

## HIGH FREQUENCY WORDS AND FOREIGN LANGUAGE COMPETENCE DEVELOPMENT IN ENGINEERING

**Tatiana Panyшева**

Peter the Great Saint-Petersburg Polytechnic University, Russia  
tpanycheva@inbox.ru

**Abstract.** The results of small-scale specialized corpora study are based on web texts in innovative engineering fields of Robotics and the Internet of Things. High Frequency word use is discussed providing with their collocation list and paying attention to frequent collocation distribution within texts relating to general robot features and use. Frequent word concordance lines were extracted from the corpus by AntConc Free Internet Tool and the possible options of collocation and concordance use in order to produce assignments for English for Specific Purposes Classroom are discussed. The aim of the paper is to demonstrate how corpora study can provide students with examples of adequate academic and terminological word use. Concordances also give a teacher a set of collocation and word patterns and help produce tasks aiming at professional vocabulary acquisition in special occupational areas. This approach can both expand the students' vocabulary and deepen understanding of professional word meaning and use in special scientific and engineering fields. Small corpora analyzes is a quick method to find and introduce new words and collocations within fast developing areas of science and engineering. The concordances retrieved from the corpora can serve multifunctional purposes providing understanding of semantics, collocation use and contextual aspects within specific occupational areas. The results of the concordances study also can provide students, lecturers and engineers with high frequency collocations and ready-to-use speech patterns for adequate communication on occupational topics.

**Keywords:** English for Specific Purposes, interdisciplinary learning, corpus linguistics, academic words.

### Introduction

The modern process of global higher education space creation and the need for academic exchange of ideas [1] highlight the importance of communicative foreign language competence development. Nowadays, worldwide communication within engineering and scientific areas can serve the purpose of further global development. The potential of this process is dependent a great deal on successful formation of occupational foreign language lexical competence [2; 3]. However, foreign language vocabulary acquisition by non-linguistic students, and teaching English for Specific Purposes to future engineers demonstrate some specific problems arising from the need to introduce the new words in a special occupational context. Nowadays, it is widely discussed that lexical competence is based not only on the new words translation and pronunciation practice but you need to know in what context the word can be introduced, word semantics and what synonyms and opposites can be used in the specific context to share the ideas successfully [4;5]. It is also important to use relevant to the professional area collocations in order to produce adequate statements and express your ideas [6]. However, both lecturers and students can face problems arising from the lack of examples of word use within the specific occupational context.

The recent developments in corpus linguistics [7-10] have provided teachers with helpful tools, which can boost the development of a new approach in lexical competence development based on production of classroom assignments with the close attention to word frequency in different occupational areas. This can help to highlight specific subject-based words and word collocations for further introduction in the classroom. The vocabulary of different professional areas in engineering can demonstrate variation in academic words and different terminological systems. So it is important for a lecturer to know the vocabulary features within each narrow professional field, which can be provided by corpus investigation. Corpus Linguistics can provide a teacher with some important frequent speech formulas and evidence of word use and can help select frequent words and frequent word collocations, which are needed to produce adequate statements.

The aim of this paper is to study words distribution and use in the informative engineering texts from the Internet devoted to the area of "The Internet of Things" (IoT) and Robotics, which cover also the innovative areas in agriculture technology, smart vehicles and related texts on robotics and computing.

## Methods

The small specialized corpora based on web pages were created to study corpora word frequency statistics and high frequency word concordances. All the texts were selected and retrieved from on-line pages within the period of 2012-2015. The papers included into the first corpus describe modern trends in Robotics and the second corpus to study is on innovations in computing and robotics tackling the topic “The Internet of things” and it also presents the texts based on extracted parts from on-line papers 2012-2015. The specialized informative texts were devoted to the innovative use of computing and robotic facilities integrated under the title of IoT or “The Internet of things”, which includes description of smart devices and the state-of-the art engineering topics of devices communication with each other so as to fulfill the tasks on their own due to innovations in computing. AntConc tool [9], designed to carry out corpus linguistic research and to provide with text statistics (word frequency and word collocation frequency), was used to study the corpora as it follows:

- firstly corpora texts words frequency statistics study was carried out according to AntConc Word list Tool [9], which provides you with the list of all the words in the corpus and reveals how frequently they occur in the text, then the high frequency terminological and academic words were selected;
- secondly, concordance lines for the high frequency academic and terminological words revealed with AntConc Word List Tool were extracted with the help of AntConc Concordance Tool [9] to see how the selected words are used in the corpus;
- finally, AntConc Plot Tool[9] was used to find the position of the word in the text so as to get additional information on the word use within the context and to check if the word use is within the investigation scope.

The collected data were analyzed in terms of possible use in the production of assignments for the English for Specific Purposes (ESP) classroom. Concordances were also studied to provide information both on words frequency and the words environment within the texts. This means that we could see contextual word meaning and use and analyze the high frequency words and their word patterns and also study which academic words and terminological words go together in the texts on Robotics and IoT.

The corpora under study can be described as a collection of authentic on-line English language texts on the following subjects:

- texts on the Internet of Things (44975 words);
- texts on Robotics (23253 words).

Both types of Corpora included texts on Agricultural innovations in IoT and in Robotics. As the main target of the research was to study the way the words are used in the texts in Robotics and IoT the high frequency words were analyzed not only as key concordance words, but they were highlighted in the texts and thoroughly analyzed in terms of their lexical environment, their meaning in the context and their semantics.

## Results and discussion

The specialized corpora based on informative online texts were analyzed according to Baker et al, 2006 approach [8], and the word frequency statistics based on AntConc Tool is presented in Table 1. As AntConc Tool provided with the number of words in the Robotics and IoT corpora the academic and terminological word frequency was normalized to the text length of 1000 words so that the results for different length texts corpora could be compared. Having analyzed the corpora under the terms of the word frequency we were able to highlight the high frequency terminological and academic words, which represent different aspects of Robotics and IoT (Table 1).

As it is demonstrated in Table 1 the most frequent words revealed in Robotics corpora were: *application, robot, robotics, human, humanoid, sensors, system*

The Internet of things corpora study demonstrated the following high frequency words: *application, data, human, robot, sensors, system, smart*

Then the concordances of the high words presented in Table 1 were carefully studied with AntCon Concordance Tool[9] to highlight the collocations that can represent authentic speech patterns used to give the general idea of Robotics, the Internet of Things and Smart Agriculture.

Table 1

### High frequency academic and terminological words in the robotics corpora and the IoT corpora

Word	Word Frequency in Robotics Corpora	Word Frequency in the IoT Corpora
application	4.36	5.25
automation	2.17	0.68
autonomous	2.12	0.79
data	0.12	7.67
human	11.43	4.58
humanoid	2.04	0.51
interaction	1.24	0.56
intelligence	1.87	0.11
robot	18.23	7.52
perform	2.04	0.98
robotics	5.33	1.44
sensors	4.65	5.60
system	6.57	4.41
smart	0.01	4.57
walking	2.75	0.01

The next step was to check the word use and the word environment in the corpora texts. So we retrieved concordance lines of the frequent words demonstrated in Table 1 to see if they could be used as recourses to create assignments for the ESP Classroom. The corpora files were inserted into the Corpora Files Field of AntCon Concordance Tool (Fig.1) and the term to be studied was entered into the query box and AntCon retrieved concordance lines from the corpora files. An example of the concordance lines for word *robot* is presented in Fig. 1. The total for concordance lines within the corpus texts for the word *robot* is displayed as Concordance Hits and equals, for example, 419 for the Robotics Corpus.

The next step was to find frequent collocations while examining the concordance windows. For example, we retrieved from the concordances and studied the following collocations, which reveal the way the words can be used in the classroom:

*Agricultural robotics;*

*Agricultural robots;*

*Agricultural robot platforms.*

We revealed that the above collocations in concordances go together with the following “chunks of words”:

*a strong trend towards;*

*enable safe, efficient, and economical agricultural production;*

*limit the growing pressure on the environment;*

*able to perform a wide range of tasks;*

*designed to perform a wide range of agricultural tasks.*

Having studied these results and being provided with the quick reference to the text provided by the AntConc Plot tool [9] we can produce the following task enabling the students to produce statements relevant to authentic English (Table 2). The task target is speech production on the Topic “There is a strong trend towards more agricultural robots. Why?”.

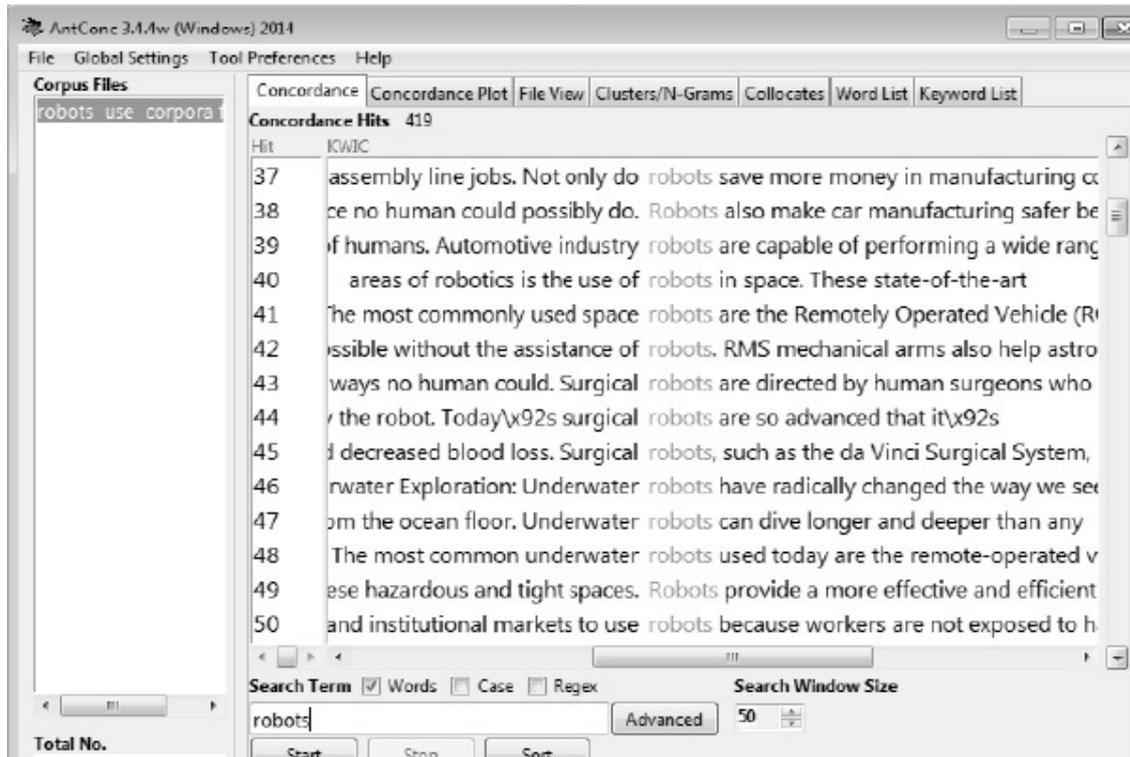


Fig. 1. Example of concordances: 37-50 concordance lines for a high-frequency term *robots*

Table 2

**There is a strong trend towards more agricultural robots. Why?  
Match the phrases in the columns (A, B, C)**

A	B	C
-Agricultural robotics -Agricultural robots -Agricultural robot platform -Agricultural Io T solutions	Can Are Is Provide (s)	Enable safe, efficient, and economical agricultural production Limit the growing pressure on the environment Able to perform a wide range of tasks Designed to perform a wide range of agricultural tasks Transforming agricultural industry Farmers with new facilities

If we take the engineering areas of Robotics and IoT we can see that frequent words can be categorized by different semantic fields.

Words *autonomous, intelligence and humanoid* were revealed to be high frequency words in the Robotics Corpora.

Their high frequency collocations with the word *robots* extracted from the corpora texts are the following:

*autonomous robots, robots Intelligence and humanoid Robots.*

IoT concordances demonstrate high frequency collocations: IoT technology, data-gathering sensors, machine-generated data, data privacy, to transfer data, human-to-computer interaction, robot-human interaction

The attempt to integrate high frequency words into special semantic fields can also be demonstrated by the most frequent collocations describing the semantic field of Robots categorized by

their use: *cleaning robots, industrial robots, medical robots, military robots, production robots, space robots, surgical robots, tactical robots, underwater robots.*

To put the robots into categories according to their movements we highlighted the following collocations retrieved from concordances: *wheeled robots, mobile robots, unmanned robotic vehicles, autonomous robotic vehicle.*

The concordances also helped us retrieve some useful word patterns with frequent collocations, which can be exemplified by the pattern with the frequent collocation “*human intervention*”, which was found to go together with other collocations forming frequent patterns in the studied texts:

*make decisions without human intervention;*

*to automate processes without human intervention;*

*work without direct human intervention.*

The high-frequency words *human* and *humanoid* in Robotic corpora and also the word *human-like*, which gives frequent collocations within the semantic fields of the robot description, are:

*Human-like appearance;*

*To be human-like;*

*Human-looking robots;*

*Human-like behaviour;*

*Humanoid body form;*

*Super- human lifting power.*

The same way the frequent collocations and word patterns with the high frequency words from Table 1 were retrieved from concordances with the aim to discuss the problem of Human- Robot interaction. The collocations retrieved from the corpora are:

*assist a human user, displace human jobs, replace humans, replace human employees, occasional human involvement, hazardous to human workers, keep human workers safe.*

The above collocations and word patterns not only represent frequent words but grouped in the special way acquaint briefly, like key words and key collocations, both the teacher and the student with the problems of human-robot interaction and can be used in the classroom discussions and assignments production. Additionally, the observed approach supports the idea of interdisciplinary learning in English for Special Purposes classroom, which nowadays is considered to be an essential element within modern curriculum [11-13].

## Conclusions

1. Word frequency count in small specialized corpora on engineering with further collocations study can serve as a tool providing with academic and terminological words within the special occupational area of foreign language acquisition.
2. The concordances retrieved from the corpora serve multifunctional purposes providing bases for discussion of semantics, collocations and contextual aspects.
3. The results of the concordance study also can provide students, lecturers and engineers with collocations and ready-to-use speech patterns for adequate communication on occupational topics.
4. Assignments based on corpora analyses of recent publications meet the demand for interdisciplinary learning and increase the students' awareness not only of word patterns and their use, but also of modern trends in engineering and science.

## References

1. EHEA Ministerial Conference “Beyond the Bologna Process: Creating, and connecting national, regional and global higher education areas”. Bucharest, 2012.[online] [3.03 2015]. Available at: [http://www.ihea.info/uploads/documents/BPF\\_Statement\\_27042012with add](http://www.ihea.info/uploads/documents/BPF_Statement_27042012with_add).
2. Robinson P.J. A rich view of lexical competence. *English Language Teaching Journal*, vol. 43(4), 1989, pp. 274-282.

3. Besukladnikov K.E., Shamov A.N., Novoselov M.N. Modeling of Educational Process Aimed at Professional Foreign Language Acquisition. *World Applied Science Journal*, vol. 22(7), 2013, pp. 903-910.
4. Chunming Gao, Bin Xu. The Application of Semantic Field Theory to English Vocabulary Learning. *Theory and Practice in Language Studies*, vol. 3, 2013, pp. 2030-2035.
5. Панышева Т.Ю. Использование семантических полей в формировании иноязычной лексической профессиональной компетенции у студентов неязыковых вузов (Semantic Fields and Foreign Language Professional Competence Development among Non-Linguistic University Students) *Вопросы методики преподавания в вузе, Teaching Methodology in Higher Education*, vol. 17, No 3, 2014, pp. 314-317.
6. Ozola E., Corpus Based Comparative Analysis of English and Latvian Terms of land Administration. *Proceedings of International Conference "Rural Environment. Education. Personality 2015"*. ISSN 2255-808x, pp. 429-424.
7. Baker P., Hardie A., McEnery A. *A Glossary of Corpus Linguistics*. Edinburgh University Press. 2006. 192 p.
8. Baker P. Quering Keywords, Questions of Difference, frequency and Sense in Keywords Analysis. *Journal of English Linguistics*. vol. 32, 2004, pp. 346-359.
9. Anthony I.A. *Freeware Corpus Analysis Toolkit for Concordancing and Text Analysis*, 2013. [online] [06.2015]. Available at: <http://www.laurenceantony.net/software/antconc>
10. O’Keeffe A., McCarthy M., Carter R. *From Corpus to Classroom. Language Use and Language Teaching*. ISBN 9780521616867. Cambridge University Press, 2007. 333 p.
11. Попова Н. В. Дидактическое моделирование вузовского учебника по иностранному языку с учетом междисциплинарных связей (Didactic Modeling of University Foreign Language Course Book with Interdisciplinary Links ). *Научно-технические ведомости СПбГПУ*. No 2, 2010, pp.102–109. (In Russian ).
12. Байденко В.И. Болонский процесс, проблемы, опыт, решения (Bologna Process, problems, experience and solutions.). Москва. Исследовательский центр подготовки специалистов, 2006. 128 p. (in Russian).
13. Caviglia-Harris J., Hatley J. Interdisciplinary Teaching, Analyzing Consensus and Conflict in Environmental Studies. *International Journal of Sustainability in Higher Education*. vol. 5, 2004, pp. 395-403.