A Corpus-Based Study: Authenticity Analysis of Technical Vocabulary from an In-House Textbook and Documents Used in Import and Export Procedures*

Kusuma Pitukwong
Nakhon Pathom Rajabhat University, Thailand.
Corresponding Author's Email: Kusuma.pw@gmail.com

Abstract

Textbooks are fundamental elements in teaching-learning encounters and therefore should provide real language input to students in order to enhance the effectiveness of their language learning. Through the use of corpus-based analysis, more authentic, content-rich and up-to-date textbooks can be produced that will provide further opportunities for students to experience input from real life rather than language items that have been artificially created. The main purpose of this study was to compile a small sized Textbook Corpus and a small sized Authentic Corpus to explore the authenticity levels of the in-house textbook used in an English for Import and Export course compared to the documents used in actual import and export procedures, which function as the source of authenticity in this context. Additionally, the frequent technical vocabulary lists in the Textbook Corpus and the Authentic Corpus were identified by using a hybrid method (a corpus-based approach combined with the use of a technical dictionary).

The results show that there were 96 technical vocabulary items, which covered approximately 1.23% of the running words in the Textbook Corpus. For the Authentic Corpus, there were 198 technical vocabulary items, which covered approximately 9.44%. In the aspect of the authenticity levels of the textbook, it was found that the in-house textbook has little similarity to authentic language in terms of technical vocabulary. As a result, the material writers should add these technical vocabulary items found in the top ranks of the Authentic Corpus into the textbook in order to achieve a higher degree of authenticity, and to help the learners focus on more important and frequent items found in real-life situations.

Keywords: Corpus; Technical Vocabulary; Authenticity

*Received November 13, 2020; Revised February 10, 2022; Accepted February 15, 2022

Introduction

Specialized or technical vocabulary has always been a major concern in the teaching of English for Specific Purposes (ESP) (Nation, 2001; Hsu, 2011; Wanpen, Sonkoontod, and Nonkukhetkhong, 2013). As stated by Robinson (1991), "It may often be thought that a characteristic of ESP is that a course should involve specialist language and content." Consequently, it is critical to prepare learners to acquire technical vocabulary in their texts because each field of study has its own characteristic and specialized vocabulary. Nation (2001) mentions that technical vocabulary consists of words that appear frequently in a specialized text or subject area but appear or occur infrequently in other fields. When used in a specific field, it can be a common word with a specialized meaning that differs from its common meaning (Lui, and Lei, 2020). It also has its own distinct character in relation to its area. There are five different approaches for creating a list of these words: 1) using a rating scale, 2) using a technical dictionary, 3) using clues provided in the text, 4) using a corpus-based approach or keyword analysis, and 5) using a hybrid method (Kwary, 2011; Tongpoon-Patanasorn, 2018; Ha, and Hyland, 2017).

The first three approaches rely on the intuition of the subject experts. In order to overcome this problem, the experts developed the corpus-based approach. This approach compares word frequencies in a technical corpus with the number of word occurrences in a reference corpus using automatic term extraction tools. (Chujo, and Utiyama, 2006; Lui, and Lei, 2020). The software will list words according to their keyness values (Ha, and Hyland, 2017). Positive keyness value, in this sense, refers to vocabulary items that appear more frequently in a technical corpus than in a reference corpus. These words are regarded as technical vocabulary (Anthony, 2004).

The corpus-based approach offers a number of advantages in the teaching of ESP. For example, Crawford (2008) used a corpus-based approach to analyze the transcripts of twelve business studies lectures, concentrating on linguistic features associated with academic, disciplinary, professional, cultural, and individual dimensions of identity. In addition, Durrant (2007); Walker (2011) used a corpus analysis to generate lists of important collocations for EAP students. Chang, and Kuo (2011) gathered a corpus-based list from three major journals of computer science in order to develop teaching materials for writing research articles. From the research mentioned, it can be concluded that the corpus-based approach is very beneficial for both teachers and learners of ESP, especially in developing materials.

In Thailand, the corpus-based approach is not a new idea; however, there have been few research studies using corpora (Pitukwong, 2012; Supatranont, 2005; Tongpoon-Patanasorn, 2018). In 2002, Poonpon (2002) used a science corpus in order to analyze the vocabulary input of English for Science courses at Mahidol and Khon Kaen Universities. The purpose of the research was to find the highly frequent vocabulary that science students were actually taught at the tertiary level. In addition, Supatranont (2005) employed the use of a concordance-based method in teaching academic reading for engineering students, in which corpus analysis was used to create teaching materials; for example, by using concordance output as gap-fill exercises, for a Vocabulary Level Test, and for the study of lexical features. It was concluded that the use of the concordance-based method provides crucial and abundant data on linguistic features, which helps to improve students' knowledge and proficiency in English learning.

Learners in ESP classes are exposed to a variety of texts, particularly texts in a specific context. Technical vocabulary, according to Sutarsyah, Nation, and Kennedy (1994), comprises terms that appear more frequently in specialized texts than in non-specialized texts, and it has been discovered that some learners struggle with this vocabulary while reading texts. Pitukwong (2012) evaluated students' knowledge of technical vocabulary in business texts, and the results showed that the scores achieved by this group of students were lower than 50%. Hence, it might be assumed that this group of students lacks relevant and sufficient expertise with technical vocabulary in business texts.

The findings of the previous research raise interesting questions concerning how technical language is taught and learned in Thailand. What technical vocabulary should be learned? How should it be taught to the students? Is the technical vocabulary taught in the classroom suitable for students' real-world implementation?

In order to help learners learn technical vocabulary, many researchers suggest that the materials and textbooks should be authentic. Authentic materials, according to Nunan (1989), are any materials that have not been created explicitly for the aim of language teaching but were created to fulfill some social purpose in the language community in which they were produced. Authentic materials, according to Peacock (1997), are crucial in the teaching of a foreign language because they prepare students for real-life situations. Moreover, a lack of authentic materials can cause confusion about what they have learned in the classroom and what they are going to use in the real-world tasks. In order to expose students to real-life

applications, many researchers suggest teachers apply authentic materials in the actual classroom.

In the Thai context, several studies have explored authenticity (Tanasavate, and Chinwonno, 2013), yet not many studies have investigated the degree of authenticity of the textbooks used in the actual classroom in relation to the language used in real-world situations. As a result, an attempt has been made to investigate the authenticity degree of textbooks used in ESP classrooms and the documents used in real-world implementation in terms of the technical vocabulary in an English for Import and Export Course, with the aim to develop more effective teaching materials to help students to improve their technical vocabulary learning.

Objectives of the Research

1. To explore the frequent technical vocabulary items in the Textbook Corpus and the Authentic Corpus.

2. To identify the similarities and differences between the Textbook Corpus and the Authentic Corpus, focusing on technical vocabulary.

Research Methodology

1st Step: The Design of the Corpora

This research is a descriptive study, with the purpose to explore similarities and differences between the Textbook Corpus and the Authentic Corpus focusing on technical vocabulary. By using the data acquired from the comparison of both corpora, the authenticity level of the English-language textbook used for the Import and Export Course was determined with the Authentic Corpus. There were three main corpora: the Textbook Corpus, the Authentic Corpus (as specialized corpora) and the British Academic Written English (BAWE) Corpus (as a reference corpus).

1. The Textbook Corpus

The first specialized corpus used in this study is entitled the “Textbook Corpus,” which was used as the main source to identify a list of frequent technical vocabulary, and also to compare with the Authentic Corpus to identify similarities and differences in terms of technical vocabulary. The Textbook Corpus was compiled from the in-house textbook and materials used in the English for Import and Export Course at Nakhon Pathom Rajabhat University in Thailand. This in-house textbook was developed by the researcher. Therefore, the size of the

Textbook Corpus is relatively small (49,695 words) because this textbook and materials were taught to a group of students who enrolled in the course English for Import and Export in one semester. However, many small sized corpora can provide effective results in specialized areas. Kwary (2011); Bowker, and Pearson (2002) both stated that smaller sized corpora with a few thousand to a few hundred thousand words are sufficiently useful in the study of language for specific purposes.

2. The Authentic Corpus

The second specialized corpus is named the "Authentic Corpus." This corpus is recognized as the main source for authentic discourse in real life since it was compiled from documents used in import and export procedures. It was compiled to identify the frequent technical vocabulary list and was also used to compare the similarities and differences with the Textbook Corpus. In compiling the Authentic Corpus, four main criteria were taken into consideration: representativeness, balance, size, and chronology (Mosel, 2018).

2.1 Representativeness

In terms of representativeness in the field of import and export, the boundaries of the target population were determined by consulting two experts in the import and export field who teach at the tertiary level. They were chosen because they can effectively provide lists of documents used in the import and export procedures that students in this area are likely to encounter in their future careers. These experts were asked to classify the document categories that should be included in the Authentic Corpus. The resulting list contained five document categories: commercial documents, banking documents, import documents, export documents, and customs documents. Consequently, there were 103 documents that were analyzed for the 50,536 running words in this corpus.

2.2 Balance

To reach the goal of balance, each subcategory contains a proximal number of running words. Approximately 10,000 running words were included in each category, for a resulting subcategory of approximately 50,536 running words in the Authentic Corpus.

2.3 Size

There are no fixed rules to determine the exact size of a corpus since it depends on the purpose of the study. As stated by O'Keeffe and Farr (2003) "the size of a corpus must be large enough to fairly represent the vocabulary occurrence within a language and can consist of millions of words, especially when analyzing collocation." Nevertheless, approximately 10,000 words are also acceptable in the case of a specialized corpus. Bowker, and Pearson

(2002) state that “a few thousand and a few hundred thousand words are just as useful in the study of languages for specific purposes.” Accordingly, the Authentic Corpus consisted of 50,536 running words.

2.4 Chronology

As stated by Pearson (1998), the samples of the text in the corpus should have been delivered in the last 10 years prior to the date of compilation. Hence, the Authentic Corpus in this study was compiled from documents dating from the period of time ranging from 2009 to 2019 in order to avoid obsoleteness.

3. The British Academic Written English (BAWE) Corpus

The British Academic Written English (BAWE) Corpus was used as a reference corpus for comparison to the words in the Textbook Corpus and the Authentic Corpus in order to identify technical vocabulary. This reference corpus consists of 6,506,995 words from 2,761 pieces of academic written English words from four broad disciplines: Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences. It was chosen as a reference corpus because it provides a reasonable number of words, which allows for more chances of vocabulary occurrences when compared with the Textbook Corpus and the Authentic Corpus.

2nd Step: Research Instruments

The researcher adopted a hybrid approach (using a corpus-based approach and a technical dictionary) in order to identify a list of frequently used technical words in the field of import and export.

1. The Dictionary of Import and Export

“The Dictionary of Import and Export” compiled and edited by Mike (2007) was used to identify the technical words. If a word occurs in any technical dictionary, it is considered to be a technical word. While it is easy to conduct, this method is likely to ignore some higher and some lower-frequency technical words. In this sense, these lists include no information about word frequency; therefore, this approach offers no guidance about how learners should focus their efforts in the area of import and export.

2. AntConc Software (version 3.5.8)

Because using a dictionary to identify technical words does not show word frequency, the researcher used AntConc 3.5.8 (Anthony, 2019) in the present study to identify frequent technical vocabulary in the Textbook Corpus and Authentic Corpus. This program is a freeware program that offers the essential tools needed for corpus analysis such as keyword frequency generators, tools for cluster and lexical bundle analysis, concordancing functions and word

distribution plotting. This concordancing software was developed by Professor Laurence Anthony of Waseda University, Tokyo, Japan, and it can be downloaded from the Laurence Anthony Laboratory website. This program is easy-to-use, multi-purpose, and can run on any Windows operating system, including Windows 98, ME, NT, and XP. It will also run on Macintosh OSX and Linux operating systems. The latest version is AntConc 3.5.8, which was released in 2019.

3rd Step: Data Analysis

There were three main corpora used in the present study: the Textbook Corpus, the Authentic Corpus (as a specialized corpora), and the British Academic Written English (BAWE) Corpus (as a reference corpus). In order to identify frequent technical vocabulary lists from the Textbook Corpus and the Authentic Corpus, the researcher employed the "hybrid method" by using a corpus-based approach and the technical dictionary. The rationale for selecting the "hybrid method" was because it can overcome the weaknesses of singular analytic methods (Tongpoon-Patanasorn, 2018).

Initially, AntConc software version 3.5.8 (Anthony, 2019) was employed to automatically compare the degree of word frequency in the Textbook Corpus with the word frequency in the BAWE Corpus. Similarly, to identify the technical vocabulary in the Authentic Corpus, the Keyword List tool of AntConc software was used to compare the list of words from the Authentic Corpus with that of the reference corpus (BAWE Corpus). The Keyword List tool of AntConc offers the function of comparing the frequency of the wordlist based on the specialized corpus with the frequency of a wordlist based on the reference corpus. The tool then lists words according to their "keyness values." In this sense, words that have higher or positive keyness are words that are significantly more frequent in the specialized corpus than in the reference corpus. These words are considered to be technical vocabulary. In contrast, "negative keyness" refers to words that appear significantly less frequently in the specialized corpus than in the reference corpus. These words are not included in the technical vocabulary list. Therefore, in this study, it was required that technical vocabulary show positive keyness values when compared to a reference corpus (BAWE Corpus).

However, this analysis only results in single-word units. This is a problem in determining the technical vocabulary of the import and export area because most terms are multi-word units. Consequently, it is necessary to include the multi-word units instead of only the single-word units. Therefore, the use of keyword analysis alone is inadequate to determine the technical vocabulary of import and export. To deal with this point, the AntConc program has

another menu, called Clusters, which can generate multi-word units. However, the researcher must decide based on his/her own judgment which words are to be shown in the clusters. Some of the results are not particularly appropriate for inclusion as technical vocabulary. As a result of this flaw, the researcher decided to use a secondary method; that is, using a technical dictionary to determine whether these words are technical or not. If a vocabulary item appears in the technical dictionary, it is recognized as a technical word. Therefore, these words were checked against the "The Dictionary of Import and Export" of Mike (2007), in order to ensure that the terms are relevant to the particular subject field. After identifying the technical vocabulary by using the hybrid method, the lists of frequent technical vocabulary from the Textbook Corpus and the Authentic Corpus were used for comparison of the similarities and differences.

Research Results

The 1st objective is to explore the frequent technical vocabulary items in the Textbook Corpus and the Authentic Corpus.

Table 1 The 20 Most Frequent Technical Vocabulary Items in the **Textbook Corpus**

Rank	Technical Vocabulary	Occurrences	Rank	Technical Vocabulary	Occurrences
1	Insurance	48	11	Letter of credit	17
2	Bill of lading	30	12	Open account	17
3	Currency	25	13	Tariff	17
4	Delivery	25	14	Customs	15
5	Cargo	22	15	Packing list	14
6	Devaluation	21	16	VAT	14
7	Logistics	21	17	Cash in advance	11
8	Commercial invoice	19	18	Collection	11
9	Consignment	19	19	Exchange rate	11
10	Carrier	17	20	Proforma invoice	11

In the aspect of the frequent technical vocabulary lists in the Textbook Corpus, Table 1 shows that there were 96 technical vocabulary items. These words occurred repeatedly 612 times (word occurrences), so they covered approximately 1.23% of the running words in the Textbook Corpus. It can be seen that the top 10 most frequently occurring words were insurance, bill of lading, currency, delivery, cargo, devaluation, logistics, commercial invoice, consignment, and carrier, respectively.

Table 2 The 20 Most Frequent Technical Vocabulary Items in the **Authentic Corpus**

Rank	Technical Vocabulary	Occurrences	Rank	Technical Vocabulary	Occurrences
1	Customs	626	11	Letter of credit	108
2	Duty	368	12	Declaration	101
3	Application	229	13	Carrier	92
4	Applicant	170	14	Agent	83
5	Currency	165	15	Guarantee	82
6	Draft	160	16	License	81
7	Insurance	151	17	Bill of lading	70
8	Receipt	143	18	Collection	70
9	Default	122	19	Cargo	69
10	Beneficiary	121	20	Documentary	69

As shown in Table 2, there were 198 technical vocabulary items in the Authentic Corpus. These technical vocabulary items occurred repeatedly 4,771 times (word occurrences), so these words covered approximately 9.44% of the running words in the Authentic Corpus. The top 10 most frequently occurring words were customs, duty, application, applicant, currency, draft, insurance, receipt, default, and beneficiary, respectively. These words are highly topic-related to the documents used for imports and exports procedures, such as application, currency, draft, receipt, beneficiary, letter of credit, and bill of lading.

The 2nd objective is to identify the similarities and differences between the Textbook Corpus and the Authentic Corpus, focusing on technical vocabulary.

Table 3 The 10 Overlapping and non-overlapping Technical Vocabulary in the Textbook Corpus and the Authentic Corpus

No.	Overlapping Technical Vocabulary	Occurrences in the Textbook Corpus	Occurrences in the Authentic Corpus	No.	Non-Overlapping Technical Vocabulary	Occurrences in the Authentic Corpus
1	Acceptance	3	21	1	Application	229
2	Advance payment	5	10	2	Default	122
3	Advising bank	4	10	3	Agent	83
4	Air waybill	1	26	4	Documentary	69
5	All risks	2	5	5	Damages	53
6	Alongside	1	4	6	Amendment	48
7	Applicant	1	170	7	Principal	47

8	At sight	3	14	8	Negotiation	43
9	Availability	1	2	9	Discount	38
10	Bank guarantee	1	1	10	Payee	37

In order to explore the similarities and differences between the technical vocabulary in the Textbook Corpus and the Authentic Corpus, the researcher compared the overlapping and non-overlapping parts of these two vocabulary lists (see Table 3). The 81 overlapping words contribute to the coverage of 40.9%. The examples of overlapping words were applicant, beneficiary, bill of lading, cargo, carrier, currency, customs, delivery, declaration, duty, draft, letter of credit, etc. The words that are in the Authentic Corpus but are not in the Textbook Corpus, accounting for 59.1%, include 117 words. The top 10 non-overlapping words were application, default, agent, documentary, damages, amendment, principal, negotiation, discount, and payee, respectively.

The New Body of Knowledge

The proportion of technical vocabulary of the Textbook Corpus is slightly different from that of the previous study conducted by Nation (2001), which claimed that “technical vocabulary covered more than 5% of running words in specialized texts, and was made up of words that occurred frequently in a specialized text or subject area but did not occur or were of very low frequency in other fields.” In the present study, these words covered only 1.23% of the running words in the Textbook Corpus. In contrast, the percentage of technical vocabulary in the Authentic Corpus covered about 9.44% of the running words. This result is relatively similar to the previous research in terms of the overall characteristics of technical vocabulary. For example, Romer and Wulff (2010) compiled a biology corpus from 64 papers written by students in Biology in order to identify technical vocabulary. These technical vocabulary items covered about 5.7% of the running words in the biology corpus and are rarely found in other disciplines. The findings mentioned were relatively similar to the results of the current study since 9.44% of the words found in the Authentic Corpus were technical words. These results support the overview of Hyland, and Tse (2007): “technical vocabulary which is dictated by the subject area and typically covers around 5% of academic texts.” In other words, it can be explained that technical vocabulary is a small group of highly topic-related words that occur frequently, accounting for around 5% of the running words in the corpus.

Discussion of the Research Results

1. Frequent Technical Vocabulary Items in the Textbook Corpus and Authentic Corpus

In the aspect of the frequent technical vocabulary list in the Textbook Corpus, it was shown that there were 96 technical vocabulary items. These words occurred repeatedly 612 times (word occurrences), and thus, these words covered approximately 1.23% of the running words in the Textbook Corpus. It can be seen that the top 10 most frequently occurring words were insurance, bill of lading, currency, delivery, cargo, devaluation, logistics, commercial invoice, consignment, and carrier, respectively. For the Authentic Corpus, there were 198 technical vocabulary items. These technical vocabulary items occurred repeatedly 4,771 times (word occurrences), and therefore, these words covered approximately 9.44% of the running words in the Authentic Corpus. It can be seen that the top 10 most frequently occurring words were customs, duty, application, applicant, currency, draft, insurance, receipt, default, and beneficiary, respectively. These words are highly topic-related and are a direct reflection of the field of imports and exports. The findings of the present study closely correspond to Kwary's (2011) results in terms of highly topic-related words, in which the hybrid method to identify technical vocabulary in the business corpus was also used, and it was found that the hybrid method is more effective and accurate in identifying technical vocabulary than other methods used in isolation. As mentioned by Chung, and Nation (2004), the corpus-based method provides a consistency rate of around 91.9% in identifying technical terms, and it also gives the frequency of technical words, which can help teachers and students make more effective decisions when selecting the words to teach and to learn.

However, in this study it was found that the use of a computer-based approach only results in single units. This is the main problem in identifying the technical vocabulary in the area of import and export because the majority of the technical terms are multi-word units, such as "letter of credit", "bill of lading", "commercial invoice", etc. To solve this problem, Cluster, the subprogram of the AntConc software was used to identify technical multi-word units as it allows users to search for multi-word expressions immediately to the left or right of the search term. However, the results of using the Cluster subprogram might be vague because there are some multi-word units that do not really make sense with regard to being included in a list of technical vocabulary. For example, the word "Clean bill of" should not be included because it does not have a specific meaning. It is thus necessary to use our individual intuition to select which words are to be shown in the cluster, the cluster size and the minimum

clusters frequency. Due to this drawback, using a specialized dictionary is the second method to identify a list of technical vocabulary. In this method, unknown vocabulary from the specialized corpora was checked against the Dictionary of Import and Export, compiled and edited by Mike (2007). If the vocabulary item appears in a technical dictionary, it is a technical word. Using the specialized dictionary also provides examples of usage, grammatical information, and the pronunciation about each word. Therefore, it can be concluded that each method used to identify technical vocabulary has different benefits and drawbacks. To overcome these weaknesses, using a hybrid method can strengthen and provide more validity than using a single analytic method.

However, the proportion of technical vocabulary of the Textbook Corpus is slightly different from the previous study conducted by Nation (2001), who claimed that "technical vocabulary covered more than 5% of the running words in specialized texts, and was made up of words that occurred frequently in a specialized text or subject area but did not occur or were of very low frequency in other fields." In the present study, these words covered only 1.23% of the running words in the Textbook Corpus. The slightly different occurrences of the technical vocabulary percentages might result from the unrepresentativeness of the Textbook Corpus. As a result, the possibility of technical vocabulary occurrences might be lower than 5% of the running words in specialized texts. According to Biber (1993), a corpus must be 'representative' in order to be appropriately used as the basis for generalizations concerning a language as a whole. In contrast, the percentage of technical vocabulary in the Authentic Corpus covered approximately 9.44% of the running words. This result is relatively similar to the previous research in terms of the overall characteristics of technical vocabulary. For example, Romer, and Wulff (2010) compiled a biology corpus from 64 papers written by students in Biology in order to identify technical vocabulary. This reveals that wordlists derived from the biology corpus are from the discipline of biology, including such words as "species," "genetic," "plague," and "mutation." These technical vocabulary items covered approximately 5.7% of the running words in the biology corpus and are rarely found in other disciplines. The findings mentioned were relatively similar to the results of the current study since 9.44% of the words found in the Authentic Corpus were technical words. These results support the overview of Hyland and Tse (2007): "technical vocabulary which is dictated by the subject area and typically covers around 5% of academic texts." In other words, researchers can explain that technical vocabulary is a small group of highly topic-related words which occur frequently, accounting for around 5% of the running words in the corpus.

2. The Number of Overlapping and Non-Overlapping Technical Vocabulary in the Textbook Corpus and the Authentic Corpus with regard to the authenticity levels of the textbook used in the English for Import and Export course at Nakhon Pathom Rajabhat University, the findings showed that there were 81 similar technical words in the Textbook Corpus and the Authentic Corpus, which contribute to the 40.9% coverage. This study reveals that the technical vocabulary in the Textbook Corpus and the Authentic Corpus are obviously different. According to Barbieri, and Eckhardt (2007), "corpus-based analysis is an ideal tool to re-evaluate the order of presentation of linguistic features in textbooks and to make good decisions about what to prioritize in textbook presentations" (p. 322). From this statement, we can state that the textbook used in English for the Import and Export Course does not greatly reflect authentic language use in the target language. The partial use of technical vocabulary may provide the material developers with an awareness to review the textbook and to give more attention to the explored technical vocabulary lists.

Since textbooks are the main components of teaching and the main source of input for students, their closeness to real life-language use is very crucial in the acquisition of a correct understanding of the target language. If the technical vocabulary are taught unsystematically, students might become confused with the target language and discover that their learning is irrelevant to the real-world job tasks. Consequently, students should receive more opportunities to be exposed to real-life input rather than mechanically produced language items (Peksoy, and Harmodlu, 2017). Therefore, the explored technical vocabulary and their weight in the textbooks should be reconsidered and readjusted in order to have more authentic course materials. For example, the word "application", which occurred as the third rank in the Authentic Corpus at 229 times did not occur in the Textbook Corpus. In general, "application" means a formal (often written) request for something, such as a job, permission to do something or a place at a college or university (Oxford Dictionary, 2019). As a technical word in the area of import and export, it refers to "instructions from the applicants to the issuing bank to open letter of credit" (Mike, 2007). In addition, the word "default" came as the ninth rank and occurred repeatedly 122 times in the Authentic Corpus but did not occur in the Textbook Corpus. In a general sense, it means "what happens or appears if you do not make any other choice or change, especially in a computer program" (Oxford Dictionary, 2019), but it means "the failure to meet the legal obligation of a loan" (Mike, 2007) in the import and export field. Based on these examples, the material writers should add those technical vocabulary items found in the top ranks of the Authentic Corpus into the

textbook in order to achieve a higher degree of authenticity, and help learners focus on more important and frequent items in real-life situations. Consequently, the learners will have more opportunities and the possibility to be prepared for real-life communication situations.

Conclusions

The study aimed to compile the Textbook Corpus and the Authentic Corpus to investigate the authenticity levels of the in-house textbook used in the English for Import and Export course at Nakhon Pathom Rajabhat University compared to the documents used in actual import and export procedures, which functioned as the source of authenticity in this context. Additionally, the frequent technical vocabulary lists in the Textbook Corpus and the Authentic Corpus were identified by using a hybrid method. In terms of corpus analysis, this study provides some initial practical input for materials development in English for Import and Export courses. The lists of technical vocabulary obtained by using AntConc can be implemented as a resource in future ESP materials development. The high frequency technical vocabulary from the corpus analysis should be used as a criterion for selecting vocabulary introduced in the English for Import and Export course, especially those vocabulary items in the top rank. The course instructor can easily and practically design many kinds of vocabulary exercises such as gap-filling, multiple-choice, or even cloze passages. For example, the teacher can generally blank out selected words and ask the students to fill in the blanks. The important thing is that the students can find the answers in authentic texts, which helps them improve their guessing of the meaning of technical vocabulary. Apart from vocabulary exercises, the teacher may also design a technical vocabulary test. However, some of the contexts of the selected vocabulary items may need to be changed in cases where the original texts are sometimes too difficult for the learners. Because of all these benefits, it can be suggested that the most frequent technical vocabulary items should be used as a resource for future materials development in English for Import and Export courses in order to motivate the students to learn vocabulary effectively in an authentic context.

Suggestions

From the results of the research, the researcher has suggestions as follows:

1. The suggestions from the research

1.1 The study was limited to a particular textbook used for a small group of students. In the future, compiling a larger size and more representative corpora might be helpful to gain more insightful about the effectiveness of these textbooks and materials.

1.2 For the ESP teacher, it is suggested that the lists of technical vocabulary obtained by using AntConc, especially the words in the top rank, should be used as a resource for future materials development in English for Import and Export courses in order to motivate the students to learn vocabulary effectively in an authentic context.

2. The suggestions for future research

2.1 Apart from focusing on a list of technical vocabulary, other areas of language might be investigated. There are various lexico-grammatical features whose analyses would be valuable to teaching. For instance, further research might compare the syntactic and semantic functions of passive in the two languages in greater depth.

2.2 Further research may explore vocabulary learning strategies which high and low English proficiency students employ to improve their technical vocabulary acquisition based on the acquired technical vocabulary list from the corpus analysis.

References

Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-platform Corpus Analysis toolkit, *Proceedings of IWLeL: An Interactive Workshop on Language E-learning* (pp.7-13). Toky: Japan.

Anthony, L. (2019). *AntConc (version 3.5.8) [computer program software].* Retrieved November 3, 2019, from: http://www.laurenceanthony.net/

Barbieri, F., & Eckhardt, S. E. B. (2007). Applying Corpus-Based Findings to Form-focused Instruction: The Case of Reported Speech. *Language Teaching Research,* 11(3), 319-346.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing,* 8(4), 243-257.

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. New York: Routledge.

Chang, C.F., & Kuo, C.H. (2011). A Corpus-Based Approach to Online Materials Development for Writing Research Articles. *English for Specific Purposes,* 30(3), 222-234.

Chujo, K., & Utiyama, M. (2006). Selecting Level-Specific Specialized Vocabulary Using Statistical Measures. *System,* 34(1), 255-269.

Chung, M., & Nation, P. (2004). Identifying Technical Vocabulary. *System,* 32(2)*,* 251–263.

Crawford, W. (2008). Place and Time Adverbials in Native and Non-native English Student Writing. In A. Ädel & R. Reppen (Eds.), *Corpora and Discourse: the Challenges of Different Settings* (pp. 267-89). Amsterdam & Philadelphia: John Benjamins.

Durrant, P. (2007). Review of Nadja Nesselhauf, 'Collocations in a Learner Corpus'. *Functions of Language,* 14(2), 251-261.

Ha, A., & Hyland, K. (2017). 'What is Technicality? A Technicality Analysis Model for EAP Vocabulary'. *Journal of English for Academic Purposes,* 28(2), 35-49.

Hsu, W. (2011). The Vocabulary Thresholds of Business Textbooks and Business Research Articles for EFL Learners. *English for Specific Purposes*, 30(2), 247–257.

Hyland, K., & Tse, P. (2007). Is there an Academic Vocabulary? *TESOL Quarterly*, 41(1), 235–253.

Kwary, D. (2011). A Hybrid Method for Determining Technical Vocabulary. *System*, 39(2), 175-185.

Lui, D., & Lei, L. (2020). *Technical Vocabulary*. Oxfordshire: Routledge.

Mike, R. (2007). *The Dictionary of Import and Export*. New York: Lotus Press.

Mosel, U. (2018). *Corpus compilation and exploitation in language documentation projects.* Oxford: Oxford University Press.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nunan, D. (1989). *The Learner-Centred Curriculum: A Study in Second Language Teaching*. Cambridge: Cambridge University Press.

O'Keeffe, A., & Farr, F. (2003). Using Language Corpora in Language Teacher Education: Pedagogic, Linguistic and Cultural Insights. *TESOL Quarterly,* 37(3), 389-418.

Peacock, M. (1997). The Effect of Authentic Materials on the Motivation of EFL Learners. *ELT Journal*, 51(2), 144–156,

Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company.

Peksoy, E., & Harmodlu, O. (2017). Corpus Based Authenticity Analysis of Language Teaching Course Books. *International Journal of Languages' Education and Teaching,* 5(4), 287-307.

Pitukwong, K. (2012). *An Investigation of Vocabulary Learning Strategies Employed by High and Low English Proficiency Students based on Selected Technical Vocabulary from a Business Corpus. (*Master's Thesis). Mahidol University. Nakhon Pathom.

Robinson, J. (1991). *ESP today: A Practitioner's Guide*. Hemel: Hempstead.

Romer, U., & Wulff, S. (2010). Applying Corpus Methods to Writing Research: Explorations of MICUSP. *Journal of writing research,* 2(2), 99-127.

Stevenson, A. (2019). *Oxford Dictionary of English*. Oxford: Oxford University Press.

Supatranont, P. (2005). *A Comparison of the Effects of the Concordance-Based and the Conventional Teaching Methods on Engineering Students' English Vocabulary Learning*. (Doctoral Dissertation). Chulalongkorn University. Bangkok.

Sutarsyah, C., Nation, P., & Kennedy, G., (1994). How Useful is EAP Vocabulary for ESP? A Corpus -based Study. *RELC Journal,* 25(1), 34–50.

Tanasavate, K., & Chinwonno, A. (2013). A Study of Thai EFL Learners' and Teachers' Use of Authentic Materials. *An Online Journal of Education,* 8(2), 178-190.

Tongpoon-Patanasorn, A. (2018). Developing a Frequent Technical Words List for Finance: A Hybrid Approach. *English for Specific Purposes*, 51(3), 45-54.

Wanpen, S., Sonkoontod, K., & Nonkukhetkhong, K. (2013). Technical Vocabulary proficiencies and Vocabulary Learning Strategies of Engineering Students. *Procedia-Social and Behavioral Sciences,* 88(1)*,* 312-320.

Walker, C. (2011). How a Corpus-Based Study of the Factors which Influence Collocation can Help in the Teaching of Business English. *English for Specific Purposes,* 30(4), 101–112.