A corpus-based study of specialized vocabularies in industrial work for developing industrial word lists of English for science and technology class

การศึกษาคลังข้อมูลคำศัพท์เฉพาะในงานอุตสาหกรรมเพื่อพัฒนารายการคำศัพท์อุตสาหกรรมสำหรับวิชาภาษาอังกฤษวิทยาศาสตร์และเทคโนโลยี

Received: *March 4, 2020*
Revised: *July 9, 2020*
Accepted: *October 29, 2020*

Patcha Bunyarat
พัทชา บุญยะรัตน์
Faculty of Humanities and Social Sciences, Thailand
คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม, ประเทศไทย
patcha_bunyarat@hotmail.com

Abstract

In English for Specific Purposes (ESP), area-specific language and vocabulary are key success factors for learners. This paper aimed to investigate the most frequently-occurring specialized vocabularies in industrial sectors based on authentic materials, in an effort to create Industrial Word Lists (IWL) that can be used in English for science and technology classes. The corpus was developed from three sub-fields of industry, and includes 19,582 running words. AntConc (version 3.5.8) was used to determine the viability of both the Industrial Word List and of English for science and technology textbooks. The 439 words included in the Industrial Word List (IWL) were collected and analyzed by comparing them against three word lists that are currently used. Overlap was found between the IWL and these other word lists, namely, the Outside Word List (OWL) with an overlap of 132 words (31.21%), the General Service List (GSL) with an overlap of 222 words (50.57%) and the Academic Word List (AWL) with an overlap of 80 words (18.22%). This comparative study of keyword lists with authentic industrial materials and English for science and technology textbooks utilizing AntConc reveals that most authentic industrial materials include specialized language. However, there is a shortage of specialized vocabularies for specific careers in textbooks. Finally, the pedagogical implications are offered for English for science and technology classes by improvement of the English for science and technology textbooks.

Keywords: corpus-based study, word lists, specialized vocabulary, industrial sector, AntConc program

บทคัดย่อ

ในการจัดการเรียนการสอนภาษาเพื่อวัตถุประสงค์เฉพาะ ภาษาเฉพาะที่เกี่ยวข้องกับสายงานต่าง ๆ เป็นสิ่งสำคัญในการเรียนรู้ งานวิจัยนี้มีวัตถุประสงค์เพื่อสืบค้นคำศัพท์เฉพาะที่ใช้บ่อยในภาคอุตสาหกรรม โดยศึกษาจากคลังข้อมูลภาษาที่เป็นเอกสารที่ใช้จริงในภาคอุตสาหกรรม และพัฒนาเป็นรายการคำศัพท์เฉพาะอุตสาหกรรม เพื่อใช้ในวิชาภาษาอังกฤษวิทยาศาสตร์และเทคโนโลยี แหล่งคลังข้อมูลภาษาที่ใช้ในการสืบค้นมีขนาดคำศัพท์จำนวน 19,582 คำ โปรแกรม AntConc (เวอร์ชั่น 3.5.8) ถูกใช้เพื่อวิเคราะห์ความถี่และรายการคำศัพท์เฉพาะในงานอุตสาหกรรมและในหนังสือเรียนภาษาอังกฤษวิทยาศาสตร์และเทคโนโลยี ผลการวิจัยได้รายการคำศัพท์เฉพาะอุตสาหกรรมที่ปรากฏบ่อยจำนวน 439 คำ นำมาวิเคราะห์ประเภทรายการคำศัพท์ว่าเป็นกลุ่มคำศัพท์เฉพาะ (Outside Word List: OWL) หรือคำศัพท์ทั่วไป (General Service List: GSL) หรือคำศัพท์วิชาการ (Academic Word List: AWL) ผลปรากฏว่า คำศัพท์เฉพาะอุตสาหกรรมที่พัฒนาได้มีลักษณะผสมประเภท คือ เป็นคำศัพท์ทั่วไปจำนวน 222 คำ (ร้อยละ 50.57) คำศัพท์วิชาการจำนวน 80 คำ (ร้อยละ 18.22) และมีลักษณะเป็นคำศัพท์เฉพาะจำนวน 132 คำ (ร้อยละ 31.21) ผลการเปรียบเทียบประเภทรายการคำศัพท์ที่ปรากฏในเอกสารที่ใช้จริงในภาคอุตสาหกรรมกับในหนังสือเรียนภาษาอังกฤษวิทยาศาสตร์และเทคโนโลยีโดยใช้โปรแกรมความเป็นศัพท์เฉพาะหัวข้อ (keyness) ใน AntConc แสดงให้เห็นว่าคำศัพท์ในเอกสารที่ใช้จริงในภาคอุตสาหกรรมส่วนใหญ่เป็นคำศัพท์เฉพาะ ส่วนในหนังสือเรียนยังขาดแคลนคำศัพท์เฉพาะที่จำเป็นต้องใช้ในการทำงานเฉพาะทาง ผลการวิจัยนำไปประยุกต์ใช้ในการปรับปรุงหนังสือเรียนในรายวิชาภาษาอังกฤษวิทยาศาสตร์และเทคโนโลยี

คำสำคัญ : การศึกษาตามแนวทางคลังข้อมูลภาษา, ประเภทรายการคำศัพท์, คำศัพท์เฉพาะ, ภาคอุตสาหกรรม, โปรแกรม AntConc

1. Introduction

Thailand aims to encourage foreign investment in Thailand by developing New S-Curve industries as part of its Thailand 4.0 policy. Ten targeted industries were divided into two broad categories:

1) The First 5 S-Curve industries: Next-Generation Automotive, Intelligent Electronics, Advanced Agriculture and Biotechnology, Food Processing and Tourism.

2) The Second 5 New S-Curve industries: Digital, Robotics and Automation, Aviation and Logistics, Biofuels and Biochemicals, and Medical Hub.

There are several issues that need to be addressed in order to increase foreign investment in Thailand. One such issue concerns the English proficiency of Thai workers. The Bureau of Academic Affairs (2018, pp.13-14) produced a report that identified the skill expectations of employers regarding their workers. This report found that there is an enormous gap between the expectations of employers, and their workers' English language proficiency. This is similar to Suwanawela (2019) who pointed to the low English proficiency of graduates and students in vocational training colleges.

One difficulty Thai workers face is the difference between the structure and pronunciation of vocabulary. Vocabulary is considered the language element that is key to successful language learning. Many researchers have reported on the significance of vocabulary knowledge (Nation, 2001). According to Wilkins (1972, pp.111-112, cited in Ibrahim, 2015, p.1), the importance of vocabulary is that "while without grammar very little can be conveyed, without vocabulary nothing can be conveyed". In the case of specialized vocabularies or technical words used in industries, there are many words which are words not found in daily life and found only occasionally in more than one word form. When a specialized vocabulary term appears in more than one word form or compound form, it tends to confuse students, because in English, such as in the case of compound words, the head is on the right or final position, while in Thai, the head is on the left. The difference in compound order could have negative effects on language comprehension.

There are various studies which have indicated that learning English vocabulary is a major problem among Thais (Promwatcharanon & Chatreepinyo, 2018; Saenpakdeejit, 2014). Awareness of the importance of teaching and learning English vocabulary and the process of making vocabulary lists are still based on international standard levels, such as vocabulary wordlists based on the CEFR (Common European Framework of Reference for Languages), wordlists designed for standardized tests, such as TOEIC or IELTS (Chujo & Oghigian, 2009), or wordlists compiled for use in the social sciences (Chanasattru & Tangkiengsirisin, 2016). Wordlists have been found to be useful for English language learning, especially in general English

classes. However, as it pertains to teaching students in specific fields, especially in relation to vocabulary teaching, previous research from Ward (2009) has shown that lexical collocations and technical vocabulary make it difficult for learners to attain mastery of the terminology of their specific fields. Hsu (2014) also identified the same problem, that students not majoring in English lacked the lexical knowledge necessary to fulfill the needs of their respective careers.

To respond to the demand for skilled workers and mechanics, many universities in Thailand open courses for developing students in various industrial fields. At Pibulsongkram Rajabhat University of Thailand, there are five such majors: industrial technology, logistics engineering, computer engineering, industrial product design, and ceramics. All students in these fields have to register for English for science and technology as one of their core courses. In addition, at the nearby vocational training college, there is a course called English for Industrial Technology, intended to increase students' proficiency in English.

In ESP classes, students studying in industrial fields typically lack adequate English proficiency and vocabulary knowledge. This may lead to these students ultimately being unqualified for industries and employers who expect English proficiency from them. As mentioned in Wu (2014, p.120), a limited vocabulary is the primary obstacle for technical college students who have difficulty comprehending English textbooks. In reality, students tend to be taught general English vocabulary. However, ESP classes must focus their efforts on teaching the vocabulary that is necessary for a particular topic or field. Doing so can benefit students in many ways, such as decreasing students' learning loads. Instead of having to remember words that are not used frequently in their specific fields, they can instead focus on the vocabulary which is necessary to meet the needs of the industries in which they are going to enter. Furthermore, as seen in the following previous research, the research on word lists for Thai students is not carried out in various fields and is constructed based on Thai contexts.

This paper focuses on identifying the specialized vocabularies in industrial sectors by utilizing the corpus method, which involves comparing them to words already appearing in textbooks and offering suggestions on how to apply them for course design or instructional techniques. In a nutshell, the specialized vocabularies found in industrial authentic materials in the Thai context and in textbooks of English for science and technology are the target groups of this research. Furthermore, some tips for teaching English for science and technology are provided for pedagogical purposes. This study focuses on specialized vocabularies in industries since they play a key role in language use for communication.

2. Literature review

2.1 Corpus linguistics

Corpus, or Corpora in plural, is a collection of naturally occurring language texts, chosen to characterize a state or variety of a language (Sinclair, 1991). Therefore, corpus linguistics is not a branch of linguistics, such as sociolinguistics or psycholinguistics that studies about a specific field. It relates to language studies by using the corpora as a data-driven tool for language learning (Aroonmanakun, 2010, p.3).

Corpus linguistics makes us understand what patterns are associated with lexical and grammatical features and helps us answer various questions. One of these questions is "What are the most frequent words and phrases in English?" (Bennett, 2010). It has also been influential in informing dictionaries, grammar books, usage manuals, syllabuses, evaluation methods, and course materials (Boulton & Tyne, 2013).

From the characteristics and importance of corpus linguistics explained above, we can say that corpus linguistics is a significant field that influences various language studies and researches as seen from the following review.

2.2 The role of corpus-based study of English for specific purposes (esp)

Data-driven learning, learning in which teachers or learners are able to make direct use of corpora themselves (Boulton & Tyne, 2013), is an influential method that enhances second language and foreign language learning due to its focus on authentic language usage. Corpora allow researchers to analyze language in a number of different ways. Corpora can be built up by collecting related language materials and recording the target language, words, grammar, and collocations within. According to Hunston (2005), a corpus can be used to show word frequency, phraseology, and collocation. Additionally, corpora can generate frequency lists which can be useful for identifying possible differences between the corpora. The categories of corpora are distinguished based on their particular purposes as stated below.

A specialized corpus is a corpus of texts from a particular genre, such as newspaper editorials, geography textbooks, and academic articles in a particular subject, lectures, casual conversations, or essays written by students. It aims to be representative of a given type of text.

As opposed to a specialized corpus, a general corpus is a corpus of many different types of texts, including written and/or spoken language. It may include texts produced in one or more particular languages, and it can include as many texts as possible to identify. A general corpus is usually much larger than a specialized corpus and is often used as a base-line in comparison with more specialized corpora.

In addition to specialized and general corpora, Hunston offers other types of corpora, such as comparable corpora, parallel corpora, learner corpora, pedagogic corpora, historical or diachronic corpora, and monitor corpora. (For further details, please refer to Hunston, 2005)

Corpus-based language teaching is gradually becoming mainstream, especially in ESP, as can be seen by the number of research papers written on the topic (Salager-Meyer, 1984, cited in Nguyen, 2010; Ward, 2009; Valipouri & Nassaji, 2013; Peter & Fernandez, 2013; Hou, 2014; Hsu, 2014; Lei & Liu, 2016). As will be discussed in the next section, the findings of these papers indicate that corpora, in particular frequency data, should be applied in language studies, if for no other reason than most research related to discipline-specific word lists focuses primarily on the frequency aspect of corpora.

2.3 Three categories of word lists

There are three categories of word lists for teaching English language learners. The three word lists are (1) the General Service List (GSL); (2) the Academic Word List (AWL); and (3) specialized or technical vocabulary (Hou, 2014, p.26).

1) The General Service List (GSL) is a list of roughly 2,000 words collected by West (1953, cited in Hou, 2014). Words were collected from a corpus of written English for teaching English language learners and teachers. There have been two major updates of the GSL. The first, revised by Browne, Culligan & Phillips (2013), is called the NGSL (The New General Service List), and another, revised by Brezina & Gablasova (2015), is known as the new-GSL (the New General Service List).

2) The Academic Word List (AWL) was developed by Coxhead (2000), and it contains 570 word families that account for 10.0% of the total words (tokens) in academic texts. The AWL was developed from different academic journals and university textbooks in 28 disciplines from four main areas: the arts, commerce, law, and natural science. The AWL collected words which are mostly useful for learners studying with high academic goals. The list also provides a useful basis for further research into the nature of academic vocabulary.

3) Specialized or technical vocabulary is the vocabulary used in certain fields. They are one of the three categories of word lists called the Outside Word List (OWL) (Panjanon & Soranasathaporn, 2016; Coehead & Hirth, 2017). Specialized vocabulary are words that have specific meanings within a specific field of expertise and can also lead to massiveness of text. They are called technical words by Nation (2001) and cover about 5% of the running words in specialized texts, occurring frequently in a specialized text or subject area, but not occurring, or of

very low frequency, in other fields (Nation, 2001, pp. 18-19). These characteristics are typically difficult for students who are non-native speakers to understand (Coxhead, 2018).

Technical words are also divided into four sub-types by their degrees of technicalness (Cowan, 1974; Nation, 2001, cited in Menon & Mukundan, 2010). The four sub-types are described below.

1) Highly technical words – they are found rarely outside their specific field.

2) Sub-technical words – they appear with high frequency across disciplines and meanings relate to the specific field and are also understood outside the field, such as the word 'memory' in the computing field (Nation, 2001, p.199).

3) Semi-technical words – they have one or more general English language meanings and in technical contexts take on extended meanings.

4) Non-technical words – they are quite common words which have little specialization of meaning, for example 'hospital' and 'judge'.

As seen above, specialized vocabularies are one of the three main word lists, and they are also divided into sub-groups by technicalness. In this paper, the sub-groups categorized by technicalness are not the main objective, but the concept of technicalness is useful when considering its individual characteristics. The characteristic that it is used outside the GSL and the AWL is considered and emphasized in order to establish word lists for industrial works.

2.4 Corpus-based vocabulary research for discipline-specific purposes

In this section, previous researches can be divided into two types of content relating to research on word lists in science and words, collocation, and phraseology in science.

Nowadays, there are many previous studies relating to word lists in science fields. In the field of engineering, Ward (2009) and Hsu (2014) developed the Basic Engineering List (BEL) and Engineering English Word List (EEWL) to serve individual aims such as the studies of first-year engineering students and for covering students' lexical shortage. Other studies on word lists – such as Chemistry Academic Word List (CAWL), a specialized vocabulary word list of 113 words in a composition culinary course, in–house specialized corpora on wine tastings, and a new medical academic vocabulary list (MAVL) (Valipouri & Nassaji, 2013; M. Nordin et al. (2013); Hou, 2014; Lei & Liu, 2016) – are also typical of research on the word lists of sciences.

Apart from the above, there are several other research articles relating to word lists in the fields of social science. Moini & Islamizadeh (2008) conducted a study of the 1,263 most frequently used academic words in linguistics, in order to develop the Linguistics Academic Word List (LAWL) divided into four main sub-disciplines: phonology, morphology, syntax, and semantics. Regarding

research on word lists in the Thai context, Chanasattru & Tangkiengsirisin (2016) developed a high frequency content word list discovered in social science research articles, called the Social Science Word List (SSWL) for undergraduate and graduate students in the fields of social science. The SSWL contains 394 content words from 64 open-access English research papers, and all 394 content words can be found in the NGSL (New General Service List) and the AWL (Academic Word List).

In addition to research on word lists, there are many studies relating to frequently-appearing words and their collocations. Panjanon & Soranasathaporn (2016) studied collocations using the high-frequency content words of the Outside Word List (OWL) from 100 research articles published in the Institute of Laboratory Animal Research (ILAR) journal between 2010 and 2014. A laboratory animal science word list was created and the most frequently-appearing word, *zebrafish*, and its lexical collocation were discussed, while Salager-Meyer (1984, cited in Nguyen, 2010, p.22) studied the complex nominal phrases in English scientific-technical texts in comparison with general English, in terms of frequency of occurrence, average length and proportion.

There are also studies about the length of technical words. Salager-Meyer (1984, cited in Nguyen, 2010, p.22) found that the average length of complex nominal phrases in technical English was at 2.61 whereas in medical texts it was 2.55. Justeson & Stava (1995) explored different aspects of technical terminology, focusing on their linguistic properties, as well as developing an algorithm for identification in a test. The structure of technical terms could be summarized by length, which was found to be between 1 and 4 or more than 4. The majority of terms appeared to prefer phrase structures or noun phrases containing adjectives, nouns, and occasionally, prepositions. Rarely did these terms contain verbs, adverbs, or conjunctions.

The employment of corpora is widespread in English language teaching, though the target of a specific corpus depends on the aim of the research associated with it. The corpus is particularly promoted in English for specific purposes, where the use of authentic materials is fundamental. Previous researchers mostly aimed to create specialized vocabularies found in academic papers and textbooks. As mentioned above, Valipouri & Nassaji (2013) analyzed a corpus of 1,185 specialized vocabularies from chemistry research articles. Ward (2009) designed word lists for engineering students by collecting 25 textbooks in related fields of engineering. Hsu (2014) also created the EEWL based on 100 college textbooks across 20 engineering subfields. Panjanon & Soranasathaporn (2016) cross-referenced the Outside Word List (OWL) with 100 research articles published in the Institute of Laboratory Animal Research (ILAR) journal.

Though the results of these research papers have been considered useful for ESP or EAP classes, the word lists themselves may not reflect the words used in the workplace if for no other

reason than these lists were constructed from textbooks and academic papers as opposed to authentic sources. Of the articles that were referenced in the current study, only Hou (2014) appears to have used authentic materials. Hou (2014) studied from authentic materials, such as wine tasting notes and information available on the official website of the Liquor Control Board of Ontario, in Canada. The current study has attempted to follow the example set by Hou to create specialized vocabularies based on the authentic materials that are currently in use in various industrial sectors. In summary, the current study will attempt to offer discipline-specific word lists in industrial fields for enhancing students' English performance so that they may be ready for the Eastern Economic Corridor (EEC) (the EEC is a special economic zone of three provinces in Eastern Thailand). The construction of the corpus is based on authentic materials from two sub-fields of the 10 S- Curve industries, Automotive and Automation, and three departments of the manufacturing process: production, technology and quality control. Thailand is the largest automotive producer in Southeast Asia and 12th in the world (Thailand Board of Investment, 2017), and in order to accelerate its next-generation automotive industry as one of the S-Curve industries, one system, robotic automation, is in the lead. Automation technologies tend to be more and more significant in the automotive industry (Top 4 Automation Technologies Used in Automotive Industry, 2020). Besides the core knowledge of industrial work (Automotive and Automation), English proficiency is also required of office workers (The Bureau of Academic Affairs, 2018, p.14). This is why they need to know the technical words in each organizational structure, such as production, technology, and quality control.

2.5 Word Formation for ESP

Morphology is the component of grammar that is devoted to the study of the internal structure of words (Culpeper et al., 2009, p.77). Morphology is the relevant discipline to deal with the increase of new words in a language as a result of the rapid development in the fields of science and technology. This is due to the number of new terms in various fields of science and technology that have been invented to explain new concepts and ideas in these fields. This is true not only in the science and technology fields, but in daily life as well. For example, the word *selfie* is a relatively new term used to describe a photograph that people take of themselves, usually with a mobile phone. *Selfie* has become one of the most frequently used words in social media because of the proliferation of camera-equipped mobile phones.

In morphology, the smallest units which have meanings or possess grammatical functions are called morphemes. Morphemes are divided into free morphemes and bound morphemes. The presence of only one free morpheme can be considered a word. When at least two free morphemes

or two words combine, compounds are formed. This process is known as compounding. The focus of discussion was on compounding since, with regard to Crystal (1997, p.384 cited in Nagy, 2014, p.265.), compounds are one of the main structures in 'science specific grammar'. He observes that most technical vocabulary is in the form of compounds. According to Culpeper et al. (2009, p.102), compounds are sub-classified as noun compounds, verb compounds, and adjective compounds. The following are some examples taken from Culpeper et al. (2009, p.102):

Noun-compounds

N + N [school]$_N$ + [boy]$_N$	[schoolboy]$_N$
A + N [wet]$_A$ + [lands]$_N$	[wetlands]$_N$

Verb-compounds

N + V [house]$_N$ + [train]$_V$	[housetrain]$_V$
Prep + V [under]$_{Prep}$ + [sell]$_V$	[undersell]$_V$

Adjective-compounds

Adj + Adj [yellow]$_{Adj}$ + [green]$_{Adj}$	[yellow green]$_{Adj}$
N + A [razor]$_N$ + [sharp]$_{Adj}$	[razor-sharp]$_{Adj}$

Wu (2014) insisted on the significance of word formation knowledge for ESP learners. Word formation should be taught to help learners memorize and be able to recall vocabulary more effectively.

> Word formation is a crucial strategy for learners to memorize related words. The knowledge of basic affixes, roots, suffixes, and word formation enables learners to decode them. Decomposing an unknown word into its structural components to infer its meaning is called structural analysis. ESP vocabulary learners should be taught the basic strategies of decomposing a word into parts to guess its meaning. Several studies have reported using various affixes to teach students to structurally analyze ESP vocabulary. The instruction resulted in large gains in student abilities to infer word meanings.
>
> (Wu, 2014, p.123)

Crystal (1997, p.384 cited in Nagy, 2014, p.265) proposed the structures of science-specific grammar by describing that the large technical vocabulary, apart from being based on Latin or Greek words, is in the form of compounds and also long sentences with complex internal structures. Besides, he insisted that science is the major source of the formation of new words.

In this paper, compound words are identified from the specialized vocabularies gathered from authentic industrial materials. The reason that compounds are the only form being identified is because compounds are the main structure of technical vocabulary (Crystal, 1997, p.384 cited in Nagy, 2014, p.265). Additionally, noun compounds are selected from these authentic materials as part of a discussion on methods for instruction in ESP classes.

3. Research Questions

1) What are the most frequently-occurring specialized vocabularies in authentic materials produced for the industrial sectors?

2) How does the vocabulary used in authentic materials compare to a reference corpus (English for Science and technology textbooks). What is the proportion of specialized vocabularies in industrial sectors when compared to the vocabularies in English for Science and technology textbooks?

3) What form of specialized vocabularies are most commonly found in authentic materials in industries?

4. Methodology

4.1 Procedure

The study is composed of specialized vocabularies in two sub-fields of the 10 S-Curve industries, Automotive and Automation, and three departments of a manufacturing process. The procedure included the following steps:

1) Creating specialized vocabulary corpora based on authentic materials from industrial sectors around EEC, using software called AntConc.

2) Making lists of specialized vocabularies in industrial sectors (Industrial Word Lists: IWL), with their ranking in the lists based on their frequency (this will help answer RQ1).

3) Investigating the characteristics of words used in English for Science and technology textbooks by comparing them to the reference corpus (authentic industrial materials) (this answered RQ2).

4) Identifying the word formation of specialized vocabularies and making pedagogical recommendations for ESP classes (this answered RQ3)

In this present study, function words, such as *out, to, up, the, on, off,* and *by*, were not analyzed, including the letters, *i, o, s,* and *d*. These letters, which are found frequently in authentic materials, tend to be used as symbols for identifying some pictures in the materials.

Function words were omitted from the IWL as they do not bear any special meaning.

4.2 Data Resources

To gather and analyze data, the following resources and tools were used.

4.2.1 The specialized vocabulary corpora

To compile the specialized vocabulary, authentic materials were collected from two sub-fields of the 10 S-Curve industries, Automotive and Automation, and three departments of the manufacturing process (briefly categorized under three sub-fields).

1) Automotive

2) Automation

3) Production, technology, and quality control

The materials for the automotive and automation industries were taken from manuals and authentically used documents from each industry located around the Eastern Economic Corridor (EEC), whereas the materials related to production, technology, and quality control come from an Interpreter Dictionary in Japanese, Thai and English (2015).

Table 1 Corpora construction of authentic materials

corpora	word types	word tokens
materials relating to automotive (=2199)	4082	19582
materials relating to automation (=4013)		
materials relating to production, technology, and quality control (=3080)		

4.2.2 The textbooks of English for science and technology

Two English for science and technology textbooks and one workbook were selected (all are the official textbooks used in English for science and technology classes at Pibulsongkram Rajabhat University) in order to identify specialized vocabularies and compare the characteristics of the words in the textbooks with the words appearing in authentic materials.

1) Technical English (course book) compiled by Bonamy (2014)

2) Technical English (workbook) compiled by Jacques (2014)

3) English for science and technology complied by Sirijaruwong & et al. (2010)

Three textbooks are used as the main resources for the core course, English for science and technology, for all industrial technology students in Pibulsongkram Rajabhat University.

Table 2 Corpora Construction of Textbooks

corpora	word types	word tokens
Three Textbooks	5245	44280

4.2.3 Analytical tool

AntConc is a tool used in research related to word frequency, concordance and comparative corpora analysis. This tool helps researchers navigate corpora and was developed by Anthony (2014). This program is available for free on the internet.

4.3 Word formations for analysis

To answer the research questions, word formations of specialized vocabularies were analyzed. The scope of analysis is based on the two following criteria: word length and compounding. Compounding was the focus for the present study since it is a common method of word formation for specialized language. As Crystal (1997, p.384 cited in Nagy, 2014, p.265) pointed out, most technical vocabulary is in the form of compounds. He further stated that science is considered to be a primary source of the formation of new words. Compounding is assumed to be a common word formation process, especially for terminology or technical vocabulary. Compounding, regarding Nakov (2013, p. 2), is considered a 'text compression device' since it makes long explanation shorter by reducing words. However, the core information still exists.

5. Findings and discussion

In order to build corpora of specialized vocabularies intended to be used in industrial sectors, three different examples of authentic materials (as seen in Table 1) were collected. The total number of word types is 4082 and there are 19582 word tokens. During the AntConc analysis, the word list program was used to generate specialized industrial vocabulary items. The criteria for selecting words for constructing the Industrial Word List (IWL) is that the frequency of word appearance must be at least 10 times. Finally, 439 of the most frequently-occurring word types, with many occurring at least 10 times in the Industrial Word List (IWL), were ultimately chosen, and the first 50 words of all subfields and each field were selected, as represented below.

Table 3 The first 50 most frequent specialized vocabulary items from *all fields.*

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	184	module	18	70	power	35	55	setting
2	160	control	19	69	brake	36	54	circuit
3	158	data	20	69	product	37	54	machine
4	133	system	21	69	production	38	54	program
5	148	valve	22	67	gauge	39	53	work
6	133	number	23	65	output	40	52	instruction
7	101	device	24	65	quality	41	51	limit
8	93	time	25	64	speed	42	49	fuel
9	85	engine	26	63	motor	43	49	remote
10	83	process	27	62	file	44	48	function
11	79	value	28	60	memory	45	48	inspection
12	76	switch	29	59	high	46	47	line
13	76	type	30	59	mode	47	46	link
14	75	operation	31	58	air	48	45	termial
15	75	test	32	58	pressure	49	44	screw
16	71	error	33	56	gear	50	43	bearing
17	71	input	34	55	oil			

These first 50 words are the specialized vocabularies found frequently in the previously identified industrial sectors. For the purposes of this research, the above words can be categorized into specialized vocabularies based on authentic materials used in present research. Many of the words in Table 3 found in the GSL and AWL. The proportions between the GSL, AWL, and OWL (specialized or technical terms) are as follows:

Table 4 Industrial Word List (IWL = 439 words)

OWL		GSL		AWL		GSL&AWL	
137	31.21%	222	50.57%	80	18.22%	2	0.46%

Two words found in both the GSL and AWL are *network* and *mechanism*. Most words are categorized under GSL, OWL and AWL, respectively. This means that specialized vocabularies in industrial sectors are a mixture of real technical terms used especially in industries, academic words (AWL) and general words (GSL). This includes:

1) Real technical terms: *module, valve, switch, brake, gauge, gear, circuit, instruction, fuel, remote, inspection, terminal, pulse, monitor, torque, relay, ignition, alarm, analog, clutch, voltage, cable, battery, cylinder, filter, menu, transmission, sampling, communication*

2) Words overlapping with the GSL: *control, system, number, time, engine, value, type, operation, test, power, product, production, quality, speed, motor, memory, high, air, pressure, oil, setting, machine, program, work, limit, line, screw, bearing, management, ring, current, signal, tool, position, spring, cost, drive, information, material, start, supply, check, list, resistance, steering, gas, angle, block, station, point, frequency, connection*

3) Words overlapping with the AWL: *data, device, process, error, input, output, file, mode, function, link, processing, area, status, code, parameter, ratio, cycle, contact*

The words that were not listed in the GSL and AWL accounted for 31.21% (137 words) of the 439 words.

All of the most frequently occurring specialized vocabularies appear in one word form, and the meanings are not hard for students to understand. Basically, students start with learning general words related to everyday life, such as food or directions, and then start to learn more specialized vocabularies. Most specialized vocabularies relate to the words commonly used in daily life. This can be seen from the number of words overlapping with the GSL (50.57%). The specialized vocabularies used in the industrial sectors investigated in this research could be taught in English for science and technology classes because most of the terms appear in short form and link to the general words.

Specialized vocabularies can be categorized into sub-fields as can be seen in the following tables:

Table 5 The top 50 most frequent specialized vocabulary items from the automotive industry

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	87	valve	18	24	**control**	35	16	drive
2	81	engine	19	22	***axle***	36	16	gauge
3	65	**system**	20	22	***power***	37	16	light
4	50	brake	21	22	wheel	38	15	pin
5	44	gear	22	20	piston	39	15	shaft
6	43	oil	23	20	pump	40	15	tyre
7	42	***air***	24	19	suspension	41	14	***cycle***
8	40	***fuel***	25	19	vacuum	42	14	head
9	39	ignition	26	18	idle	43	14	ratio

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
10	34	steering	27	18	rod	44	13	four
11	32	clutch	28	17	advance	45	13	lock
12	28	***bearing***	29	17	heat	46	13	***motor***
13	27	cylinder	30	17	plate	47	13	stroke
14	27	***pressure***	31	17	plug	48	13	***torque***
15	27	***spring***	32	17	tire	49	12	ball
16	25	***ring***	33	16	arm	50	12	cap
17	25	spark	34	16	back			

In Table 5, we can see the word types used in the automotive field. The most frequently appearing words are *valve, engine, brake, gear, oil, steering,* and *clutch*.

Table 6 The top 50 most frequent specialized vocabulary items from the automation industry

Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word
1	181	module	18	51	program	35	35	status
2	144	***data***	19	50	**system**	36	35	***terminal***
3	91	**control**	20	49	***motor***	37	34	relay
4	91	device	21	48	***speed***	38	34	***test***
5	68	error	22	47	instruction	39	33	area
6	68	input	23	47	remote	40	33	parameter
7	67	**number**	24	43	link	41	33	position
8	67	***switch***	25	43	pulse	42	32	menu
9	65	***type***	26	41	monitor	43	31	list
10	64	***time***	27	40	***power***	44	31	***processing***
11	64	***value***	28	39	***function***	45	31	start
12	62	output	29	39	limit	46	31	voltage
13	59	file	30	37	alarm	47	30	code
14	58	mode	31	37	analog	48	30	station
15	58	***operation***	32	37	information	49	29	communication
16	54	setting	33	37	signal	50	29	connection
17	53	memory	34	35	***circuit***			

In Table 6, we can see the word types used in the automation field. Some of the most frequently appearing words explain the manufacturing systems or production processes that are carried out automatically, including the following: *module, device, error, input, output, file,* and *mode*.

Table 7 The top 50 most frequent specialized vocabulary items from production, technology, and quality control

Rank	Freq	word	Rank	Freq	word	Rank	Freq	word
1	61	quality	18	21	**number**	35	14	***line***
2	58	product	19	19	***angle***	36	14	resistance
3	52	production	20	18	part	37	14	review
4	49	***process***	21	18	**system**	38	14	working
5	45	**control**	22	17	inventory	39	14	yarn
6	44	gauge	23	17	manufacturing	40	13	***bearing***
7	44	work	24	17	***operation***	41	13	***circuit***
8	41	machine	25	17	sampling	42	13	defect
9	36	*test*	26	17	tool	43	13	design
10	34	cost	27	16	capacity	44	13	***gas***
11	34	inspection	28	16	machining	45	13	maintenance
12	31	management	29	16	plan	46	13	pipe
13	28	material	30	15	analysis	47	12	***air***
14	27	***time***	31	15	***pressure***	48	12	***frequency***
15	26	steel	32	15	rate	49	12	**screw**
16	23	equipment	33	14	***data***	50	12	***supply***
17	23	**high**	34	14	factory			

From Tables 5-7, one point can be seen. The underlined and italicized words are the words found most often in the three previously mentioned sub-fields. Some of these terms appear in two sub-fields (the italicized words) while others appear in all sub-fields (the underlined words).

In Table 7, we can see the word type used in production, technology and quality control. The words that appear most frequently are the following: *quality, product, production, gauge, work* and *machine*.

Finally, it can be concluded that 439 words of the IWL are a mixture of the OWL, GSL, and AWL. This differs from the CAWL (Chemistry Academic Word List) in Valipouri & Nassaji (2013) that has not been listed in the GSL and AWL. The IWL also has different criteria to the result in Ward's Engineer Corpus (EC). In the EC, only 10% of the running words are from the AWL whereas in the IWL, 18.22% of the running words are from the AWL.

To address the results of RQ2, the keyword list program in AntConc was used as a statistical analysis tool. Keywords are words which appear at a significant frequency in a corpus when compared to other words. During this process, the target corpus is based on the words appearing in the authentic materials, while the reference corpus is comprised of words taken from the English for science and technology textbooks. As a result, it can be said that most of the words in the target corpus are words which appear in the Outside Word List, while most of the words in the English for science and technology textbooks, as seen from the statistical results, are also not specialized vocabulary items. The statistical measure was defaulted by the setting of Log Likelihood. The significance value, 99th percentile; 1% level; $p< 0.01$; critical value = 6.63, was applied for assessing the keyness of the target file words in AntConc. It means that the difference between the two corpora is highly significant. The results show that there are 1038 keyword types [including positive and negative keyness, which means overuse or underuse of the target corpus relative to the reference corpus (Lancaster University, n.d.)]. The positive keyness has 926 keyword types (including all appearing-frequency) from 26220 word tokens. The representative words from the target corpus, ranked from 1-50 [ranking criteria is Keyness (LL2)], are shown as follows:

> *module, data, control, error, device, value, file, gauge, input, mode, operation, production, program, memory, output, quality, limit, setting, system, number, processing, brake, pulse, instruction, product, management, remote, test, type, time, analog, bearing, terminal, clutch, status, ring, inspection, parameter, ratio, gear, screw, torque, angle, valve, motor, process, sampling, relay, code, mounting*

By showing the negative keyword option, words that are unusually infrequent in the target corpus compared with the reference corpus were also displayed. The target corpus revealed that negative keywords appear with unusually low frequency when compared with the frequency at which they appear in the reference corpus. It found 113 keyword types and 26220 tokens from the present study. Some are shown below:

do, how, words, use, water, complete, not, make, used, take, same, go, computer, box, fly, outlet, tank, all, each, large, rotate, down, find, groups, kg, refer, wheels, truck, match, ask, printer, different, plastics, cause, push, rise, heavy, new, bend, robot, electricity, form, radio, cool, description, parts, people, year, per, reading, workshop, wind, pull

The words above did not appear frequently in the target corpus or the authentic materials. By the way, it could be said that the above negative keywords are frequently appearing words in the English for science and technology textbooks. From this, it can be implied that in the English for science and technology textbooks, though they are specific textbooks for specific fields, a large proportion of the vocabulary is general, not specific.

This means that although the textbooks were written for the purpose of teaching students majoring in science and technology, the emphasis was on non- specialized vocabularies. Owing to a shortage of specialized vocabularies in textbooks used in English for science and Technology, students majoring in these fields might encounter these vocabularies only after graduating and entering the workforce. In other words, they may not be equipped with an understanding of their meanings, nor of how to use them.

Additionally, the negative keywords were run against the authentic materials used in this study to confirm the coverage. This yielded coverage results of 89.21% (926: 113) of word types. This also implied that the authentic materials studied possess a lot of specialized vocabularies.

To find the answer for the third research question (RQ3), the length and compounding of specialized vocabulary were analyzed.

Table 8 The length of specialized vocabularies

Length	vocabulary amounts (9,292)						Total	
	automotive (2,199)	100 %	automation (4,013)	100 %	production, technology, quality control (3,080)	100 %	9,292	100 %
1	706	32.1	850	21.2	1,440	46.8	2,996	32.2
2	1,085	49.3	1,969	49.1	1,286	41.8	4,340	46.7
3	352	16.0	831	20.7	293	9.5	1,476	15.9
≥4	56	2.5	363	9.0	61	2.0	480	5.2
			Length Average		(1x2,996)+(2x4,340)+(3x1,476)+(4x480)			
					2,996+4,340+1,476+480			
					= 1.94			

The average length of specialized vocabularies in industrial authentic materials is 1.94, whereas in English scientific-technical texts, the average length is 2.61. In medical texts, the average length is 2.55, as stated by Salager-Meyer (1984, cited in Nguyen, 2010, p.22). The results appear to imply that the technical terms used in the industrial fields are not too long when compared to other fields, such as English scientific-technical texts and medical texts.

From the results in Table 8, we can see specialized vocabularies in the form called a *compound*. Based on the length of words above, two-word compounds appeared in 4,340 words, or 41.8% of the words investigated. This is the highest proportion of all specialized vocabularies. It can be said that two-word compounds are most commonly found in authentic materials. Two-word compounds will be mainly discussed as follows.

The word formation process is divided into inflection and derivation. Ward (2009, p.175) pointed out the difficulty of inflected forms of words for Thai learners who have a very limited knowledge of inflections. The inflected forms are inflectional suffixes (-s affix), inflected -ed and -ing forms, and the different distributional profiles of each inflectional form. Additionally, Wu (2014, p.123) discussed the importance of understanding how words are formed, which will enable

students to understand the meaning of specialized vocabularies. In the case of noun-compounds, they can be divided based on word structure, such as *noun + noun*, *adjective + noun*, and *verb + noun*. Most of these words follow the *noun + noun* word structure. This appears to agree with Justeson & Stava (1995) who identified that the specialized vocabularies or terminology prefer noun phrases containing nouns, adjectives and prepositions. In contrast, it is rare to find verbs in noun phrases. As a result, in this present study, there were two-word compounds including verbs as modifiers. The verb + noun structures were found in *-ing + noun* and *-ed + noun* forms. The *-ing + noun* form was found in 297 words, whereas the *-ed + noun* form was found in 105 words.

When -ing forms are used as modifiers, it can happen both with a noun-like -ing form (gerund) or an adjective -like -ing form (participle). The two structures have different kinds of meaning. In this present study, the separation of a gerund and participle is not the key issue, as stated by Swan (1995, p.277) that the distinction is not really as simple; some grammarians prefer to avoid the terms participle and gerund. The meaning of *-ing + noun* could be divided in two ways. Typically, *-ing + noun* compounds indicate the function of something or the ongoing quality or state of something, such as the *waiting room* that means a room for waiting and the *waiting train* that means a train that is waiting. The following *-ing + noun* compounds occurred with *the -ing* that appears more than 10 times which are *working, machining, sampling,* and *mounting*.

Table 9 -ing + noun compounds beginning with *working* and *machining*

-ing **+ noun compounds beginning with *working***	**freq.**	***-ing*** **+ noun compounds beginning with *machining***	**freq.**
working day	2	machining design	2
working process	1	machining process	1
working area	1	machining cost	1
working ratio	1	machining time	1
working posture	1	machining sequence	1
working condition	1	machining accuracy	1
working speed	1	machining data	1
working standard	1	machining expense	1
working expenses	1	machining defect	1
working hours	1	machining center	1
working group	1	**Total**	**11**
working point	1		
Total	**13**		

Table 10 -ing + noun compounds beginning with *sampling* and *mounting*

ing + noun compounds beginning with sampling	**freq.**	**ing + noun compounds beginning with mounting**	**freq.**
sampling inspection	2	mounting screw	2
sampling frequency	2	mounting hole	1
sampling plan	1	mounting bracket	1
sampling check	1	mounting environment	1
sampling times	1	mounting stuff	1
sampling period	1	mounting tab	1
sampling value	1	mounting pitch	1
sampling trace	1	mounting direction	1
sampling processing	1	mounting surface	1
total	**11**	mounting rail	1
		total	**11**

Here are *some -ing* + noun compounds with *-ing* that have an appearance frequency of less than 10 times.

stirling engine, bearing heater, braking system, timing mark, jacking point, metering unit, bearing shell, warning light, wiring harness, cranking motor, milling machin

Here are some of their meanings taken from the Lexico Dictionary (n.d.).

Stirling engine: a machine used to provide power or refrigeration, operating on a closed cycle in which a working fluid is cyclically compressed and expanded at different temperatures

Working point: the point or place in a machine at which the work or task it is used for is performed (now rare). A point (as on a graph or chart) corresponding to satisfactory working

Machining center: A machine tool which (originally under numerical control and now under computer-numerical control) is capable of performing automatically, any or all of the functions of storing, selecting, and changing the various cutting tools required at different stages of a manufacturing process

Based on the results, all *-ing + noun* compounds in specialized vocabularies could be categorized into the indication of function or naming something. In conclusion, -ing forms in the specialized vocabularies are forms of modifiers that are used to modify the head nouns and are formed as new words.

By the way, *-ed + noun* compounds also indicate an existing state or a completed state. Here are some *-ed + noun* compounds found in this study.

The following *-ed + noun* compounds occurred with *-ed* that appears more than 8 times which is *fixed*.

Table 11 -ed + noun compounds beginning with *fixed*

ed + noun compounds beginning with *fixed*	**freq.**
fixed bracket	1
fixed temperature	1
fixed cost	1
fixed value	1
fixed length	1
fixed ratio	1
fixed point	1
fixed distance	1
total	**8**

Here are *some -ed* + noun compounds with *-ed* that appear less than 8 times.

forced termination, planned result, machined surface, graduated pipette, accumulated weight, distilled water, discontinued model, threaded rod, dotted line, finished product, staggered tooth, completed car, bored pile, biased sampling

Here are some of their meanings taken from the Automotive Dictionary (n.d.), the Lexico Dictionary (n.d.) and the Cambridge Dictionary (n.d.).

Machined surface*:* a smooth surface of metal such as the top of a cylinder block (Automotive Dictionary)

Dotted line*:* a line made up of dots or dashes (often used in reference to the space left for a signature on a contract); an indirect, informal, or secondary line of responsibility within an organization (as represented by a dotted line on an organizational chart)

Distilled water*:* water that has been made purer by being heated until it becomes a gas and then cooled until it becomes a liquid again

Each compound has only one meaning related to a tool, a process, or a state of something in industry. This is similar to Aleson (2013, cited in Nagy, 2014, p.268) who defines

specialized language as a specialized lexicon that has only one meaning. These -ed + noun and -ing + noun compounds should be taught to learners, with emphasis placed on how they are formed and how their meanings differ from the meanings they possess in daily life. For example, the meaning of *bored pile* does not describe a pile that feels bored, but instead describes a pile that has been cast on a construction site. This is because the word "bore" carries the meaning of making a hole in something using a tool. In daily usage, many verbs ending in -ed (used as past participles) or -ing (used as the present participle) can be used passively or to convey feelings, such as when they are used like adjectives or adverbs (Swan, 1995, p.402). All *-ing + noun* compounds and *-ed + noun* compounds tend to confuse learners if they do not understand word formations or the idea of univocity.

6. Conclusion and pedagogical implications

The industrial words, called Industrial Word Lists, found in this paper include 439 words selected by having appeared more than 10 times. Most of the IWL terms are in the GSL, OWL and AWL.

As the results have shown above, it can be said that in English for science and technology textbooks, specialized vocabularies are found less frequently than they are in authentic materials. Although some specialized vocabularies appear in these textbooks, their appearance is miniscule when compared to the frequency at which they appear in authentic materials.

Teaching specialized vocabularies needs to be conducted with a greater emphasis placed on both the usage and the context of specialized vocabularies, and English for science and technology textbooks should reflect this need. The specialized vocabularies identified in this research can and should be included in textbooks. Furthermore, teaching the collocates of specialized vocabularies is highly recommended. For example, *module,* a word that is in the most frequently-occurring term in the IWL, means a part of a machine, and *module* collocates with adjectives like *command-, lunar-, memory-, RAM-, software-* and with verbs like *add, configure, install* and *load*. These should be learned in context rather than in isolation.

Another technique that can be applied to English for science and technology classes is to divide specialized vocabularies, collocates, and context into sub-fields relating to a specific field of study. For example, in the case of all industrial technology course students, the specialized vocabularies in Table 3 can be included in instruction. Specialized vocabularies from authentic materials are necessary to learn because of the impact they will have on the future careers of students preparing to enter into industrial fields.

The specialized vocabularies from this present research can also be applied to classes after the students have reached a basic 2000-word vocabulary offered in Coday & Huckin (1997, cited in Wu, 2014, p.122). The curriculum of English learning in Pibulsongkram Rajabhat University is similar to this. In other words, after learning four courses of English language from the General Education courses, science and technology students will register for English for science and Technology. As mentioned above, the 100 specialized vocabularies, including compound nouns and knowledge of -ed and -ing modifiers and other word formation processes, can be brought into classes for enhancing students' comprehension of specialized vocabularies. Although there are students in various majors who might not find IWL useful in English for science and technology classes, as seen from Table 3 and the Appendix, the technicalness of the IWL tends to fall to sub-technical words, semi-technical words, and non-technical words, not highly technical words. This relates to the proportions between the GSL, AWL, and OWL (Table 4) in the IWL. For this reason, teachers might first give an explanation of what the IWL is and what its characteristics are, and then, enhance students' learning motivation by adopting the IWL in various practical forms or mixing it with other content for classes. Finally, this research is limited by the amount of authentic materials because most industries do not allow for the publishing of their documents. As a result, for future research, a larger scale of word tokens from authentic materials and more various specific fields are necessary for further studies due to the appropriate-sized vocabulary depending on the technical vocabulary knowledge requirements of a given profession or industry. All results from the present study can also be applied for improving English for science and technology textbooks.

References

Anthony, L. (2014). *AntConc (Version 3.5.8) [Computer Software]*. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net

Aroonmanakun, W. (2010). *Corpus linguistics: Theory and usage.* (2nd edition). Bangkok: Faculty of Arts. (in Thai)

Automotive Dictionary. (n.d.). Retrieved from http://www.automotivedictionary.org/

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: Michigan Publishing.

Bonamy, D. (2014). *Technical English (course book).* (9th edition). England: Pearson Education Limited.

Boulton, A. & Tyne, H. (2013). Corpus linguistics and data-driven learning: A critical overview. *Bulletin suisse de linguistique appliquée*, *97*, 97-118.

Brezina, V. & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36*(1), 1-22.

Browne, C., Culligan, B. & Phillips, J. (2013). *New general service list*. Retrieved from http://www.new generalservicelist.org/

Cambridge Dictionary. (n.d.). Retrieved from https://dictionary.cambridge.org/us/dictionary/english/

Chanasattru, S. & Tangkiengsirisin, S. (2016). Developing of high frequency word list in social sciences. *Journal of English Studies, 11*, 41-87.

Chujo, K. & Oghigian, K. (2009). How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary. *The Journal of Asia TEFL, 6*(2), 121-148.

Coxhead, A. & Hirth, D. (2017). A pilot science-specific word list. *Revue Française de Linguistique Appliqueé xii, 2*, 65-78.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A. (2018). *Specialized vocabulary in English for academic purposes in research and teaching*. Retrieved from https://www.youtube.com/watch?v=_w2WYFueUyg&feature=youtu.be

Culpeper, J., Katamba, F., Kerswill, P., Wodak, R., & McEnery, T. (2009). *English language: Description, variation and context*. UK: Palgrave Macmillan.

Hou, H-I. (2014). Teaching specialized vocabulary by integrating a corpus-based approach: Implications for ESP course design at the university level. *English Language Teaching, 7*(5), 26-37.

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

Hunston, S. (2005). *Corpora in applied linguistics.* (3rd edition). UK: Cambridge University Press.

Ibrahim, M. A. (2015). Vocabulary input in English language teaching: Assessing the vocabulary load in spine five. *International Journal of English Language and Linguistics Research, 3*(1), 1-14.

Interpreter Dictionary in Japanese Thai English. (2015). Bangkok: TPA Press. (in Thai)

Jacques, C. (2014). *Technical English (workbook).* (7th edition). England: Pearson Education Limited.

Justeson, S. J. & Stava, M. K. (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering, 1*(1), 9-27.

Lancaster University. (n.d.). *Log-likelihood and effect size calculator*. Retrieved from http://ucrel.lancs.ac.uk/llwizard.html

Lei, L. & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes, 22*(2), 42-53.

Lexico Dictionary. (n.d.). Retrieved from https://www.lexico.com/

M. Nordin, N. R., Stapa, S. H., & Darus, S. (2013). Developing a specialized vocabulary word list in a composition culinary course through lecture notes. *Advances in Language and Literary Studies, 4*(1), 1-11.

Menon, S. & Mukundan, J. (2010). Analysing collocational patterns of semi-technical words in science textbooks. *Pertanika J. Soc. Sci. & Hum, 18*(2), 214-258.

Moini, R. & Islamizadeh, Z. (2008). Do we need discipline-specific academic word lists? Linguistics Academic Word List (LAWL). *Journal of Teaching language Skills (JTLS), 35*(3), 65-90.

Nagy, I. K. (2014). English for special purposes: Specialized language and problems of terminology. *Acta Universitatis Sapieniae, Philologica, 6*(2), 261-273.

Nakov, P. (2013). On the interpretation of noun compounds: Syntax, semantics, entailment. *Natural Language Engineering, 1*(1), 1-40.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nguyen, M. L. (2010). *Noun-Noun combinations in technical English* (master's thesis). Suranaree University of Technology, Nakhon Ratchasima.

Panjanon, V. & Soranasathaporn, S. (2016). A corpus-based study on high-frequency content words and collocations in the Outside Word List (OWL) in laboratory animal research articles. *NIDA Journal of Language and Communication, 21*(29), 1-20.

Peters, P., & Fernández, T. (2013). The lexical needs of ESP students in a professional field. *English for Specific Purposes, 32*(4), 236-247.

Promwatcharanon, K. & Chatreepinyo, P. (2018). An analysis of challenges and requirements of using communicative English for developing a self-access learning package for traffic police officers in upper northern Thailand. *Humanities & Social Sciences, 35*(1), 163-195.

Saengpakdeejit, R. (2014). Awareness of vocabulary learning strategies among EFL students in Khon Kaen University. *Theory and Practice in Language Studies, 4*(6), 1101-1108.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Sirijaruwong, P., Khajornchaikul, K., Krusong, S., Thongrit, A., Sai-ngern, S. & Wongworapanit, P. (2010). *English for Science and Technology.* (3rd edition). Bangkok: Triple Group Co., Ltd.

Suwanwela, C. (2019). English language acquisition problems and strategies for revolution. In *The seminar of English proficiency of Thai people*. Bangkok: Thailand Science Research and Innovation.

Swan, M. (1995). *Practical English usage.* (2nd edition). Oxford: Oxford University Press.

Thailand Board of Investment. (2017). *Thailand' s automotive industry the next generation*. Retrieved from https://www.boi.go.th/index.php?page=business_opportunities_detail&topic_id=117517

The Bureau of Academic Affairs (The Secretariat of the House of Representatives). (2018). *The needs for Thai labor in EEC (2018-2027).* Retrieved from https://library2.parliament.go.th/ejournal/content_af/2561/aug2561-1.pdf

Top 4 automation technologies used in automotive industry. (2020). Retrieved from https://www.plantautomation-technology.com/articles/top-automation-technologies-used-in-automotive-industry

Valipouri, L. & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248-263.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineer undergraduates. *English for Specific Purposes, 28*(3), 170-182.

Wu, L-F. (2014). Motivating college students' learning english for specific purposes courses through corpus building. *English Language Teaching, 7*(6), 120-127.

Appendix (IWL No. 51-439, referring to *Table* 3)

pulse/ management/ monitor/ ring/ current/ processing/ signal/ torque/ relay/ tool/ ignition/ position/ spring/ alarm/ analog/ area/ cost/ drive/ information/ material/ clutch/ start/ supply/ voltage/ cable/ check/ code/ parameter/ ratio/ cycle/ list/ method/ resistance/ steering/ battery/ cylinder/ filter/ gas/ angle/ block/ menu/ station/ point/ transmission/ contact/ frequency/ communication/ connection/ display/ equipment/ manual/ mounting/ no/ order/ pin/ rate/ safety/ wire/ capacity/ converter/ heat/ spark/ specification/ steel/ user/ base/ buffer/ controller/ flow/ level/ part/ stop/ bit/ inventory/ noise/ open/ servo/ center/ count/ execution/ light/ logging/ main/ model/ pump/ cup/ diagram/ plate/ range/ shaft/ standard/ stroke/ vacuum/ wheel/ common/ external/ load/ maintenance/ network/ piston/ axle/ connector/ counter/ detection/ end/ lock/ plug/ refresh/ register/ reset/ rod/ selection/ change/ command/ idle/ laser/ lead/ lower/ multiple/ scan/ screen/ source/ suspension/ arm/ back/ digital/ head/ master/ measurement/ name/ serial/ temperature/ unit/ address/

ball/ design/ electronic/ latch/ low/ switching/ advance/ backup/ bar/ cam/ cutting/ defect/ electric/ gain/ interface/ loop/ machining/ plan/ request/ size/ tube/ water/ assembly/ chart/ coil/ factory/ failure/ frame/ hot/ injection/ intelligent/ manufacturing/ metal/ meter/ pipe/ sensor/ timing/ tire/ total/ trace/ transfer/ zero/ analysis/ automatic/ carbon/ constant/ cross/ curve/ delay/ local/ nut/ phase/ self/ upper/ weight/ working/ auto/ capacitor/ direct/ double/ encoder/ fixed/ impact/ joint/ maximum/ pattern/ procedure/ project/ release/ review/ trigger/ tyre/ welding/ ac/ access/ axis/ board/ bolt/ box/ cap/ character/ clear/ condition/ construction/ distance/ friction/ fuse/ hole/ inverter/ operating/ password/ precision/ programmable/ return/ running/ scale/ side/ surface/ technical/ thread/ top/ yarn/ beam/ card/ clearance/ comment/ confirmation/ dead/ development/ drawing/ dynamic/ element/ initial/ life/ offset/ points/ programming/ protocol/ response/ sample/ sequence/ service/ set/ shift/ software/ special/ tag/ tester/ thermal/ variable/ width/ wrench/ write/ actual/ adjustment/ belt/ coating/ combustion/ compression/ conversion/ core/ cover/ delivery/ distributor/ dual/ efficiency/ energy/ feed/ flat/ horsepower/ hydraulic/ internal/ lathe/ length/ mechanism/ minimum/ negative/ option/ radius/ read/ ready/ resistor/ reverse/ silicon/ step/ storage/ timer/ alloy/ ascii/ batch/ braking/ breaker/ bus/ compound/ configuration/ coupling/ crimping/ deviation/ free/ glass/ index/ industrial/ lamp/ language/ linear/ logic/ monitoring/ paper/ parallel/ pinion/ plastic/ requirement/ run/ sheet/ small/ space/ stock/ strength/ target/ throttle/ tolerance/ window/ wiring/ acceleration/ accuracy/ bottom/ caliper/ car/ carburetor/ component/ continuous/ cutter/ differential/ direction/ discharge/ distribution/ drop/ field/ front/ goods/ hand/ induction/ insertion/ item/ jig/ key/ log/ measuring/ milling/ mixture/ motion/ overhead/ panel/ period/ port/ positive/ radiator/ random/ raw/ reduction/ registration/ rolling/ semi/ short/ slip/ socket/ surge/ synthetic/ thermometer/ torsion/ transistor/ vertical/ word