

UDC 81

<https://doi.org/10.33619/2414-2948/48/51>

**CORPUS-BASED STUDY OF ENGINEERING TERMS:
LINGUISTIC AND PEDAGOGICAL IMPLICATIONS**

©*Nurmatova G.*, ORCID: 0000-0002-4301-8089, Navoi State Mining Institute,
Navoi, Uzbekistan, g.nurmatova77@gmail.com

**КОРПУСНОЕ ИССЛЕДОВАНИЕ ИНЖЕНЕРНЫХ ТЕРМИНОВ:
ЛИНГВИСТИЧЕСКИЕ И ПЕДАГОГИЧЕСКИЕ РАССУЖДЕНИЯ**

©*Нурматова Г. Х.*, ORCID: 0000-0002-4301-8089, Навоийский государственный горный
институт, г. Наваи, Узбекистан, g.nurmatova77@gmail.com

Abstract. The article discusses the corpus study of engineering terms for linguistic research — the extraction and classification of terms identified by the frequency of words in the corpus and pedagogical implications — the application of effective results to the teaching and learning of engineering terms in engineering universities of Uzbekistan. In the course of the study, computer programs AntGorGen were used to create the target corpus and AntConc for analyzing the lexis of research articles. This article offers further corpus-based studies with consideration of the acquiring and comprehension of engineering terms for developing engineering students' research skills.

Аннотация. В статье рассматривается корпусное исследование инженерных терминов для исследования лингвистических — извлечение и классификации терминов выявленных частотности слов в корпусе и педагогических проблем — применение эффективных результатов, в преподавание и обучение инженерных терминов в инженерных вузах Узбекистана. В ходе исследования были применены компьютерные программы AntGorGen для создания намеченного корпуса и AntConc для анализов лексики научных статей. Данная статья предлагает дальнейшие корпусные исследования с рассмотрением проблем по овладению и пониманию инженерных терминов у обучающихся для развития научных навыков.

Keywords: engineering, terms, corpus-based study, linguistics, pedagogy.

Ключевые слова: инженерия, термины, корпусное исследование, лингвистика, педагогика.

Introduction

In the modern world, in the era of the progress of science and technology, we are increasingly confronted with the words of different industries and spheres, gradually appropriating and using them not only in professional, but even in common speech. These words may refer to different spheres, among which engineering terminology is the most progressively developing.

In spite of the globalization of the English, there are still actual problems in teaching and learning of field related lexis in non-linguistic universities of Uzbekistan. As one speaker of the seminar in Samarkand organized by British Council, said: “We, researchers speak in different languages about the same things”. In spite of the widely usage and strong command of English

language, it is still difficult to understand the communication of people i.e. researchers' language. Students and academics with already existing general language knowledge have already aptitude to explore further perspectives of their study field referring to foreign publication. Therefore, the second language learner (as an engineering specialist) will not be able to demonstrate his full potential or receive and use expectable information from the source without appropriate interpretation of the term in the English language. If we refer to the quotation in terms of teaching English for research, we can obviously see there is a real matter to think. In modern linguistics, as A. Laurence (2013) notices, corpus-based studies of engineering and related sciences are less studied by linguists than in other social (business, legal) fields [1]. According to Lawrence's corpus analysis, the percentages of linguists' studies in Asian ESP, ESPJ, ESPJ, ESP World (included in SCOPUS data base) journals in engineering and engineering related fields has been given less attention in comparison with other fields, such as Business or Medicine [2]. Since the aim of this research is to extract engineering terms from a relevant source, it would be important to mention how multifaceted the engineering field. According to Wikipedia, there are 174 fields of engineering. However, in our study all these engineering fields will not be included for corpus analyses, only the relevant fields of engineering. The advantageous of corpora is, it allows analyzing massive texts and only via corpus-based studies assists in exploring big number of selected articles as well as maintains balance and representativeness of a corpora [3].

Although, the problem of terminology is not new, but the engineering terminology, especially on corpus-base has rarely been of great importance for linguistic researchers.

Among the studies on terms and terminology, the works of O. A. Makarihina is worth attention, where the terms are divided according to the following signs:

1. Term - a word or phrase;
2. The term expresses a special concept;
3. The term is used in a special field

Analyzing O. A. Makarihina's criteria, terminology should be considered from both sides: the terms of the language and the terms of speech. The boundaries between them are determined by their definition and functionality. However, Portuguese linguist M. Cotter (2006) promotes two problems related with terms extraction on corpus-base study [2]:

1. Identification of terms in a data;
2. Extraction of terms from data

In order to find solution to this problem, he suggests to analyze *Word Frequency List* according to the highest, the medium and the least frequently usage, it is possible to identify if the lexical unit is used as a *word* or a *term*. Following the Cotter's idea O. Mudraya (2006) distinguished terms into the following categories [4]:

- Technical words;
- Non-technical words;
- Sub-technical words.

Materials and Methods

156 journals of PLOS ONE¹, that have Info Factor above and publish original research papers in the field of engineering science, and specifically one of its most common areas taught in

¹ PLOS applies the Creative Commons Attribution (CC BY) license to articles and other works we publish. If you submit your paper for publication by PLOS, you agree to have the CC BY license applied to your work. Under this Open Access license, you as the author agree that anyone can reuse your article in whole or part for any purpose, for free, even for commercial purposes. Anyone may copy, distribute, or reuse the content as long as the author and original source are properly cited. This facilitates freedom in re-use and also ensures that PLOS content can be mined without barriers for the needs of research.

Uzbekistan engineering universities in the 18 engineering fields. Using the AntCorGen and AntConc² software program, the most frequently used terms are identified, this in its turn will be useful for scholarly writing and developing comprehension skills of ESP students, that are both will serve for career prospects and eventually will assist in developing materials and syllabi for ESP classroom.

The most quantity of terms according to J. M. Swales and C.B. Feak appear in research articles, where the structure of IMRD is a matter of importance and its parts can be considered as the data for corpus formation and even design [6]. Since the author used a different approach with the aim to match domain categories of the corpus with the Uzbek engineering universities' fields and academic subject areas, this corpus is considered as a *domain-specific* or *genre-specific* that involves developing static corpus, which is more appropriate for this type of corpora.

Therefore, the first theme of our study is directly based on corpus-based analyses, mainly based on scholarly articles of the engineering fields. Since terms are regarded as academic words, but with the less frequency usage, we decided to study them according to the frequency of terms pursuing to develop better materials for engineering learners with the aim to comprehend scholarly articles as well for their future career. It has also been observed in works of the above mentioned scholars where they give clear analyses of terms frequencies extracted from different corpora. Eventually, corpus-based study represents authentic materials that allow delivering the lessons of ESP in a more natural contexts and its effectiveness has already been discussed by several scholars [5].

However, the effectiveness of it, particularly of comprehension those terms in the Uzbek learners classroom has not been tested yet, which we are going to conduct in our research later as well. So, the themes of our study are: corpus analyses of terms and creating wordlist of most frequently used terms, secondly, developing corpus-based materials, and thirdly, comprehension of terms by Uzbek learners. In our study there the research question is the following:

1. Can the most frequently used lexis of engineering research articles be considered as terms?
2. If yes, how they can be mined from corpus?

Having studied Makarihina's and Mudraya's views on classification of terms, we decided to study engineering terms both from the linguistic and speech point view. For the language sphere, we will analyze engineering terms with an approach to traditional linguistics, and for the speech sphere — from non-traditional — corpus linguistics, which will be discussed in the further steps of the study [6].

Results and Discussions

Nearly four million words have been analyzed and frequency list has been created and analyzed. After cleaning our raw data (cleaning from functional words such as articles, prepositions, conjunctions), we manually extracted the first 100 frequently words of each domain. These 100 words have been selected as patterns for further implications of our project. The result shows that not all of them can be regarded as “terms”; therefore we continued analysis with concordance tool and extracted terms according to the context displayed on the concordance screen. From the frequency list we generated the list of terms of each domain, then we elicited the terms that belong to all domains (Table).

² Anthony, L. (2019). AntCorGen (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>

Table.

EXCERPT OF WORDS AND DOMAINS

<i>Term</i>	<i>Appearing Frequency in these domains</i>			
Control	Automotive	Control	Mining	Mechanical
Ecological	Ecology	Architecture	Mining	Transport
Energy	Energy and power	Electrical	Industrial	Mechanical
Fracture	Mining	Chemical	Ecology	Fracture
Power	Architecture	Energy and power	Electrical	Industrial
System	Aerospace	Automotive	Control	Electronic
Thermal	Energy and power	Geological	Technological	Mechanical
Wave	Geological	Mining	Environmental	Electrical

This type of analysis, furthermore, give possibility to analyze the other part of the frequency list and match them with concordance screen context in order to elicit term, additionally, to identify in which domain a particular word most often appear, or to develop lists according to the domain [7].

Students of non-linguistic university are tend to obtain information of their field of study in the English language as it is easier to digest it for them. Moreover, students of higher level, masters and PhDs are more even interested in studying of research of other scholars and even writing research articles in English with the aim to introduce and share their innovative ideas to other researchers of other countries.

References:

1. Anthony, L., & Bowen, M. (2013). The language of mathematics: A corpus-based analysis of research article writing in a neglected field. *Asian ESP Journal*, 9(2), 5-25.
2. Cotter, M. J. (2006). Teaching terms: a corpus-based approach to terminology in ESP classes. *In Proceedings of the 5 th International AELFE conference*.
3. McEnery, T., McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.
4. Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25(2), 235-256. <https://doi.org/10.1016/j.esp.2005.05.002>
5. Maher, P., & Milligan, S. (2019). Teaching master thesis writing to engineers: Insights from corpus and genre analysis of introductions. *English for Specific Purposes*, 55, 40-55.
6. Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills (1)*. Ann Arbor, MI: University of Michigan Press.
7. Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes*, 33, 54-65. <https://doi.org/10.1016/j.esp.2013.07.001>

Список литературы:

1. Anthony L., Bowen M. The language of mathematics: A corpus-based analysis of research article writing in a neglected field // *Asian ESP Journal*. 2013. V. 9. №2. P. 5-25.
2. Cotter M. J. Teaching terms: a corpus-based approach to terminology in ESP classes // *Proceedings of the 5 th International AELFE conference*. 2006.
3. McEnery T. et al. *Corpus-based language studies: An advanced resource book*. Taylor & Francis, 2006.
4. Mudraya O. Engineering English: A lexical frequency instructional model // *English for Specific Purposes*. 2006. V. 25. №2. P. 235-256. <https://doi.org/10.1016/j.esp.2005.05.002>

5. Maher P., Milligan S. Teaching master thesis writing to engineers: Insights from corpus and genre analysis of introductions // English for Specific Purposes. 2019. V. 55. P. 40-55.

6. Swales J. M., Feak C. B. Academic writing for graduate students: Essential tasks and skills. – Ann Arbor, MI: University of Michigan Press, 2004. V. 1.

7. Hsu W. Measuring the vocabulary load of engineering textbooks for EFL undergraduates // English for Specific Purposes. 2014. V. 33. P. 54-65. <https://doi.org/10.1016/j.esp.2013.07.001>

*Работа поступила
в редакцию 14.10.2019 г.*

*Принята к публикации
19.10.2019 г.*

Ссылка для цитирования:

Nurmatova G. Corpus-based Study of Engineering Terms: Linguistic and Pedagogical Implications // Бюллетень науки и практики. 2019. Т. 5. №11. С. 410-414. <https://doi.org/10.33619/2414-2948/48/51>

Cite as (APA):

Nurmatova, G. (2019). Corpus-based Study of Engineering Terms: Linguistic and Pedagogical Implications. *Bulletin of Science and Practice*, 5(11), 410-414. <https://doi.org/10.33619/2414-2948/48/51> (in Russian).