dictionaries, identify and analyse the primary meanings;
2. Choose the meaning most appropriate to the initial context from several options offered in dictionaries;
3. Compose a query in the search bars in the suggested corpora. Get an understanding of the frequency of the term and the concordance lines provided;
4. Determine the meaning and frequency of the term by analysing the natural context of the term in the concordances.

Implementation stage:

1. Translate the term according to the identified meaning that best fits its natural context of use.

Revision stage:

1. Present the results of the work with the terms identified in the original text in the form of a table of meanings and examples of their use in the natural context, including the distribution of their frequency;
2. Form a list of the most frequent terms of the subject area, including their meanings and contexts of use on the basis of the table for further inclusion in the final basic vocabulary.

The following assessment criteria were applied to the exercises:

1. Lexical aspects of the presentation - appropriateness of the presentation in terms of lexical and terminological units. No omissions of information in the final text of the translation, no redundant information or inaccurate interpretation of the text contravening the meaning of the original.
2. Grammatical aspects of the presentation - the adequacy of the presentation in terms of grammatical form and syntactic structures. No errors in the final text of the translation in regard to the choice of grammatical tense, gender, number or case.
3. Stylistic identity of the text - compliance with professional stylistics and generally accepted norms of literary language. No misuse of syntactic constructions of the original text in the final text of the translation.

The following scoring scale was used in the assessment:

From 10 to 19 points - failed. The text is not fully expounded (less than 30%) within the specified time period (60 minutes). The coherence of the text is breached. Mostly incorrect translation of the terms (less than 29% correct). Gross lexical, grammatical and stylistic errors in the text.

From 20 to 39 points - unsatisfactory. The text is not fully expounded (30-49%) in the specified time period (60 minutes). The coherence of the text is breached. Few terms are correctly translated (30-49% correct). Gross lexical, grammatical and stylistic errors in the text.

From 40 to 59 points - satisfactory. The text is not fully expounded (50-69%) in the specified time period (60 minutes). The text is coherent. At least half of the terms are correctly translated (50-69% correct). Minor lexical, grammatical and stylistic errors in the text.

From 60 to 79 points - good. Most of the text is expounded (70-89%) in the specified time period (60 minutes). The text is coherent. Most of the terms are translated correctly (70-89% correct). No lexical, grammatical and stylistic errors in the text.

From 80 to 100 points - excellent. The text and terminology contained therein are fully expounded (90-100%) in the specified time period (60 minutes). All terms are correctly translated (90-100%). No lexical, grammatical and stylistic errors in the text.

RESULTS

Within the framework of the experiment in English for specific purposes classes, the control group of 25 Master's degree students (17 males and 8 females), was taught using the traditional teaching method. Table 1 shows the results of the introductory and final testing of the control group of students in the programmes: *Automation of Technological Processes and Production* and *Oil and Gas Engineering* with the level of foreign language proficiency - intermediate plus.

Table 1: The results of the introductory and final testing of the control group.

Master's programmes	Students	Males	Females	Introductory testing - sum of points	Final testing - sum of points	Increase in the amount of points based on the test results
Automation of Technological Processes and Production	10	5	5	713	817	104
Oil and Gas Engineering	15	12	3	757	885	128
Total	25	17	8	1,470	1,702	232

According to the data shown in Table 1, one can observe an increase in the sum of points in the control group of the Master's degree students in the *Automation of Technological Processes and Production* major, as well as the Master's degree students in *Oil and Gas Engineering*. However, in order to confirm the validity of the data obtained in the experiment, the authors carried out the following calculation to determine the existence of significant differences in the students' performance (student's *t*-test). This calculation of *t*-distribution was carried out for two dependent samples - the results of introductory and final testing of Master's degree students of the control group. The absence of differences in the test results was taken as the null hypothesis (the mean value of the difference between the observations at the introductory and final test stages is zero):

$$H_0: \bar{Z} = 0, \quad z_i = x_{1i} - x_{2i} \quad H_1: \bar{Z} \neq 0$$

The calculation was based on the formula:

$$t = \frac{\bar{Z}}{s_z/\sqrt{n}} \sim t_{n-1} \quad (1)$$

Where:

$$s_z = \sqrt{\frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n-1}} \quad (2)$$

The result of the calculations was the value of $t_{Emp} = 7.7$. The calculated value of *t* was checked for falling into the critical area of significance. Its critical values t_{Cr} for $p \leq 0.05 = 2.06$ and for $p \leq 0.01 = 2.8$, thus the obtained empirical value of *t* (7.7) is in the zone of significance. Therefore, the null hypothesis is rejected in favour of the presence of significant differences in the test results for the control group. Based on the performed calculations, it is fair to say that the traditional teaching method used in the control group confirmed its effectiveness in teaching technical translation.

For the completeness of the teaching experiment, the data for the 25 Master's degree students in the experimental group (14 males and 11 females) is also presented. These classes were conducted according to the traditional method with an addition of linguistic corpus-based searches including the appropriate formation of search queries. In table 2, the authors present the results of the introductory and final testing of the experimental group of students in the programmes: *Automation of Technological Processes and Production* and *Oil and Gas Engineering*, with the level of foreign language proficiency - intermediate plus.

Table 2: The results of the introductory and final testing of the experimental group.

Master's programmes	Students	Males	Females	Introductory testing - sum of points	Final testing - sum of points	Increase in the amount of points based on the test results
Automation of Technological Processes and Production	7	4	3	651	792	141
Oil and Gas Engineering	18	10	8	845	1,184	339
Total	25	14	11	1,496	1,976	480

According to the data shown in Table 2, one can observe a considerable increase in the sum of points in the experimental group, both among Master's degree students in the *Automation of Technological Processes and Production* programme, as well as the Master's degree students in *Oil and Gas Engineering*. However, in order to confirm the validity of the data obtained in the experiment, the authors carried out the calculation according to student's *t*-test (as for the control group above).

The result of the calculations was the value $t_{Emp} = 10.1$. The calculated value of t was checked for falling into the critical area of significance. Its critical values t_{Cr} for $p \leq 0.05 = 2.06$ and for $p \leq 0.01 = 2.8$, thus the obtained empirical value of t (10.1) is in the zone of significance. Therefore, the null hypothesis is rejected in favour of the presence of significant differences in the test results for the experimental group. The summarised data of the teaching experiment is presented in Table 3.

Table 3: Summarised data of the teaching experiment.

Groups	Introductory testing - sum of points	Final testing - sum of points	Introductory testing - arithmetic mean	Final testing - arithmetic mean	Introductory and final testing - deviation	Introductory and final testing - squared deviation	Introductory and final testing - t value
Control	1,470	1,702	58.8	68.08	-232	3,038	7.7
Experimental	1,496	1,976	59.84	79.04	-480	11,420	10.1

According to the data shown in Table 3, demonstrating the results of the introductory and final tests, out of the two groups, the experimental group has the highest value of the sum of points, arithmetic mean and the t distribution. This indicates a greater effectiveness of teaching English for specific purposes supported by the corpus-based addition in the experimental group, compared with the traditional model of teaching used in the control group. The students in the experimental group not only identified keywords in the proposed tasks much faster than the students in the control group, but also determined their collocations, according to which they carried out the translation process and formed their basic vocabulary.

CONCLUSIONS

Based on the performed calculations, it can be concluded that in comparison with the results of the control group, the classes enhanced by the addition of the linguistic corpus-based searches used in the experimental group not only confirmed their effectiveness in teaching technical translation, but also significantly increased the acquisition of professionally relevant basic vocabulary for corresponding majors.

The type of tasks included in the experimental group, assists in the formation of professional skills and abilities necessary for Master's degree students of engineering and technical majors. Through these tasks, students can expand their vocabulary, lexical skills, and learn various meanings of words in different situations. They can determine the appropriate meaning according to the original context, identify the most frequent words and word combinations, and develop a professionally significant basic vocabulary that will contribute to effective communication in their professional practice.

REFERENCES

1. Akhiroh, N.S., Teaching English in non-English departments: empowering teacher towards improvement. *Indonesian J. of English Educ.*, 4, 1, 49-62 (2017).
2. Rafiq, K.R.M., Hashim, H. and Yunus, M.M., Sustaining education with mobile learning for English for specific purposes (ESP): a systematic review (2012-2021). *Sustainability*, 13, 9768 (2021).
3. Alizadeh, I., Discovering the identity and suitability of electronic learning tools students use in English for specific purposes (ESP) programs. *Call-Ej*, 19, 2, 213-229 (2018).
4. Murzo, Y.E. and Chuvileva, N.V., Use of information technologies in developing foreign language competence for professional interaction of undergraduate and postgraduate students specializing in mineral resources. *Inter. J. of Emerging Technologies in Learning*, 3, 144-152 (2021).
5. Mulyadi, D., Wijayatiningih, T.D., Singh, C.K.S. and Prastikawati, E.F., Effects of technology enhanced task-based language teaching on learners' listening comprehension and speaking performance. *Inter. J. of Instruction*, 14, 3, 717-736 (2021).
6. Huhta, M., Vogt, K., Johnson E. and Tulkki, H., *Needs Analysis for Language Course Design: a Holistic Approach to ESP*. Holl, D. (Ed). Cambridge: Cambridge Univ. Press (2013).
7. Pushmina, S.A., *Teaching EMI and ESP in Instagram*. In: Anikina, Z. (Eds), *Integrating Engineering Education and Humanities for Global Intercultural Perspectives*, Lecture Notes in Networks and Systems. 131, Springer, Cham, 475-482 (2020).
8. Butarbutar, R., Learner's perception of task difficulties in technology-mediated task-based language teaching. *Englisia: J. of Language, Educ., and Humanities*, 9, 1, 129-144 (2021).
9. Vyatkina, N., *Language Corpora for L2 Vocabulary Learning: Data-driven Learning across the Curriculum*. In: Ecke, P. and Rott, S. (Eds), *Understanding Vocabulary Learning and Teaching: Implications for Language Program Development*. Boston, MA: Cengage Learning, 121-145 (2018).
10. Boulton, A. and Cobb, T., Corpus use in language learning: a meta-analysis. *Language Learning*. 67, 2, 348-393 (2017).
11. Ratkaine Jablonkai, R. and Cebon, N., Corpora as tools for self-driven learning: a corpus-based ESP course. In: Khosrow-Pour, M. (Ed), *Language Learning and Literacy: Breakthroughs in Research and Practice*, 1, 166-190 (2020).

12. Plungyan, V.A., Corpus as a tool and as an ideology: on some lessons of modern corpus linguistics. *Russkij yazyk v nauchnom osveshchenii*, 2, **16**, 7-20 (2008).
13. Evert, S., *Corpora and Collocations. Corpus linguistics: an International Handbook*. Berlin, Germany: Mouton de Gruyter, 1212-1248 (2008).
14. Conrad, S. and Biber, D., *Real Grammar: a Corpus-based Approach to English*. Pearson, Longman (2009).
15. McEnery, T. and Hardie, A., *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press (2012).
16. Enayati, F. and Pourhosein Gilakjani, A., The impact of computer assisted language learning (CALL) on improving intermediate EFL learners. Vocabulary learning. *Inter. J. of Language Educ.*, 4, 96 (2020).
17. Tafazoli, D., Huertas Abril, C.A. and Gómez Parra, M.E., Technology-based review on computer-assisted language learning: a chronological perspective. *Pixel-Bit. Revista de Medios y Educación*, 54, 29-43 (2019).
18. Boulton, A., Corpora in language teaching and learning. *Language Teaching*, 50, **4**, 483-506 (2017).
19. Bennet, G.R., *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Michigan: University of Michigan Press (2010).
20. Krueger, R., Working with corpora in the translation classroom. *Studies in Second Language Learning and Teaching*, 2, **4**, 505-525 (2012).
21. Sturza, A. and Supuran, A., English for specific purposes courses for engineering sciences, a necessary requirement. *Annals of Faculty of Economics*, 1, **1**, 707-711 (2017).
22. Akmal, A.N., Hidayati, N. and Farah, R.R., Exploring the English needs of mechanical engineering students in ESP class: Indonesian university context. *English Learning Innovation*, 1, **1** (2020).
23. Rasyimah, Ahyar, J. and Sari, D.K., Challenges in designing ESP course for first year students of engineering faculty at Malikussaleh University. *Emerald Reach Proc. Series*, 1, 563-568 (2018).

BIOGRAPHIES



Stepan A. Boyko is an Associate Professor of English at Saint Petersburg Mining University in Saint Petersburg, Russia. He graduated a specialist of foreign languages teaching from the Tomsk State University, Russia, and attained a PhD in pedagogy from the Tomsk State University in 2014. He has published a number of articles dealing with pragmatics, discourse studies and text linguistics. His sphere of interest includes engineering education, translation studies, corpus-based approach to teaching translation and cognitive aspects of language teaching.



Elena A. Koltsova is an Associate Professor of English at Saint Petersburg Mining University in Saint Petersburg, Russia. She graduated a specialist of foreign languages teaching from the Ivanovo State University, Russia, and attained a PhD in philological sciences from the State University in Saint Petersburg in 2009. She has published a number of articles dealing with pragmatics and communication in politics, business and interpersonal relations. Among her other research interests are the cognitive functions of language and metacognition in teaching English for academic and specific purposes.



Valentina A. Spiridonova is an Associate Professor of English at Saint Petersburg Mining University in Saint Petersburg, Russia. She graduated a specialist of foreign languages teaching from the Ivanovo State University, Russia, and attained a PhD in philological sciences from the State University in Saint Petersburg in 2017. Her research expertise is in the interdisciplinary area, including such disciplines as Internet linguistics, psycholinguistics, media and discourse studies, text linguistics and English linguistics.