Global Business Languages

Volume 20 Article 9

Vocabulary demands for engineering students studying English in Russia: Comparing ESP course materials across three engineering disciplines

Tatiana Nekrasova-Beker
Colorado State University

Recommended Citation

Nekrasova-Beker, T. (2020). Vocabulary demands for engineering students studying English in Russia: Comparing ESP course materials across three engineering disciplines. *Global Business Languages*, 20, 159-171.
Available at (DOI): https://doi.org/10.4079/gbl.v20.9

Tatiana Nekrasova-Beker
Colorado State University

VOCABULARY DEMANDS FOR ENGINEERING STUDENTS STUDYING ENGLISH IN RUSSIA: COMPARING ESP COURSE MATERIALS ACROSS THREE ENGINEERING DISCIPLINES

ABSTRACT
With English language instruction becoming increasingly more specialized in higher education institutions around the globe, English for Specific Purposes (ESP) practitioners are facing a unique challenge in developing language courses that require considerable knowledge of a discipline in order to make it applicable to students and to meet their specific language needs. In the case of Engineering, which is the target discipline in this study, substantial empirical research has been published describing general discipline-specific requirements as well as common challenges that second language (L2) students face in various pedagogical contexts (e.g., Kaewpet, 2009; Pritchard & Nasr, 2004; Rowley-Jolivet, 2015; Rozycki & Johnson, 2013). Yet research investigating the vocabulary demands of pedagogical materials utilized in various sub-fields within the same discipline is limited. Therefore, the present study examined the extent to which the vocabulary demands of the pedagogical materials employed in ESP courses in Thermal-Power, Computer, and Chemical Engineering in Russia were comparable across the courses and achievable for the students. The results indicated that vocabulary coverage varied considerably across the three disciplines, with Chemical Engineering texts requiring the largest vocabulary size for adequate comprehension. The implications of the study for materials development and teaching ESP courses in various Engineering sub-fields are discussed.

KEYWORDS: corpus-based, engineering discourse, English for Specific Purposes, pedagogical materials, vocabulary load

INTRODUCTION

One of the four main goals that the Association for Engineering Education of Russia (AEER) specified in their recent goal statement is to promote international connections in engineering education and to help integrate Russian scholars and engineers into international research initiatives. This is not a small task to undertake. In addition to the discipline-specific knowledge and expertise necessary for such levels of involvement, it is also the ability to effectively communicate this knowledge and expertise in a different language that is often more challenging to acquire. Therefore, the AEER argues for the necessity of “targeting the language training to a fluent use of foreign language as a tool of international communication . . . to improve the quality of engineering education” (Association, 2018).

When it comes to providing language instruction to university-level students majoring in Engineering, a current view shared by many language practitioners is that such courses should be geared towards specialized content that is relevant to students’ area of study rather than offering general-purpose language courses that target broad topics which can be accessible to students across different disciplines (Hsu, 2014; Kanoksilapatham, 2015; Ward, 1999, 2009a). While many polytechnic institutions in Russia (as well as globally) have answered the call for offering specialized language instruction to their students, this initiative, while being timely and forwardlooking, has not been an easy process. One of the main challenges encountered at the institution targeted in the present study was related to materials development for LSP courses. With many engineering disciplines taught at the university (from chemistry and applied biomedical sciences to computer science and robotics) as well as the fact that commercially available textbooks met the content requirements and students' general proficiency level only for a few proposed LSP courses, language instructors were tasked with developing packets of course materials that included authentic discipline-specific texts and supplementary activities. The present study was motivated by the most immediate question about whether the outcome of this materials development effort was successful. Specifically, the study explored the level of vocabulary difficulty of the pedagogical materials used in three different English for Specific Purposes (ESP) courses by carrying out a series of corpus-based analyses of these texts (a collection of texts in each discipline is referred to as corpus). A corpus-based analysis enables a researcher to identify patterns in language data based on their frequency of occurrence. A variety of language structures can be targeted in such analyses, including exploring the most common collocations used in texts, identifying a list of words that are "key" to a given genre or author, or investigating the vocabulary load of texts, to list a few examples. In order to contextualize the present study, the following section offers a brief discussion of previous research that examined various features of engineering discourse, focusing specifically on corpus-based analyses of engineering texts.

ENGINEERING DISCOURSE: EXPLORATIONS OF DISCIPLINE-SPECIFIC FEATURES

There is a substantial body of empirical research that has explored various language-related skills, abilities, and practices that are necessary for future engineers and, therefore, should be considered in an engineering curriculum. Researchers have examined a plethora of topics, including the development of receptive and productive skills in engineering students (e.g., Kaewpet, 2009; Pritchard & Nasr, 2004), discipline-specific genre requirements in engineering and challenges associated with teaching them (e.g., Archer, 2008; Wolfe, 2009), as well as analyses of the specific communicative needs of the engineering industry and the extent to which they are addressed during training (Spence & Liu, 2013). Within this body of research focusing on language use in engineering, corpus-based investigations of engineering discourse is an area of study that has gained a lot of interest from researchers and language practitioners alike, especially within the past ten years, due to a unique methodology that allows one to reveal usage patterns that, otherwise, would not be discovered. For LSP practitioners—those who are involved in designing and teaching LSP courses—familiarity with corpus-based methodology and its applications in language classrooms not only informs the development process, but also offers tools to engage students in data-driven language learning which is meaningful, content- and level-appropriate, and promotes student autonomy (Bárcena, Read, & Arús, 2014).

With respect to Engineering, a number of previous corpus-based analyses have focused on investigating the use of specific language structures in different engineering text types, including multiword constructions (Chen, 2010; Hyland, 2008; Ward, 2007; Wood & Appel, 2014), language structures that express certain concepts or functions (e.g., expressions of quantification in Rowley-Jolivet, 2015, or the lexis of construction in Orna-Montesinos, 2008), as well as the patterns of textual organization of various genres in engineering (e.g., Kanoksilapatham, 2015). The goal of these analyses is to provide detailed descriptions of the language patterns of engineering discourse in order to demonstrate its specificity compared to the typical patterns identified in general academic discourse. More recent investigations of this kind have also compared different sub-disciplines within the engineering field in order to examine the extent to which individual areas of study have impacted the discourse practices established in the discipline. For example, Kanoksilapatham (2015) investigated whether research articles written in three engineering sub-disciplines—civil, software, and biomedical engineering—shared the same textual organization. The results of the study indicated that, while all three sub-disciplines shared a set of common discourse practices, each sub-discipline also exemplified unique organizational patterns to construct the individual sections of a research article. One of the important implications of this study is that it highlighted the diversity of the engineering field (not only in terms of the content topics, but also with respect to the multitude of patterns of social interaction and the various mechanisms that exist to transmit discipline-specific knowledge) and justified why it is important for ESP practitioners to recognize the specificity of individual engineering disciplines in teaching and curriculum development. The argument put forward by Kanoksilapatham (2015) and other researchers (e.g., Hsu, 2014; Ward, 2009a; 2009b) provided additional justification for the questions explored in the present study by comparing the vocabulary load (as discussed in more detail below) in three different engineering corpora.

In addition to corpus-based research examining specific language structures and discourse patterns utilized in engineering texts, a number of studies have been conducted on engineering lexis, including studies which have investigated the vocabulary load of various types of engineering texts and pursued a specific goal of producing word lists that can be targeted during pedagogical instruction. This research, which is reviewed below, has greatly informed the questions asked in the present study as well as the methodology employed to answer these questions.

As noted above, the research exploring the vocabulary load of engineering texts is very practice-oriented. It is an often-cited challenge for students majoring in engineering, both English L1 and L2 learners, to be able to read and successfully comprehend discipline-specific texts (e.g., Anderson, 2015; Evans & Green, 2007; Pritchard & Nasr, 2004). When it comes to English language learners, their deficiencies in L2 reading skills are even more evident and might prevent them from being able to succeed at the university as well as to “effectively transition into the workforce” (Hartshorn, Evans, Egbert, & Johnson, 2017, p. 37). While the majority of content instructors expect that students are able to use discipline-specific course materials as a resource, an empirical investigation of first-year students’ (including ESL learners’) university experiences majoring in different disciplines (including engineering) suggests that students across disciplines encounter a number of reading challenges, from an inability to read discipline-specific texts and visual information to not understanding key concepts (Anderson, 2015). Therefore, it is not surprising that many ESP researchers (many of whom have an extensive experience teaching ESP formally) focus on issues of vocabulary coverage in various types of engineering texts.

Corpus-based research on vocabulary coverage was pioneered by Nation and colleagues who argued that a reader will be able to achieve successful comprehension of a text only if they know a certain percentage of running words in it, also referred to as the lexical coverage of a text (Liu & Nation, 1985; Nation, 2001; 2006). Two figures of 95% and 98% are typically cited in this type of research, with the former representing the lower threshold at which readers are likely to gain minimal comprehension of a text and to guess words from the context and the latter representing the upper threshold required for optimal (also referred to as ideal and unassisted) comprehension. While different researchers argue for using one figure over the other, with some evidence provided for an even lower vocabulary threshold required for minimum comprehension for some students (see Hu & Nation, 2000), it seems logical, for the purposes of the present study, to focus on 95% coverage as a more feasible benchmark, especially in the pedagogical context where English is taught as a foreign (as opposed to second) language (for a similar argument, also see Hsu, 2014; Ward, 1999; 2009a).

Nation and his colleagues have investigated lexical coverage of texts in a variety of genres, including newspapers (Hwang & Nation, 1989), short novels (Hirsh & Nation, 1992), a general English textbook (Matsuoka & Hirsh, 2010), and a variety of texts including novels, newspaper texts, a graded reader and a movie (Nation, 2006). The figures varied across different genres, with about 2,500 word families needed to achieve 95% lexical coverage of an English language textbook to about 4,000 word families plus proper nouns and marginal words needed to read English fiction and newspapers (95%). When it comes to discipline-specific texts, especially texts containing large quantities of technical vocabulary (i.e., words that require specialist knowledge to be understood), we can rightfully expect that these numbers increase (but see Ward's (1999) earlier claim about only 2,000 most common engineering word families needed to provide 95% lexical coverage of basic engineering texts).

A more recent study that provided a very comprehensive treatment of the topic of lexical coverage in engineering texts was conducted by Hsu (2014) who analyzed a corpus of commercially available textbooks which covered 20 different subject areas within engineering. The results indicated that while knowledge of the most frequent 5,000 word families (plus proper nouns, apparent compounds and abbreviations) should be sufficient to be able to comprehend an engineering textbook at a minimally acceptable level (95%), the vocabulary thresholds varied greatly across the engineering disciplines. Specifically, the first 3,500 most frequent word families was sufficient to comprehend texts in civil and mechanical engineering, while marine and biochemical engineering texts were the most lexically demanding and required the knowledge of 8,500 word families. The findings of the study illustrated a substantial discrepancy in lexical demands across different engineering disciplines, providing additional evidence of linguistic diversity within the engineering domain and the need to target each content area separately in ESP courses.

A natural outcome of the research conducted on lexical coverage of discipline-specific texts is the development of discipline-specific word lists that students need to know in addition to the first (K1) and second (K2) most frequent 1,000 word families in English—also known as the General Service List of English Words (GSL, West, 1953)—in order to be able to comprehend texts in their respective disciplines. There are multiple word lists that have been developed in the field, including the Academic Word List (AWL, Coxhead, 2000) which is interdisciplinary in nature and includes the 570 most frequent academic word families occurring in general academic texts. In terms of discipline-specific word lists, there exist the Medical Academic Word List (MAWL) (Wang, Liang, & Ge, 2008), the Business Word Lists (BWL for postgraduates, Hsu, 2011 and BWL for undergraduates, Konstantakis, 2007), and several word lists for engineering students, including the Student Engineering Word List (Mudraya, 2006), a Basic Engineering List (BEL) (Ward, 2009b), and an Engineering English Word List (EEWL) (Hsu, 2014). While each engineering word list mentioned here has its strengths (e.g., providing comprehensive coverage, being tested with larger engineering corpora) as well as weaknesses (e.g., including too many items or being rather narrowly focused on a specific collection of texts), Ward's (2009b) BEL was used in the present study due to its compactness (BEL contains

299 English word types) and its applicability to many engineering disciplines. Furthermore, BEL was extracted from a corpus of 25 engineering textbooks which were commonly used in the third and fourth years of undergraduate studies in engineering. It is for these reasons that BEL was utilized in the present study as one of the baseword lists to explore the extent to which the three engineering corpora provided the general coverage of the engineering field.

THE PRESENT STUDY

PEDAGOGICAL CONTEXT

The current study examined pedagogical materials used in three ESP courses taught at a large public polytechnic university in Russia which is nationally classified as a university with a high concentration of research activity by the National Accreditation Agency of the Russian Federation. The university provides training in a number of engineering disciplines supplemented with a strong emphasis on foreign language training. During the first two years of L2 instruction, students acquire general-purpose language competence, followed by an additional two years of discipline-specific language training. While the university has been experimenting with various formats to deliver discipline-specific instruction (from various content-based models to collaborative teaching with content instructors to requiring content instructors to incorporate discipline-specific language instruction in their courses), the present study was conducted in a context when all ESP courses were offered through a language unit and followed the same set of standards specifying the language skills targeted during the program of study (e.g., preparing an oral presentation on a discipline-specific topic, reading and annotating an article from a professional journal). The courses were taught by language instructors who worked closely with content specialists to ensure that the content topics targeted in their ESP courses were appropriate for the students and relevant for their program of study. Furthermore, content instructors provided additional assistance to language specialists regarding discipline-specific concepts and terms, as requested. Therefore, the degree of involvement of content specialists in course planning and materials selection varied across different engineering disciplines.

The materials examined in the present study were employed during the second semester of the four-semester ESP courses in Thermal-Power (TPE), Computer (CE), and Chemical Engineering (ChE). According to the general course description, the major goal of ESP instruction was to develop students' speaking and writing skills to be able to communicate ideas and concepts related to their discipline in English. Students were assumed to be at the intermediate level in terms of their English proficiency (institutionally defined in terms of number of semesters studying English). During the semester, students spent about three hours of English instruction in a classroom (for a total of 52 hours of instruction) and were also expected to spend an additional three hours to complete homework assignments. In terms of the content coverage, students in all three ESP courses studied four modules targeting four different discipline-specific topics. At the end of the semester, students took a final exam which included three sections: writing an annotation in English for a Russian text of about 400 words in length, preparing a literature review of about 2000 words in length on a topic related to the content covered during the semester, and preparing a three-to-four-minute presentation on the topic explored in the course.

RESEARCH GOAL

Considering the unique aspects of the course design and material selection and adaptation processes, the goal of the present study was three-fold. First, the study examined the extent to which pedagogical materials employed in three different engineering courses were comparable in terms of providing access to general academic, basic engineering, as well as more specialized vocabulary. It should be clarified that the materials used in the courses, including course textbooks, were developed by the ESP instructors who were guided by the content specialists. Specifically, the textbooks (as well as a large portion of additional classroom materials) represented a collection of texts and supplemental activities gathered from various English-based resources, such as websites, textbooks, professional journals, and popular science magazines. Therefore, the most immediate research goal in this specific pedagogical context was to gather additional evidence that would help evaluate the effectiveness of teaching materials and the courses in general. Related to this issue is the question of how unique (i.e., discipline-specific) the vocabulary included in the three courses is and whether (and to what extent) the materials employed in the three disciplines provided the general coverage of the field in addition to highlighting the discipline-specific coverage. Finally, the study also explored the vocabulary size needed to achieve 95% comprehension of the texts in each of the three corpora in order to evaluate the difficulty level of texts included in the corpora and to gauge whether those difficulty levels were comparable and realistic for the students enrolled in the three courses. The three research questions asked in the study were:

1) What are the vocabulary demands of pedagogical materials included in the three corpora?
2) What is the coverage of basic engineering words (BEL) in the three corpora?
3) What is the lexical threshold necessary for reasonable comprehension (i.e., to provide 95% coverage) of discipline-specific texts included in the three corpora?

To answer the three research questions, the study utilized a corpus-based approach to investigate the vocabulary demands in the three target disciplines. While the study goal is related to a specific pedagogical context and its unique challenges, the study can be of interest to a larger professional audience in two respects. First, the study illustrates how a corpus-based methodology can be employed to evaluate pedagogical materials and provides a detailed discussion of this approach which can be further adopted to other pedagogical contexts. Second, although this study compares only three engineering disciplines and is limited to materials employed in these particular ESP courses, the study highlights the substantial differences in vocabulary profiles for the three engineering disciplines and discusses the more general implications of this finding for L2 pedagogy. The next section provides a detailed description of the methodology employed in the study.

METHOD

ENGINEERING CORPORA

Drawing on the methodology established in previous corpus-based studies targeting discipline-specific texts (e.g., Hsu, 2011; Matsuoka & Hirsh, 2010; Todd, 2017), the present study examined the overall level of vocabulary included in the three corpora (typically reported as percentages of words that belong to more and less frequent classes, such as general vocabulary, academic vocabulary, and technical or discipline-specific vocabulary), as well as the level of vocabulary load required to comprehend the majority (i.e., 95%) of texts included in these corpora. Three corpora of pedagogical materials were developed to carry out the analyses. Each corpus included all pedagogical materials written in English to which students were exposed during the semester, including textbooks, laboratory materials, tests, and other materials which were made available to the students (e.g., handouts, classroom readings and activities).

Once the materials were identified, a standard procedure was employed to prepare each corpus. First, all materials were converted into an electronic format. Second, each file was further examined and textual features that could potentially misrepresent the vocabulary distribution (e.g., headers, references) as well as features that could interfere with the analysis (e.g., graphics, tables, formulas) were removed from the file. Additionally, textual anomalies (e.g., text breaks, text repeats, hyphenated compounds) were reformatted and a “clean” version of each document was saved as a plain text file. Third, each individual type of materials (e.g., textbooks, classroom materials) for a specific ESP course was saved as a separate text file. Finally, a complete version of all pedagogical materials for an ESP course was saved as a separate file and represented a corpus of pedagogical materials for the discipline. The breakdown of the three corpora employed in the present study is presented in Table 1.

Table 1. *Description of Three Engineering Corpora*

Sections	No. of words in TPE	No. of words in CE	No. of words in ChE
Tests	2,299	6,186	2,202
Textbook chapters	23,745	20,971	35,215
Classroom materials	20,625	9,025	2,416
Lab materials	879	1,777	1,513
TOTAL	47,548	37,959	41,346

Note: As illustrated in Table 1, while all four sections of materials were represented in each corpus, there were substantial differences in the number of words included in each section among the three ESP courses. Overall, the textbook chapters constituted the largest portion of each corpus, followed by classroom materials (a section that included additional handouts, reading materials, and activity worksheets), albeit the size of this section was very comparable to the textbook chapters in TPE.

ANALYSES

A free corpus-based program, Range (Heatley, Nation, & Coxhead, 2002), was used to explore the types of vocabulary that students encountered when working with pedagogical materials used in the three ESP courses. This program scans the texts and determines the level of vocabulary by assigning each item into one of the frequency bands, from more frequent (i.e., the first (K1) and second (K2) thousands of most frequent words in English) to Academic Word List (AWL) (Coxhead, 2000), to less frequent items, such as off-list words or K3–34. In addition to scanning the texts against existing word lists (i.e., K1, K2, AWL, or K1-34), this program also allows one to use custom word lists (e.g., BEL) in the analysis. The program incorporates the British National Corpus (BNC, 2007) as well as the Corpus of Contemporary American English (COCA; Davies, 2008) as reference corpora.

To answer the research questions stated above, three analyses were carried out in the present study. The first analysis was conducted to explore the distribution of words in K1, K2, AWL, and off-list in each ESP corpus in order to evaluate the level of vocabulary exposure to more and less frequent words offered to students in the three ESP courses. The second analysis was carried out to determine the coverage of basic engineering words (using BEL as a baseword list) in the three corpora. Finally, the third analysis was performed to analyze the vocabulary load of texts included in the three corpora (using BNC / COCA as baseword lists). In order to provide a more clear depiction of the vocabulary profiles of the texts included in the three corpora, function words and proper nouns were excluded from all three analyses (i.e., only content words were analyzed).

RESULTS AND DISCUSSION

The first analysis was conducted to explore the vocabulary demands on the discipline-specific texts in the three corpora. The results of the analysis are presented in Table 2 which includes information about the number of tokens (i.e., running words) in each of the four categories (i.e., K1, K2, AWL, and off-list words), their coverage percentage of each corpus (%Cov.), as well as the cumulative coverage percentage of the corpus (Cum.%). For example, the TPE corpus included 6,109 running words that belonged to the second 1,000 of most frequent word families (K2), which accounted for 24.19% of the total words included in this corpus. Together with the words belonging to the first 1,000 of most frequent word families (i.e., K1 + K2), they accounted for 62.02% of the words in that corpus. In other words, when reading a TPE text, students with a vocabulary of 2,000 word families would be able to recognize only about 62% of the content words in that text.

Table 2. *Distribution of Words in K1, K2, AWL, and Off-list in the ESP Corpora*

Word lists	TPE			CE			ChE		
	Tokens	% Cov.	Cum. %	Tokens	% Cov.	Cum. %	Tokens	% Cov.	Cum. %
K1	9,554	37.83	37.83	9,661	43.16	43.16	4,181	25.75	25.75
K2	6,109	24.19	62.02	5,026	22.45	65.61	2,386	14.69	40.44
AWL	3,508	13.88	75.90	5,227	23.35	88.96	3,524	21.70	62.14
Off-list	6,066	24.10	100.00	2,471	11.04	100.00	6,146	37.86	100.00

Note: Counts only include content words. Therefore, number of tokens in the table does match the number of tokens in each corpus.

Comparing the results across the three corpora, it is obvious that the pedagogical materials included in the CE corpus were less demanding in terms of the vocabulary requirements. In fact, over 43% of the running words in CE belonged to the first 1,000 of most frequent words in English. Furthermore, the AWL accounted for the largest percentage of running words (23.35%) in CE materials, followed by 21.70% of lexical coverage in ChE materials and 13.88% of lexical coverage in TPE texts, suggesting that pre-teaching these words prior to students enrolling in these ESP courses can potentially increase their comprehension of the discipline-specific texts. In terms of the off-list vocabulary, a list that is likely to include a large percentage of technical terms specific to each discipline, it is clear that ChE materials contain the largest percentage of these highly specific (and, therefore, much less frequent) lexical items (although, it should be noted that off-list includes all items that could not be classified as belonging to the other three word lists).

The second research question asked about the BEL coverage in the three corpora. As illustrated in Table 3, BEL provided good lexical coverage in all three corpora, accounting for almost 12% of running words in ChE, followed by almost 17% of words included in TPE, and over 18% of running words in CE. These results are not surprising, as the numbers are largely in line with those reported in Ward (2009b), as well as the outcome of the first analysis (reported above) which suggested that chemical engineering materials were much more technical in nature and included more specialized vocabulary.

Table 3. *Distribution of Basic Engineering Words in the ESP Corpora*

Word lists	TPE			CE			ChE		
	Tokens	%Cov	Cum%	Tokens	%Cov	Cum%	Tokens	%Cov	Cum%
On-list	4,226	16.75	16.75	4,108	18.35	18.35	1,923	11.84	11.84
Off-list	21,011	83.25	100.00	18,277	81.65	100.00	14,314	88.16	100.00

Note: Counts only include content words. Therefore, number of tokens in the table does match the number of tokens in each corpus.

Furthermore, the results of the second analysis also suggested that while the vocabulary included in the three corpora reflected the lexical idiosyncrasy of the discipline-specific texts, the three engineering disciplines also shared a set of vocabulary common to the engineering field. This set of vocabulary might be a good pedagogical resource to draw on when working with prospective engineering students whose language proficiency is low to be able to successfully comprehend engineering discourse. Also, in situations when it is not practical or feasible to develop and teach an ESP course that would target highly specialized content (e.g., chemical engineering or computer engineering) and, therefore, might include students from many different engineering disciplines, focusing on BEL will still expose students to useful vocabulary and will help prepare them for the lexical demands of discipline-specific texts, albeit only to a certain extent.

The last research question asked about the lexical threshold necessary for reasonable comprehension (i.e., to provide 95% coverage) of discipline-specific texts in the three corpora. The results of the third analysis which addressed this research question are presented in Table 4.

Table 4. *Distribution of Words in K1–K34 in the ESP Corpora*

Word lists	TPE			CE			ChE		
	Tokens	%Cov	Cum%	Tokens	%Cov	Cum%	Tokens	%Cov	Cum%
1st – 1000	9,554	37.83	37.83	9,661	43.16	43.16	4,181	25.75	25.75
2nd – 1000	6,109	24.19	62.02	5,026	22.45	65.61	2,386	14.69	40.44
3rd – 1000	5,709	22.62	84.64	5,579	24.92	90.53	2,686	16.54	56.98
4th – 1000	2,007	7.95	92.59	786	3.51	94.04	1,443	8.89	65.87
5th – 1000	604	2.39	**94.98**	452	2.02	**96.06**	845	5.20	71.07
6th – 1000	560	2.22	97.20	218	0.97	97.03	712	4.39	75.46
7th – 1000	503 1	1.99	99.19	77	0.34	97.37	870	5.36	80.82
8th – 34th	3-19	.65-.07	100.00	46-20	.21-.09	100.00	110-81	.78-.57	**94.01**

Note: Counts only include content words. Therefore, number of tokens in the table does match the number of tokens in each corpus.

As shown in Table 4, both TPE and CE reached 95% lexical coverage within the range of 4,500–5,000 word families (see percentages in bold in Table 4). This means that if a student has the knowledge of about 5,000 most frequent word families in English, they should be able to adequately comprehend discipline-specific texts in thermal-power and computer engineering. The numbers for TPE in the present study were slightly higher compared to those reported in Hsu (2014) who found that civil and mechanical engineering employed a much smaller vocabulary of about 3,500 word families to provide 95% coverage of discipline-specific texts. As for the CE texts, the threshold was comparable to the one reported in Hsu (2014), with CE materials requiring about 4,500–5,000 word families to comprehend those texts in both studies.

Regarding the pedagogical materials utilized in the ESP course in chemical engineering, the results of this study depicted a rather disturbing picture. As Table 4 shows, students' familiarity with 34,000 word families in English (which is a very unlikely scenario in and of itself) would not even allow them to comprehend 95% of those texts. These thresholds were alarmingly high, especially when compared to those reported in Hsu (2014) who determined that 5,500 word families provided adequate comprehension of chemical texts in the EEWL corpus. On the one hand, such high vocabulary thresholds computed in the present study might be a reflection of the specific discipline (chemistry) demanding a wider variety of vocabulary which includes a higher percentage of extremely infrequent (discipline-specific) words. On the other hand, these thresholds also suggested that a number of texts included in the corpus were likely to be extremely difficult for the students in the ESP course and, therefore, their inclusion in the course must be reconsidered. In addition to re-examining the pedagogical materials employed in the course, another recommendation was to consider additional ways to scaffold chemical engineering texts in the ESP class. This might involve developing a discipline-specific word list that would include important (and likely much less frequent) vocabulary items that can be targeted in the ESP class in order for students to be able to attain an adequate comprehension of those texts (see Kwary (2011) and Todd (2017) for excellent discussions of methods to identify relevant vocabulary).

CONCLUSION

The present study was conducted to explore the vocabulary demands of the pedagogical materials in the three engineering ESP corpora featuring texts in Thermal-Power, Computer, and Chemical Engineering. The most immediate goal of the study was to examine the extent to which the vocabulary demands were realistic in terms of meeting the level of students' language ability. The results indicated that chemical engineering texts were the most lexically challenging, with 34,000 word families only providing up to 94% of lexical coverage of these texts, followed by thermal-power engineering texts which required the knowledge of at least 5,000 word families for students to be able to reach the 95% comprehension threshold. Texts in computer engineering were the least demanding in terms of the vocabulary knowledge, as students would need to be familiar with 4,000–5,000 most frequent word families in order to reach minimal comprehension. These findings raise concerns about the appropriateness of the pedagogical materials employed in the chemical engineering ESP course and the extent to which this course was able to set realistic expectations for the students.

The results of the study also offer implications for a larger community of ESP/ LSP practitioners. Firstly, the study employed a corpus-based methodology to evaluate the pedagogical materials utilized in the three ESP courses from the lexical coverage perspective. This methodology can be adopted to other pedagogical contexts. Secondly, the study provided additional evidence for the validity of two existing word lists—AWL and BEL—in terms of their ability to account for large percentages of vocabulary included in the three discipline-specific ESP corpora. Both of these lists can be employed as potential sources to consult when selecting vocabulary to teach in ESP courses targeting basic engineering content.

REFERENCES

Association for Engineering Education of Russia (2018). Accessed on February 2, 2018, from http://aeer.ru/en/goals.htm

Anderson, N. (2015). Academic reading expectations and challenges. In N. Evans, N. Anderson, & W. Eggington (Eds.), *ESL Readers and Writers in Higher Education: Understanding Challenges, Providing Support* (pp. 95–109). New York: Routledge.

Archer, A. (2008). 'The place is suffering': Enabling dialogue between students' discourses and academic literacy conventions in engineering. *English for Specific Purposes, 27*, 255–266.

Bárcena, E., Read, T., & Arús, J. (2014). *Languages for Specific Purposes in the Digital Era*. New York: Springer.

Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 107–125). New York: Continuum.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 21–-238. DOI: 10.2307/3587951

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 million words, 1990–present*. Available online at http://corpus.byu.edu/coca/.

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes, 6*, 3–17. DOI: 10.1016/j.jeap.2006.11.005

Hartshorn, J., Evans, N., Egbert, J., & Johnson, A. (2017). Discipline-specific reading expectation and challenges for ESL learners in US universities. *Reading in a Foreign Language, 29*, 33–60.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). Range [Computer software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation/aspx.

Hirsh, D., & Nation, I.S.P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689–696.

Hsu, W. (2011). A business word list for prospective EFL business postgraduates. *Asian-ESP Journal, 7*, 63–99.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33,* 54–65.

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*, 403–430.

Hwang, K., & Nation, I.S.P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language, 6*, 323–335.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4–21.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes, 37*, 74–86.

Kaewpet, C. (2009). Communication needs of Thai civil engineering students. *English for Specific Purposes, 28*, 266–278.

Konstantakis, N. (2007). Creating a business word list for teaching business English. *ELIA, 7*, 79–102.

Kwary, D.A. (2011). A hybrid method for determining technical vocabulary. *System, 39*, 175–185. DOI: 10.1016/j.system.2011.04.003

Liu, N., & Nation, I.S.P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*, 33–42.

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language, 22*, 56–70.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*, 235–256.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*, 59–82.

Orna-Montesinos, C. (2008). A contribution to the lexis of construction engineering textbooks: The case of building and construction. *Ibérica, 16*, 59–80.

Pritchard, R., & Nasr, A. (2004). Improving reading performance among Egyptian engineering students: Principles and practice. *English for Specific Purposes, 23*, 425–445.

Rowley-Jolivet, E. (2015). Quantification in conference talks and proceedings articles in engineering. *English for Specific Purposes, 38*, 11–22.

Spence, P., & Liu, G. (2013). Engineering English and the high-tech industry: A case study of an English needs analysis of process integration engineers at a semiconductor manufacturing company in Taiwan. *English for Specific Purposes, 32*, 97–109.

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*, 31–39.

The British National Corpus, version 3 (BNC XML Edition) (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Available online at http://www.natcorp.ox.ac.uk/

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*, 442–458.

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language, 12*, 1999.

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes, 6*, 18–35.

Ward, J. (2009a). EAP reading and lexis for Thai engineering undergraduates. *Journal of English for Academic Purposes, 8*, 294–301.

Ward, J. (2009b). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*, 170–182.

West, M. (1953). *A general service list of English words*. London: Longman, Green.

Wolfe, J. (2009). How technical communication textbooks fail engineering students. *Technical Communication Quarterly, 18*, 351–375.

Wood, D., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes, 15*, 1–13.