

COMPUTER-AIDED RESEARCH OF ESP CLASS MATERIALS: VOCABULARY POTENTIAL AND LEARNING OPPORTUNITIES

Ya. Rozanova, A. Kudryashova, O. Zamyatina

Abstract. The present study focuses on the assessment of vocabulary potential of teacher-created pedagogical materials to use in the class of English for Specific Purposes (ESP) taught to Russian undergraduates, majoring in Chemical Engineering, in a large public technical university. Specifically, the aim of the research is to examine the vocabulary distribution in the collection of course materials in terms of its range, frequency and variety. The motivation behind the study is that ESP instructors often have to rely on their own intuition during the text selection and material development due to the fact that there is the lack of pedagogical materials covering discipline-specific vocabulary of every specialization. ESP instructors teach English first, integrating the presentation of topics from subject matter classes and the major challenge consists in the lack of content-based knowledge of ESP instructors to select the domain vocabulary to L2 learners. Thus, to facilitate customized language acquisition (basic engineering, academic and specialized vocabulary) and to determine the extent to which the vocabulary contained in the texts is specialized and relevant to L2 learners' domain, corpus software "Range" is used to run the analysis. The results suggest that the ESP corpus promotes the acquisition of high frequency English words (the first and the second 1,000), basic engineering and academic vocabulary. The results also demonstrate insufficient level of teacher-created materials for specialized vocabulary development and highlight the need to include texts of varying vocabulary types, and to optimize specialized wordlist.

Keywords: computer-aided research; corpus software; text analysis; vocabulary acquisition; English for specific purposes.

Introduction

The high priority of English being the language of global communication, international scholarship and research of modern society reinforces the promotion of English language teaching due to the fact that a proficient command of EFL graduates offers them a significant opportunity to get a more prestigious job, rapid promotion and gain success in the professional field, as well as share professional and research findings by being involved in international scientific events. Against this background, Russian universities place a great emphasis on teaching EFL and English for Specific Purposes (ESP). The fact that the instructional models and the ways of learning English present an important issue nowadays is supported by numerous studies in Russia, and abroad. Specifically, teaching ESP to EFL learners has been addressed much in various scientific works. As the result, prevailing investigations focused on the development of pedagogical materials, applica-

tion of innovative methods and approaches that provide favorable learning opportunities for English language acquisition have been conducted.

Thus, the issue of qualitative teaching materials becomes rather urgent nowadays. With this regard, it is worth stating that despite a great abundance of English course materials, the published commercial textbooks are actually unclaimed in many ESP contexts nowadays. It is partly conditioned by the necessity to cover a wide range of rather specific engineering sub fields (such as Chemistry of polymers, biotechnology, chemistry of silicates, robotics, industrial design, etc.) and, consequently, develop teaching materials in the scope of specific knowledge fields. Thus, there occurs the situation when the field of ESP is witnessing a lack of authentic, discipline specific materials reflecting the context of the learners' future profession. Consequently, language instructors often rely on their own intuition during the text selection and adaptation, and have to develop pedagogical materials themselves in order to immerse learners in a professional environment by making their ESP acquisition course profession-oriented and meaningful for them.

Thus, the increased learning potential of the pedagogical materials motivate ESP instructors to use the corpus software tools for critical evaluation of how resourceful the material is before piloting it with a class. It is conditioned by the fact that being a non-native English speaking teacher (non-NEST) one tends to have linguistic insecurity and inappropriateness, and would use structures or speech patterns which native speakers would not use. To raise the non-NESTs awareness of the language taught, a corpus software with huge chunks of samples of written and spoken language embedded can be used. It contains a variety of samples (e.g. BNC – British National Corpus) which reflect the reality of today's English communication, so non-NESTs have an opportunity to proofread, compare, and evaluate whether or not the developed course materials provide favorable opportunities to expand students' knowledge of academic and specialized vocabulary.

Background

The corpus software has been used much to analyze the richness of a discipline-specific vocabulary in the corpus [1–4] for the reason that specialized terms are of paramount importance for providing fluent communication in English in a professional field of action. Based on numerous exploratory studies conducted in the field of vocabulary threshold [5–7] there has been determined a strong link between the learners' vocabulary size and their reading skills. A large vocabulary background facilitates reading comprehension whereas a limited vocabulary range often results in certain difficulties with the text comprehension. In other words, the more vocabulary is mastered beforehand and recognized while reading, the higher level of comprehension may be reached. Due to this fact, it might be assumed that reading enables to estimate how rich the learners' lexical threshold is.

As [8] claimed, knowledge of the 2,000 most frequent word families in English would make it possible for the reader to recognize 84% of the words in English texts. In [9] it is indicated that the vocabulary size of 3,000 word families would enable a reader to perform a reading task without any assistance with the use of their native language reading strategies and provide 95% of the text coverage, whereas the knowledge of 5,000 word families provides 98% of the lexical coverage (lexical coverage stands for the total percentage of the words which a reader is able to recognize in a text). Additionally, [6] claimed that while some readers required 95% of lexical coverage for adequate comprehension, it might seem more essential to tend to reach 98%, which guarantees unassisted pleasurable reading. However, according to [10] later studies, 95% of text coverage could be reached with the knowledge of 4,000–5,000 word families (including proper nouns), and on condition that a learner knows 8,000 word families (plus proper nouns), it would result in 98% of coverage.

Assuming that 95% of lexical coverage is attained by the learners with the vocabulary size of 4,000–5,000 word families, it is necessary to take into account that each word family has a headword and the number of derivatives with the same stem as well. For example, the high frequency word family *judge*, which is included in the first thousand of the most frequent words in spoken and written English, has a wide range of members (derivatives): *judged, judgement, judgemental, judgements, judges, judging, judgment, judgmental, judgments, misjudge, misjudged, misjudgement, misjudgements, misjudges, misjudging, prejudice, prejudged, prejudgement, prejudgements, prejudices, prejudging*. Considering the fact that every word family comprises a certain number of members, as it has been indicated in the example above, the total number of word families that learners in Russia should know becomes incredibly huge. We can assume that it might be hard for EFL learners to reach the vocabulary level of 4,000–5,000 words due to the fact that they have their ESP classes rather rarely, namely twice a week, which makes the situation even more complicated by the lack of linguistic environment and an additional complexity being a rather low English language proficiency level, which is according to the Common European Framework of Language predominately is majorly A2, B1 and to a lesser extent B2. On the other hand, we should take into account the fact that the learners have already been acquainted with the basic word-building rules after the years of English learning; so it might not be an arduous task for them to comprehend derived or inflected members of a word since the meaning of a headword is already known.

Nevertheless, the ESP teachers have to be selective in terms of vocabulary contained in the texts they prepare for the class. Research motivation is also conditioned by the compelling argument [11, 12] concerning the abundance of the vocabulary errors occurring in EFL learners' speech, which, in

most cases, is conditioned by the lack of vocabulary and, as a consequence, inappropriate choice of words. Therefore, the development of an optimized vocabulary word list adequate to learner's possibilities is the task of a paramount importance for ESP instructors.

According to research findings reported by [1] it was found that the most frequent words of engineering corpus, which contained nearly 2,000,000 running words, were sub-technical and non-technical words from the academic register. Moreover, according to Yang's statistical analysis [13], sub-technical or non-technical words have a very high distribution in all specialized disciplines. The abovementioned claim has driven us to the assumption that a due attention should be given to learning sub-technical and academic vocabulary along with the specific, technical vocabulary.

In the current study it is considered that the corpus software Range helps infer whether the currently used pedagogical materials provide favorable opportunities for L2 learners to acquire academic and specialized vocabulary or not. Also, corpus software applied allows examining the vocabulary size required for L2 learners to comprehend texts in English and exploring the vocabulary representativeness from the viewpoint of learner's specialization.

More specific research questions are:

1. To examine the vocabulary size the learners need to have to comprehend the texts contained in the corpus (Chemical Engineering for Polymer Studies) created by ESP teachers with use of Range (corpus software).

2. To explore the distribution of words in the corpus according to the most frequent words in English – 1K, 2K and AWL (academic word list), using GWL / or GSL (general word list / or general service list); and to determine the size of basic engineering words in the corpus using Basic Engineering List (BEL) by Ward [14].

3. To evaluate the extent to which specialized vocabulary developed by ESP instructors for chemical engineering students agrees with the widely used terms in the field of chemical engineering and chemistry of polymers.

Methodology

1. *The corpus.* The corpus was compiled basing on the course materials used in the class of Chemical Engineering. It consists of a four-skill integrated textbook, tests (progress and achievement tests, self-study activities), additional instructional materials (handouts comprising listening, grammar and vocabulary tasks, speaking and writing activities), lab materials as shown in Table 1. The total number of tokens is 141,000; however, only one part of the corpus, namely Chemical Engineering for Polymer Studies, which contains 37,959 tokens, was selected for being analyzed in the present paper. As [1] reported 'for language learning and teaching, smaller corpora can be more useful as they are designed to represent the specific part of the lan-

guage under investigation and are tailored to address the aspects of the language relevant to the needs of the learner’.

Table 1

Chemical Engineering for Polymer Studies Corpus Details

Sub-corpus details	Number of words	Comments	Examples
Tests	6,186	Listening, Reading, Grammar, Vocabulary; 3 variants	Polypeptide chains are sequenced and coiled in such a way that the hydrophobic amino acids usually face inward, giving the molecule stability, and the hydrophilic amino acids face outward, where they are free to interact with other compounds and especially other proteins
Textbook	20,971	4 units, 4 skills + grammar & vocabulary; adapted materials, teacher-created activities	Polyesters are condensation polymers, which contain fewer atoms within the polymer repeat unit than the reactants because of the formation of byproducts, such as H ₂ O or NH ₃ , during the polymerization reaction. Most synthetic fibers are condensation polymers
Supplementary materials	9,025	Grammar and Vocabulary exercises	Addition polymerization involves the linking together of molecules incorporating double or triple chemical bonds. These unsaturated monomers (the identical molecules which make up the polymers) have extra internal bonds which are able to break and link up with other monomers to form the repeating chain
Lab work materials	1,777	Testing materials	Plastic are organic substances. They are made synthetically by polymerization, and capable of being formed into an almost endless variety of products, e.g. threads, sheets, tubes and moulded objects. The ancestor of modern synthetic plastics is celluloid
Total:	37,959		

2. *Instrument and wordlists.* Range (2002) by Nation and Heatley was used to run the analysis in order to answer the research question 1 and to explore the representativeness of the vocabulary. It contains BNC (British National Corpus) and COCA (Corpus of Contemporary English) lists. These vocabulary lists consist of 25 word family frequency lists and additional lists including proper names, abbreviations, marginal words and transparent compounds. Also, it has the lists that contain one nonsense word each to provide space for additional lists [15]. The words which have not been found in the aforementioned lists are classified by Range as ‘not in the list’.

To answer the research question 2, the corpus under discussion is compared with General word list (GWL) / General Service List by West, which comprises 1K, 2K, AWL incorporated in Range. The words in the corpus are classified according to the category they belong to; they are divided into the first thousand of the most frequently used words in English, followed by the second thousand of the most frequent English words (1K and 2K), academic word list (AWL) and the 'off-list' words which were not found in the lists above. In fact, GWL / GSL containing thoroughly-made frequency lists for English learning presents the core of the general English language. In spite of the fact that it was established a long time ago GWL has already been updated several times and is still kept on updating regularly. As for AWL [16] it contains 570 word families widely used in academic genre which L2 learners actually encounter in many ESP contexts. To find the extent to which the corpus under investigation agrees with the engineering field, the 299-word list (BEL – Basic Engineering List) by Ward is also applied. BEL comprises the vocabulary frequent in a wider range of engineering sub-disciplines and it does not contain any functional words. 'BEL is short and non-technical in nature but gives excellent coverage of a wide variety of engineering textbook material; by concentrating on word types rather than lemmas or families, it encourages learning not only of individual words but also of their lexico-grammatical environments' [14].

As regards to the evaluation of the extent to which the vocabulary in Chemical Engineering for Polymer Studies corpus may be considered as professionally-oriented (research question 3), the chemical engineering vocabulary wordlist and the word list of polymer chemistry terms have been used. The first one is the wordlist of basic chemical terms used in a wide range of chemistry fields; this keyword catalogue has been developed by Wiley-VCH that is published for the scientific community, scientific societies and researchers, practitioners, teachers and students all over the world. It comprises about 760 terms widely used in the sphere of chemistry. The second wordlist, developed by the International Union of Pure and Applied Chemistry (IUPAC), comprises 850 basic terms in polymer science and demonstrates the list of the most representative key-words that cover all aspects of polymer science. Use of the abovementioned lists allow estimating potential and validity of the vocabulary in teacher-created materials used in ESP class to learn how specialized the lexis is. In addition, these two wordlists are used to examine the vocabulary potential of other materials which have been created for ESP classes by native English speakers (Handbook of Chemical Technology and Pollution Control, by Martin B. Hocking and the collection of scientific articles from the 'Journal of Polymer Research' issued by Springer Publishing House one of the leading Asian journals on Polymer research). It provides an opportunity to trace and compare common features and differences in the content of spe-

cialized vocabulary both in the materials under research and in authentic materials as well.

Results and Discussion

Table 2 addresses the research question 1 and demonstrates the figures of vocabulary size needed for L2 learners to deal with the texts in the corpus under research.

Table 2

Distribution of Chemical Engineering for Polymer Studies corpus across BNC / COCA

Basic wordlists BNC / COCA	Chemical Engineering: Chemistry of Polymers		
	Tokens	Coverage, %	Cumulative, %
1 st – 1,000	4,181	29.64	29.64
2 nd – 1,000	2,386	16.91	46.55
3 rd – 1,000	1,686	11.95	58.50
4 th – 1,000	443	3.14	61.64
5 th – 1,000	545	3.86	65.50
6 th – 1,000	612	4.34	69.84
7 th – 1,000	170	1.21	71.05
8 th – 34 th	110.....81	0.78.....0.57	82.00

Note. The counts only include content words (function words and proper nouns were excluded). Therefore, cumulative percentage does not equal 100 per cent.

As can be inferred from the table, the first 4,000–5,000 word families (in BNC / COCA) provide only 61.64–65.50% of text coverage in Chemical Engineering for Polymer Studies corpus instead of 95% mentioned by [10]. Based on the assumption that the knowledge of 8,000 word families provides 98% of text coverage and unassisted reading for pleasure, data represented in Table 2 indicate that the vocabulary size ranging from 8,000 to 34,000 in BNC and COCA yielded only 82% of coverage. The cumulative coverage 82% provided by the 34th thousand indicates that a learner is certain to have difficulties with text reading and comprehension due to vocabulary size needed to comprehend texts in the corpus. The data obtained require further research in terms of vocabulary examination to exclude not discipline specific words.

Table 3 demonstrates the distribution of 1K, 2K, AWL, and BEL in the corpus under investigation and answers the research question 2.

The results of the first analysis in Table 3 have revealed that the percentage of the basic engineering vocabulary in the corpus is insufficient 13.51% that might result in difficulties with comprehension of Chemical Engineering for Polymer Studies texts and require persistent dictionary use.

Table 3

Distribution of Chemical Engineering for Polymer Studies corpus across GSL / AWL

Basic wordlists	Chemical Engineering: Chemistry of Polymers		
Analysis 1 (BEL)	Tokens	Coverage, %	Cumulative, %
On-list	1,923	13.51	13.51
Off-list	12,312	86.49	100.00
Analysis 2 (GSL/AWL)	Tokens	Coverage, %	Cumulative, %
1 K	5,242	37.13	37.13
2 K	1,091	7.73	44.86
AWL	1,524	10.80	55.66
Off-list	6,146	43.54	99.20

To address the research question 2, it is necessary to refer to the analysis 2 (Table 3), which has shown that the words from 1K, 2K and AWL account for 55.7%. As regards to the off-list it is rather huge and indicates that there is an overwhelming number of infrequent and specialized lexis (such as *epoxy*, *acelobutyrate*, *carboxylic*, *monosaccharides*). On the one hand, the vocabulary in the corpus is diverse and representative from the perspective of its variety (general English vocabulary, academic vocabulary, specialized vocabulary). On the other hand, the quantity of specialized words seems to be rather excessive. Later off-list rigorous examinations have revealed that this off-list contains the misspelled words, proper names, and Russian words which were classified by Range as unknown ones. Thus, it is particularly important to optimize the off-list by grouping the words according to the category they belong to (specialized, misspelled, proper names, Russian words) in order to realize what it comprises indeed. The outcomes are presented in Table 4.

Off-list examination and optimization

Table 4

Specialized vocabulary	Tokens	Coverage, %
Off-list before examination and optimization	6,146	43.54
Off-list after examination and optimization	1,488	10.62

After breaking the 'off-list' into the categories and eliminating the misspelled words, proper names and Russian words it has reduced to 1,488 words (10.62%), which is acceptable relatively to general English vocabulary, academic and basic engineering lexis.

To address the research question 3, we should refer to Table 5.

Table 5 demonstrates how specialized and discipline-specific the vocabulary is in the three corpora - Chemical Engineering for Polymer Studies, Handbook of Chemical Technology and Pollution Control, and the collection of scientific articles from the 'Journal of Polymer Research'. As can be in-

ferred from the table, the first analysis has revealed that specialized vocabulary contained in Chemical Engineering for Polymer Studies corpus is 10.70%, which, in its turn, is correlated with Handbook of Chemical Technology and Pollution Control (12.56%) and 'Journal of Polymer Research' (11.3%). It indicates that the learners dealing with these corpora have adequate exposure to discipline-specific words.

Table 5

Polymer and chemical engineering terms in teacher-created and authentic materials

Analysis 1 Polymer Wordlist in the Corpora				
Corpora	Tokens number	Tokens number in percentage	Polymer Wordlist on-list, %	Polymer Wordlist off-list, %
Journal of Polymer Research published by Springer	79,779	100	9,041 (11.33)	70,738 (88.67)
Handbook of Chemical Technology and Pollution Control	28,286	100	3,554 (12.56)	24,732 (87.44)
Chemical Engineering for Polymer Studies	28,851	100	3,087 (10.70)	25,764 (89.30)
Analysis 2 Chemical Engineering Wordlist in the Corpora				
Corpora	Tokens number	Tokens number in percentage	Chemical Engineering Wordlist on-list, (%)	Chemical Engineering Wordlist off-list, (%)
Journal of Polymer Research published by Springer	79,779	100	2,764 (3.46)	77,017 (96.54)
Handbook of Chemical Technology and Pollution Control	28,286	100	1,263 (4.46)	27,023 (95.53)
Chemical Engineering for Polymer Studies	28,851	100	1,055 (3.65)	27,796 (96.34)

Note. Total number of tokens in Chemical Engineering for Polymer Studies corpus differs in Table 5 and Table 1 due to the exclusion of proper names, hyphen words and Russian words.

The second analysis has shown that the amount of words which are generally used in the sphere of chemical engineering is insufficient in Chemical Engineering for Polymer Studies corpus and equals 3.65%. Incidentally, the counts of ESP teacher-created materials and authentic materials are comparable – the Handbook of Chemical Technology and Pollution Control (4.46%) and the Journal of Polymer Research (3.46%).

Conclusion

The following conclusions were made in the course of investigation: the texts in Chemical Engineering for Polymer Studies corpus provide adequate exposure to frequently used words in everyday English, however, academic and basic engineering vocabulary is lacking and needs to be revised. The corpus was found to contain polymer science specific terms, and it is also worth noting here that the major part of the specialized vocabulary in the corpus comprises the terms of Latin origin having similar pronunciation in the learners' native language and can be easily recognized by the learners (*e.g. alkene, aminoacetic, anhydroglucose, asbestos, ascorbic, carboanions, interpolymer, keratin, melamine, phenylalanine, phenylethene, and etc.*). In this regard, we can state that the corpus offers reasonable exposure to specialized vocabulary in the field of polymer science and allows the learners to easily recognize the meaning of unknown term due to their Latin origin. The significant outcome of the study, which may be used as pedagogical implication for further work, is to proofread and amend the texts in the corpus in terms of the word spelling inasmuch as the off-list examination revealed the presence of misspelled words, which results in incredibly high figures on the learners' vocabulary size needed to comprehend the texts.

On the whole, we came to the conclusion that computer aided research, which was conducted basing on corpus software tool (Range), helped ESP teachers to critically evaluate the material they introduced in the class. As the analysis was done after piloting class materials with students, the important issue which needs to be addressed is that such expertise needs to be carried out before the classes begin. Moreover, such computer programs are excellent instruments for non-NEST teachers working in non-language environment to review and to create informative, comprehensible course materials relevant to learners' professional interests and specialization.

Acknowledgments

The research is carried out at Tomsk Polytechnic University within the framework of Tomsk Polytechnic University Competitiveness Enhancement Program grant.

References

1. Mudraya O. (2006) Engineering English: A lexical frequency instructional model. *English for Specific Purposes*. 25. pp. 235–256.
2. Matsuoka W., Hirsh D. (2010) Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language* 22/1. pp. 56–70.
3. Hsu W. (2011) The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes* 30. pp. 247–257.
4. Friginal E. (2013) Developing research report writing skills using corpora. *English for Specific Purposes* 32. pp. 208–220.

5. Laufer B. (1991) How much lexis is necessary for reading comprehension? Arnaud P. and Bejoint H. (Eds.): Vocabulary and applied linguistics. Macmillan. pp. 126-132.
6. Hu H., Nation P. (2000) Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13. pp. 403-430.
7. Qian D. (2002) Investigating the relationship between vocabulary knowledge and academic reading performance. An assessment perspective. *Language Learning* 52. pp. 513-536.
8. Hwang K., Nation P. (1995) Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23. pp. 35-41.
9. Laufer B. (1997) The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. Coady J. and Huckin T. (Eds.): *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge University Press. pp. 20-34.
10. Laufer B., Ravenhorst-Kalovski G. (2010) Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22. pp. 15-30.
11. McKay S. (1980) Teaching the syntactic, semantic and pragmatic dimensions of verbs. *TESOL Quarterly* 14. pp. 17-26.
12. Lewis M. (1993) *The lexical approach*. Hove, England: Language Teaching Publications.
13. Yang H. (1986) A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing* 1/2. pp. 93-103.
14. Ward J. (2009) A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes* 28. pp. 170-182.
15. Nation P., Webb S. (2011) *Researching and Analyzing Vocabulary*. Heinle Cengage Learning.
16. Coxhead A. (2000) A new academic word list, *TESOL Quarterly* 34/2. pp. 213-238.

Information about the authors:

Rozanova Ya., senior lecturer, Division for Foreign Languages, School of Core Engineering Education, National Research Tomsk Polytechnic University. E-mail: ioannastar@list.ru

Kudryashova A., senior lecturer, Division for Foreign Languages, School of Core Engineering Education, National Research Tomsk Polytechnic University. E-mail: english@tpu.ru

Zamyatina O., Ph.D., Associate Professor, Rector, Tomsk Regional Teachers Professional Re-training Institute. E-mail: zamyatina@tpu.ru

Received 11 May 2018