CONCORDANCES AND HIGH FREQUENCY WORDS IN ENGLISH FOR SPECIFIC PURPOSES CLASSROOM

Tatiana Panysheva[1], Sergey Obukhov[2]
[1]Peter the Great Saint-Petersburg Polytechnic University, Russia;
[2]A.F. Ioffe Physical-Technical Institute, Russia
tpanycheva@inbox.ru, sobukhov@inbox.ru

Abstract. Vocabulary acquisition in English language class is based not only on the number of words students remember but on the ability to use them in production of statements. This ability can be developed by introduction of word collocations in the classroom. Here, studying frequent terms and academic words collocations used in Microelectronics and Microelectronics application in agriculture, we demonstrate that in case of English for Specific Purposes this task is much harder, as different scientific and occupational areas can produce uncommon for everyday English collocations and new variety of word patterns. The research is based on papers from scientific journals and books dealing with the fast developing branch of semiconductors, semiconducting materials and their practical use in fabrication of microelectronic devices. The list of highlighted frequent collocations was used to develop texts and corresponding to them tasks including gap filling, word matching. The next step was to prepare texts, in which students could find collocations and form the list of collocations with high frequency academic words on their own as self-study. The presented here approach aims at the formation of lexical competence based on the academic word and term acquisition and their adequate to native speakers use within the occupational areas.

Keywords: English for Specific Purposes, lexical competence, corpus linguistics, interdisciplinary learning.

Introduction

The fast development of communication technology has boosted the need for the exchange of ideas across the scientific and engineering institutions on a world scale. We communicate ideas through words, which makes English language lexical competence so important for every explorer as English is widely recognized as the language of international meetings and conferences [1-3]. This makes nowadays English language lexical competence an important factor of success and stimulates further development of innovative approaches to vocabulary introduction in English for Specific Purposes (ESP) classroom.ESP learning means introduction of terminological and academic words and training how to use them in speech. This can be understood as not only making students write out words and memorizing them but rather teaching to find word collocations, word chains and even combinations of collocations and word chains appropriate to the specialized engineering area.

Nowadays, with the fast development of computer-based corpora research [4-7] it can become a promising and reliable tool in the vocabulary exploration and teaching. Corpus is a collection of texts (written or spoken) prepared to be processed by computer, so that the vocabulary and grammar of the language under study could be described. This can serve a solid base for the description of special feature of special texts and their vocabulary in the variety of scientific and engineering areas. So, corpora exploration can provide ESP lecturers with important information on academic words and terms frequency and the environment within the occupational area language of which they teach. The list of frequent academic words can be found in AWL [8]. This collection of words with exercises helps learners study the vocabulary items, which can be common for most academic areas. D. Porter [9] exploring AWL provided teachers and students with a wide range of exercises aiming at academic vocabulary acquisition regardless the students' future occupation.

Corpora exploration can also help reveal differences and similarities between languages we study [10], high frequency words extraction can help to build core vocabulary for SAT and ACT exams [11]. The recent computer-aided research of texts demonstrated great potential of corpora–based exploration of high frequency words and their use in written and colloquial speech [12–14]. We suggest that while studying ESP texts we should additionally reveal not only common for all academic areas words, but some special features of occupational word use. This can be achieved in case the texts are extracted from scientific journals, books and the Internet sites devoted to the chosen subject, so that they can help identify special speech patterns within the exact engineering area.

1115

In this article we aim to discuss problems arising in teaching vocabulary in ESP English classroom and to highlight the advantage of corpora based approach incorporated into tasks design and aiming at occupational vocabulary acquisition and communicative training.

Methods

The small specialized corpora was formed from informative texts of the Internet articles and web pages devoted to Smart Technologies in Agriculture and Smart Engineering Solutions (64014 words). The retrieved texts belong to the period of 2012-2017. Also a small specialised corpora of informative texts (19000 words) was used to study the vocabulary of electronic devises description (EDD corpora). The AntConc Tool, 2018 [15] was used to explore data based on text statistics providing with word frequency, frequent words collocations and their concordance patterns. First the list of word frequency was generated with AntConc Word List Tool. Then high frequency academic and terminological words were selected. The next step was to retrieve concordances lines with AntConc Concordance Tool and to select frequent words collocations. The final step was to find the collocations in the text with AntConcPlot Tool to check if the word use was within the studied occupational area scope. Then the chosen frequent words, their collocations and speech patterns were evaluated to find the best way to introduce them in the assignments for ESP classroom.

Results and discussion

The study of specialised small corpora based on texts on Microelectronic devises use in Smart Technologies for Agriculture and Smart Engineering Solutions highlighted the following most frequent academic and terminological words:

nouns: sensors, output, microchip, voltage , effect, detection, identification, implantation, current, device, figure, applications, technology, signal, data ,system, field , magnet,

verbs: to operate, to sense,

adjectives: magnetic, smart, digital.

The most frequent prepositions are: of, for, from by, in, with, off.

All words above demonstrate frequency over 1,5 per 1000 words. The most frequent words are demonstrated in Table 1, where academic and terminological word frequency is normalised to the text length of 1000 words to make the frequency results comparable to the corpora of different length.

Table 1

High frequency academic and terminological words in the corpora

Word	Word frequency	Word	Word frequency
sensor	14.0	implantation	3.0
microchip	11.5	current	2.7
magnetic	8.5	smart	2.4
output	6.1	data	2.0
field	4.6	system	2.0
effect	4.3	point	1.8
device	4.2	sensing	1.5
identification	3.3	applications	1.3
voltage	3.3	operate	1.3

To expand the results of the word frequency study and make them useful for assignment production we also studied the concordances of frequent words, so that we could see how frequent words are used in the selected texts. The study was backed up by the AntConc Text Tool [7] quickly providing you with the possibility to select the part of the text, from which the concordance was extracted. The results demonstrated that there is a wide variation in high frequency words use and gave a possibility to extract the most frequent collocations for further development of tasks. The example of concordance lines for the word 'magnet' is provided in Fig.1. The most frequent two-word collocations provided by AntConc Collocation Tool are: *magnetic field, magnetic sensors, magnetic flux, magnetic induction, magnetic system, magnetic material.*

The next step was to evaluate the patterns of speech, which these words form. The concordances provided by AntConc Concordance tool demonstrate the chains of words on the left and right of the highlighted core word. So you can analyse and choose the patterns to be introduced in the assignments. If we look at the concordance lines in Fig.1, we can find some longer collocations, which are not so frequent for the discussed corpora, but can reflect some tendencies of word use within the area of magnetic devices description. For example, after going back to the texts with AntConc Text Tool and deeper study of the collocations with the frequent word *magnetic* (159 concordance lines), we were able to highlight the collocations, which can describe the semantic field of the *magnetic field behavior* and *magnetic field features*.

According to the investigation of longer collocations (3-5 words) provided by concordance lines magnetic field behavior can be described by the following patterns:

to cause disturbance to the magnetic field, magnetic field created by…, convert magnetic field into…, increasing magnetic field, magnetic field passing through…, increase in the magnetic field flux,

to all concordance lines retrieved (159 lines) magnetic field features can be demonstrated with the use of the following collocations: *external magnetic field, generated magnetic field, negative magnetic field, positive magnetic field, strong magnetic field, weak magnetic field, one- gauss magnetic field, sensed magnetic field, magnetic field flux*
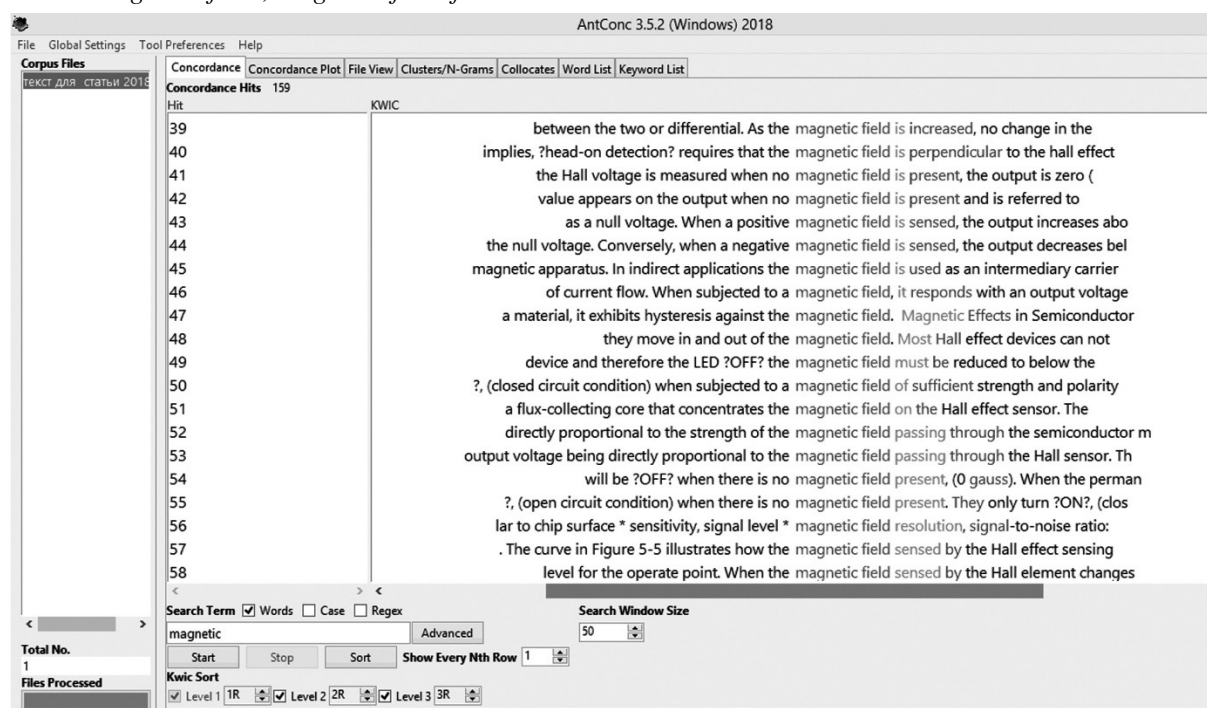

Fig. 1. **Example of concordance lines for term *magnetic* in EDD corpora** (Concordance Hits in the upper left-hand corner display the total number of concordance lines)

We also paid attention to the collocations with prepositions, which demonstrated high frequency, as it is shown in Table 2.

Table 2

Corpora preposition frequency

Preposition	Frequency	Preposition	Frequency
of	33.8	with	7.6
to	25.6	on	7.0
in	20.8	by	6.1
for	8.4	off	1.5

The most frequent patterns have been demonstrated for prepositional phrases with prepositions *of*, *to*, *in* and less frequent were collocations with the preposition *off*. If we speak about the use of

prepositions, we should pay attention to the fact that prepositions highlight the relation between the different elements in a sentence, so collocations with prepositions we can see in concordance lines for high frequency words are important and classroom practice demonstrates the frequent wrong use of prepositions in tests and in speech.

Some examples of special preposition use revealed in concordance lines are:

In the category of connection describing relationships: *characteristics of the current, recommendations on implementation of microchip technology, decreases with the magnetic field.*

In the category of origin: *the output signal from the amplifier, the electrical signal from the Hall sensor.*

In the category of instrument: *devices activated by an external magnetic field.*

In the category of place: *field density around the device.*

In the category of direction: *magnetic field is decreased to below operation point.*

In the category of reason/purpose: *to establish the reference level for the operating point.*

Preposition use in phrasal verbs they improve on the existing current sensing technology, *sensors that are switched on*, ear-splitting alarm to scare off intruders.

Set expressions: *on a large scale*, *in close proximity of, on a daily bases, switches. in the off condition.*

The above examples are only a small part of different speech patterns with prepositions revealed due to the exploration of concordance lines. These data can serve an important tool for the ESP classroom, as they provide with an enormous number of examples of preposition use in the special environment of occupational speech patterns and can be used in gap filling exercises production.

The results of concordance line study enabled us to create texts, in which students could practice to extract collocations with the highlighted terms and academic words, which we chose according to both their frequency and tendency to be used in patterns of speech together with high frequency occupational words. The first practice of texts of self-study revealed that 95 % of students were not able to retrieve word patterns longer than two-word collocations and did not pay attention to the preposition use. Still, as we can see from the examples of frequent words use provided above, longer speech patterns could be more meaningful in speech production. The ready-to-use word chains to be introduced in the ESP classroom can be as follows:

the strength of magnetic field; a circular magnetic field produced around; magnetic field perpendicular to…; converts an electric field into…

Students were asked to start vocabularies as a self study while reading texts within their occupational areas. These vocabularies included not only new words, but also new collocations of different length, so that they could use them in the classroom for speech production. Regular study of collocations based on word patterns retrieval into vocabulary as a self-study increased the students' ability to search for and extract collocations from occupational texts and facilitated word acquisition. Also special tasks based on frequent word patterns from concordance lines were developed including matching tasks and gap filling tasks. The example of a matching task targeting the development of skills of matching collocations in speech is presented in Table 3.

Table 3

Match the word patterns

Collocation 1	Collocation 2
A typical internally regulated	Magnetic field passing through
Directly proportional to the strength of	The growing plants
The ideal environment conditions for	Wireless signals from the sensor
Farmers can take advantage of	Hall effect sensor

The resulting patterns, which can be later incorporated in speaking tasks, should be:

A typical internally regulated Hall effect sensor

Directly proportional to the strength of magnetic field passing through

The ideal environment conditions for the growing plants

Farmers can take advantage of wireless signals

The approach based on exploration of frequent word concordances with further introduction of the retrieved collocations into the ESPU classroom exercises resulted in accumulation of wide variety of patterns by students and increased their awareness of the span of words in their occupational areas. Also, a new way of vocabulary acquisition was introduced through acquisition and use of collocations and long word patterns. The positive impact of the discussed above approach was demonstrated by high score for the dictation and test based on the collocations they introduced into their vocabularies (86 % of students got excellent and good marks for the tests). According to the survey the students stressed that while choosing collocations from texts they were able to remember them and searching for collocations with the computer engine "find in text" made them more motivated to study vocabulary than just translation of words and their memorising. What is more, the students concluded that searching for collocations in up-to-date ESP texts as self-study gave them extra information on the engineering subjects they study.

Conclusions

1. Small specialized corpora based on informative engineering texts can be used for high frequency words extraction and their further exploration in concordance lines, so as to get information on word patterns to be introduced in the classroom for the development of lexical communicative competence.
2. Concordance lines retrieved from occupational texts provide with wide variety of information including academic word and term frequency and their collocation patterns. The word collocations can be categorized by semantic fields and parts of speech.
3. Concordance-based ESP assignments expand the learners' vocabulary and increase their awareness of the great variety of academic and terminological word patterns within their occupational area.
4. Reading special texts both in ESP classroom and as self-study is relevant to the idea of interdisciplinary learning, while occupational corpora formation and corpora -based assignments stimulate the formation of expanded professional vocabulary.

References

[1] EHEA Ministerial Conference "Beyond the Bologna Process: Creating and connecting national, regional and global higher education area", 2012, Bucharest. [online] [3.03 2015]. Available at: http://www.ihea.info/upload/documents/BPF Statement_27042012
[2] Robinson P.J. A rich view of lexical competence. English Language Teaching Journal, vol. 43(4), 1989, pp. 274-282.
[3] 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. (In Russian).
[4] [bnc] British National Corpus-University of Oxford. [online]. available at: www.natcorp.ok.ac.uk/
[5] What is Corpus Linguistics. "Natural Language Processing". [online] [08.12.2009].Available at: languageworldofcomputingnet/linguistics/introduction/what-is-corpus.html
[6] Baker P. Querying Keywords, Questions of Difference, Frequency and Sense in Keywords Analysis. Journal of English Linguistics, vol. 32, 2004, pp. 346-359.
[7] Jurafsky. D., Martin J.H. Speech and Language Processing. Third Edition. Copyright 2017. [online] [28.08. 2017]. .Available at: https://web.stanford.edu/-jurfsky/slp3/
[8] Academic Words List. AWL [online] [15.01.2017] Available at: http://www.englishvocabularyexercises.com/AWL/
[9] Porter D. Check your vocabulary for academic English. Third edition. London: A&C Black Publishers LTD, 2007. 78 p.
[10] Ozola I., Vincela Z., Parsove V. Corpus Based Comparative Analysis of English and Latvian Terms of land Administration. Proceedings of International Conference "Rural Environment. Education. Personality", May 15-16, 2015, Jelgava, Latvia, pp. 429-434.

[11] Krieger L., Griffith.T. Direct Hits Core Vocabulary of the SAT. Third edition. Atlanta: Direct Hits Publishing, 2010. 140 p.

[12] Nation P., Anthony L. Measuring vocabulary size. E Hinkel Handbook of Research in Second Language Teaching and Learning, vol.3, New York, 2016. pp. 355-368.

[13] O'Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: language use and language teaching. First edition. Cambridge, New York: Cambridge University Press, 2007. 333 p.

[14] Tribble Ch., .Jones G. Concordances in the Classroom: a Resource Guide for Teachers. Second edition. Houston: Athelstan Publications, 1997, 114 p.

[15] Anthony I.A. Freeware Corpus Analysis Toolkit for Concordancing and Text Analysis. 2013. [online] [02.2018]. Available at: http://www.laurenceanthony.net/software