

# คำศัพท์แก่นร่วมภาษาอังกฤษที่มีความถี่สูง สำหรับนักศึกษาวิศวกรรมศาสตร์ใน มหาวิทยาลัยไทย

High Frequency English Common Core Vocabulary  
for Thai University Engineering Students

สมจินต์ เชี่ยวชาญ\*

Somjin Chiewchan\*\*

---

\* นักศึกษาปริญญาเอกสาขาวิชาภาษาศาสตร์ มหาวิทยาลัยราชภัฏสวนสุนันทา

\*\* Ph.D. Candidate (Linguistics), Suan Sunandha Rajabhat University  
Email : jchiewchan@yahoo.com

## บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ (1) ตรวจสอบหาจำนวนและความถี่ในการปรากฏใช้ของ คำศัพท์ทั่วไปที่ใช้มาก 1000 คำแรก รวมคำบอกหน้าที่ และ 1000 คำหลัง รวมทั้งคำศัพท์วิชาการ 570 คำแรก ในตำราภาษาอังกฤษด้านวิศวกรรมศาสตร์ 9 เล่มที่มีการใช้สูง (2) สร้างรายการคำศัพท์ ทั่วไป และรายการคำศัพท์วิชาการจากคลังข้อมูลที่เก็บจากตำราวิศวกรรมศาสตร์ 9 เล่ม อิงคำศัพท์ ทั่วไปที่ใช้มาก 2000 คำแรก และคำศัพท์วิชาการ 570 คำแรก (3) ค้นหาความต่างในการแจกแจง หมวดคำของคำศัพท์ประเภทเนื้อหาในรายการคำศัพท์วิชาการที่สร้างจากคลังข้อมูลที่เก็บจากตำรา วิศวกรรมศาสตร์ 9 เล่ม (4) จัดทำรายการคำศัพท์แกนร่วมภาษาอังกฤษที่มีความถี่ในการปรากฏใช้ สูงสำหรับอ่านตำราวิศวกรรมศาสตร์ ให้กับนักศึกษาวิศวกรรมศาสตร์ไทย ข้อมูลที่ใช้เป็นตำรา วิศวกรรมศาสตร์รูปแบบออนไลน์ ใช้ในวิชาพื้นฐานด้านวิศวกรรม 5 รายวิชา ได้แก่ เขียนแบบ วิศวกรรม กลศาสตร์วิศวกรรม วัสดุวิศวกรรม โปรแกรมคอมพิวเตอร์ และวิศวกรรมไฟฟ้า จำนวน 9 เล่ม รวมคำศัพท์ 1,589,065 คำ วิเคราะห์ข้อมูลด้วยโปรแกรมประมวลผลคำ แยกเป็นรายการ คำศัพท์ทั่วไปและรายการคำศัพท์วิชาการ วิเคราะห์ความสัมพันธ์คำศัพท์ประเภทเนื้อหาในรายการ คำศัพท์วิชาการกับตำรา โดยการทดสอบไคสแควร์ ประกอบค่าความถี่แบบร้อยละ เสนอผลการวิจัย โดยอิงไวยากรณ์ของเควิกและกรีนบอม (1976) โดยพบว่า มีการใช้คำศัพท์ทั่วไปที่ใช้มาก 1000 คำแรก และ 1000 คำหลัง รวมทั้งคำศัพท์วิชาการ 570 คำแรก คิดเป็น 65.4% 5.6% และ 8.6% ตามลำดับ ค่าไคสแควร์เท่ากับ 5,404.24 มีนัยสำคัญ ที่ระดับ .0001 ทั้งนี้ ผลการศึกษาแสดงให้เห็นว่าการใช้คำศัพท์ประเภทเนื้อหาในรายการคำศัพท์วิชาการที่สร้างจากคลังข้อมูลที่เก็บจากตำรา วิศวกรรมศาสตร์ 9 เล่ม แตกต่างกันอย่างมากระหว่างคำนาม 60% คำกริยา 20% คำคุณศัพท์ 17% และคำกริยาวิเศษณ์ 3% แสดงให้เห็นว่าตำราภาษาอังกฤษด้านวิศวกรรมศาสตร์มีการใช้ คำศัพท์ภาษาอังกฤษที่มีความถี่ในการปรากฏสูง

**คำสำคัญ:** คำศัพท์ที่มีความถี่สูง คำศัพท์แกนร่วมภาษาอังกฤษ นักศึกษาวิศวกรรมศาสตร์ใน มหาวิทยาลัยไทย คำศัพท์ทั่วไปที่ใช้มาก 2000 คำแรก คำศัพท์วิชาการ 570 คำแรก

## Abstract

The research purposes were (1) to investigate how often and how many of the first 1000 word families including function words, the second 1000 word families of the General Service List (GSL) (West, 1953) and 570 word families of the Academic Word List (AWL) (Coxhead, 2000) are encountered in nine highly used English engineering texts; (2) to create general word lists and academic word lists from the nine English engineering text-based corpus based on the most frequent 2000 word families of the GSL and 570 word families of the AWL; (3) to determine if the distributions of the parts of speech of the content words in the academic word list compiled from the nine English engineering text-based corpus significantly differ from one another in the nine English engineering texts; and (4) to develop a high frequency English common core vocabulary word list for Thai university engineering students in reading English engineering texts. The data were compiled from 9 engineering e-Books selected from 5 basic engineering subjects: Engineering Drawing, Engineering Mechanics, Engineering Materials, Computer Programming, and Electrical Engineering, totaling 1,589,065 tokens. The data were analyzed by the Range program, and classified into general and academic word family lists. The relationship between distributions of the parts of speech of the content words in the academic word list compiled and types of engineering texts was examined using X<sup>2</sup> test and percentages. The findings were discussed based on Quirk & Greenbaum's grammar (1976). The results show that the first 1000 word families, the second 1000 word families of the GSL and 570 word families of the AWL covered 65.4%, 5.6%, and 8.6% respectively; X<sup>2</sup> equals 5,404.24; p-value is.0001. The findings revealed that there were significant differences in the parts of speech of the content words in the academic word list compiled from the nine text-based English engineering corpora, consisting of 60% nouns, 20% verbs, 17% adjectives, and 3% adverbs. The research revealed that high frequency English vocabulary occurs in the English engineering texts.

**Keywords:** High Frequency Vocabulary, English Common Core Vocabulary, Thai University Engineering Students, the General Service List, the Academic Word List

## Statements and significance of the problem

At present, the Ministry of Education Thailand encourages all Thai students in both basic education and higher education institutions to learn English, as the English language is an essential tool for communication, education, careers, daily living, etc. (Ministry of Education Thailand, 2008). There is a special focus on improving English proficiency in communication with nationals of the Association of Southeast Asian Nations (ASEAN), comprising Indonesia, Malaysia, the Philippines, Singapore, Thailand, Brunei, Myanmar, Cambodia, Laos, and Vietnam (Association of Southeast Asian Nations, 2012).

Thus Thai university students who have been studying English as a foreign language at primary and secondary education levels should be more acutely aware of the proficiency level as to just how much and what aspects of the English language are necessary for their communication with others. They have to access academic, technical, scientific and other publications written in English. Thus, reading skills are more important for them than listening and speaking skills. Furthermore, when studying languages, it is useful to set language learning goals (Nation, 2001, p.6) such as language for the communication skills of listening, speaking, reading and writing. To achieve these goals, vocabulary is considered an extremely important and an essential part of language learning goals.

The Thai Basic Education Core Curriculum 2008 prescribes vocabulary learning goals for long term course study for primary and secondary education levels: Grade 3 graduates with around 300-450 words with concrete words; Grade 6 at around 1,050-1,200 words with concrete and abstract words; Grade 9 at around 2,100-2,250 words with words of higher abstract quality, and Grade 12 at around 3,600-3,750 words with different levels of usage (Ministry of Education Thailand, 2008). However, vocabulary learning goals have not been officially established in curricula for long term courses of study for Thai university students in higher education. Moreover, reading English language academic engineering texts is necessary for Thai university engineering students.

Based on four main kinds of vocabulary in English: high frequency vocabulary, academic vocabulary, technical vocabulary, and low frequency vocabulary (Coxhead, 2000; Chung & Nation, 2003; 2004; Nation, 2001; Nation & Meara, 2010; Nation & Waring, 1997), high frequency and academic vocabulary are necessary for students when reading academic texts, that is, the most frequent 2000 word families of the GSL (West, 1953) have been found to account for around 80% of running words in typical texts, as well as 570 word families of the AWL (Coxhead, 2000) at around 10% (Hwang & Nation, 1995; Sutarsyah, Nation, & Kennedy, 1994; Nation & Waring, 1997).

However, some students still lack knowledge of the high frequency and academic vocabulary required for university entrance. The problem comes partly from their neglect of vocabulary learning from primary to secondary education levels. Moreover, some textbooks do not provide suitable vocabulary for student levels (Matsuoka & Hirsh, 2010; Martini, 2012; Moiiinvaziri, 2012; Wan-a-rom, 2012).

Thus, these vocabulary learning goals are still unattained for most Thai university engineering students who have studied English as a foreign language. From the present researcher's experience in the Thai context, students learn more subjects through Thai lectures given in Thai and only occasionally have the opportunity to converse in English. Moreover, academic vocabulary causes a great deal of difficulty for students since they are not generally as familiar with it as with technical vocabulary in their own fields, and academic vocabulary occurs at lower frequency than general vocabulary (Xue & Nation, 1984).

Thus there is an urgent need for more Thai research to investigate high frequency and academic vocabulary in textbooks for Thai university students. Subsequently, the general word lists and academic word lists for Thai university students should be available as reference sources in setting the vocabulary learning goals for reading their academic texts. The main purpose of this research is therefore to develop a high frequency English common core vocabulary word list based on high frequency and academic vocabulary for Thai university engineering students in reading English language academic engineering texts.

## Objectives

1. To investigate how often and how many of the first 1000 word families including function words, the second 1000 word families of the General Service List (GSL) (West, 1953) and 570 word families of the Academic Word List (AWL) (Coxhead, 2000) are encountered in the nine highly used English engineering texts
2. To create general word lists and academic word lists from the nine English engineering text-based corpus based on the most frequent 2000 word families of the GSL and 570 word families of the AWL.
3. To determine if the distributions of the parts of speech of the content words in the academic word list compiled from the nine English engineering text-based corpus significantly differ from one another in the nine English engineering texts.
4. To develop a high frequency English common core vocabulary word list for Thai university engineering students to assist in reading English engineering texts.

## Research questions

1. Which words and how many words in the first 1000 words, the second 1000 words of the General Service List (GSL) (West, 1953) and 570 words of the Academic Word List (AWL) (Coxhead, 2000) are in the nine English engineering texts?

2. How many words occur in the general word lists and academic word lists compiled from the nine English engineering text-based corpus?

3. Do the nine English engineering texts differ from one another in terms of the parts of speech of the content words in the academic word list compiled from the nine English engineering text-based corpus, i.e. nouns, verbs, adjectives and adverbs?

## Hypotheses

1. The most common vocabulary in the GSL and the AWL are found in the nine English engineering texts.

2. The proportions of general and academic vocabulary in the nine engineering texts are 70% for the first 1000 word families; 5% for the second 1000 word families of the General Service List (GSL) (West, 1953) and 10% for the 570 word families of the Academic Word List (AWL) Coxhead (2000).

3. There are significant differences between the distributions of nouns, verbs, adjectives, adverbs and types in engineering texts.

## Definitions of Terms

A 'High frequency word' is frequently used with a widespread occurrence, e.g. the, of, a.

An 'Academic word' is a word which appears with high frequency in English language academic texts, e.g. evaluate, invest, valid.

A 'Technical word' is a word related to the topic and a particular subject area, e.g. static, trachea.

A 'Low frequency word' is infrequently used with a narrow range of occurrence, e.g. aired, zoned, pastoral.

A 'Word list' is a list of all the words that occur in a corpus, e.g. GSL, AWL. The list includes function words and word families.

'Thai university engineering student' refers to an undergraduate Thai student who is studying for a Bachelor's degree in engineering, namely civil, electrical, industrial, mechanical, or chemical engineering.

'Text' refers to English engineering texts commonly used by Thai university

engineering students.

'Corpus' refers to the Engineering textbook-based corpus analyzed in the research. 'GSL' is a General Service List of English words which contains 2000 high frequency words.

'AWL' is an Academic Word List of 570 words that are not in the GSL.

'Type' refers to each different word form in a text, irrespective of how many times it occurs.

'Token' refers to the total number of word forms in a text.

'Word family' is a set of related words that consist of a base word and its related inflected and derived forms even if the part of speech is not the same, e.g. add, adds, added, adding, addition, additional, additive.

'Coverage' refers to the percentage of tokens in a text which are accounted for in particular word lists.

'Part of speech' means any of the classes into which words in a language have traditionally been divided on the basis of their meaning, form, or syntactic function, as in English: noun, verb, adjective, adverb, article, demonstrative, pronoun, preposition, conjunction, and interjection.

'Open-class items' refer to the sets of items that can indefinitely be extendable by the creation of new items, e.g. nouns.

'Closed-system items' refer to the sets of items that cannot normally be extended by the creation of additional members, e.g. pronoun.

## Review of Related Literature

According to Nation (2001, p.6), there are three elements that can help determine how much vocabulary is to be learned: the number of words in the language, the number of words known by native speakers, and the number of words needed to use the language. However, native speakers do not know all the lexical items of their language, such as those used in nuclear physics or computational linguistics (Goulden, Nation & Read, 1990). Adult educated native speakers of English know around 20,000 word families, and they will add roughly 1000 word families a year to their vocabulary size. So, a twenty-year-old university student will have a vocabulary size of about 20,000 word families (Goulden et al., 1990; Zechmeister, Chronis, Cull, D'Anna, & Healy, 1995; Nation & Waring, 1997; Schmitt, 2000; Nation, 2001; Nation, & Meara, 2010; Nation, 2012).

Nation, (2006, 2012) suggested that the amount of vocabulary needed for receptive use (reading and listening) is around 6,000 word families to read novels for teenagers, to watch movies, and to participate in friendly conversation; around

8,000 to 9,000 word families are needed to read newspapers, novels, and some academic texts. These figures assume 98% coverage of the input texts, which still leaves one word in every 50, or around six words on every page (a 300 running word regular page), as unknown vocabulary. For reading texts, the most frequent 2000 word families of the GSL are the basic vocabulary needed for students to read English texts and to have other language skills as well (Hwang & Nation, 1995; Sutarsyah, Nation, & Kennedy, 1994; Nation & Waring, 1997; Nation, 2012).

Furthermore, Coxhead's study (2000) also showed that the first 1000 word families cover 70.9 %, the second 1000 word families cover 4.6 %, 570 academic word families cover 10.0 %, and other words (e.g. low frequency words and technical words) cover 14.5% in an academic corpus. Thus, the percentage of text coverage by the two main kinds of high frequency and academic vocabulary when combined cover 85.5%, while about 14.5% are not in the high frequency and academic vocabulary.

It is therefore widely accepted that the most frequent 2000 word families of the GSL (West, 1953) and 570 word families of the AWL (Coxhead, 2000) are necessary as the main vocabulary for university students in reading academic texts (Nation, 2001; Nation & Meara, 2010; Wan-a-rom, 2012; Moivaziri, 2012; Moudraya, 2006).

In addition, high frequency vocabulary word lists have been compiled from corpora of various written texts, namely the GSL (West, 1953), the UWL (Xue & Nation, 1984), and the AWL (Coxhead, 2000). These word lists were created based on range and frequency criteria by computer programs to identify the most common words in an academic text.

As mentioned earlier, the most frequent 2000 word families of the GSL and 570 word families of the AWL contain predominantly general and academic vocabulary that university students would encounter in a wide range of academic texts.

## Research framework

Figure 1 illustrates the research framework of the study.

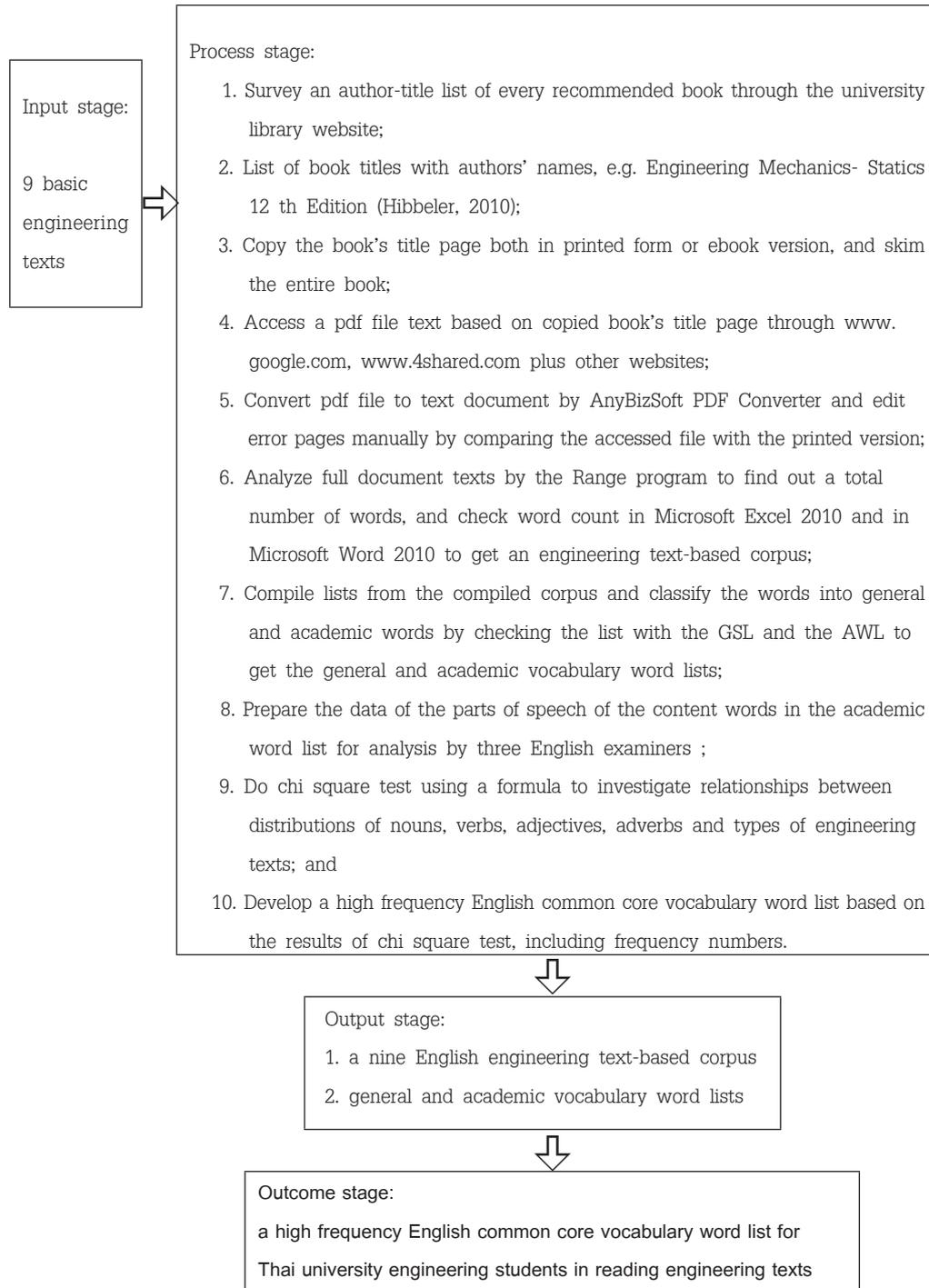


Figure 1 Research framework of the study

## Research Methods

The data were 1,589,065 words compiled from nine engineering e-Books: 1) Newnes Electrical Engineer’ s Handbook (Warne, 2000), 2) Principles of Electric Circuits - 8th Edition (Floyd, 2007), 3) Electronic Devices 7th Edition (Floyd, 2005), 4) Python Programming: An Introduction to Computer Science (Zelle, 2002), 5) Engineering Mechanics- Statics 12th Edition (Hibbeler, 2010), 6) Engineering Mechanics- Dynamics 12th Edition (Hibbeler, 2010), 7) Fundamentals of Materials Science and Engineering 5th Edition (Callister, 2001), 8) Pro/Engineer for Designers (Tickoo, 2001), and 9) Programming C# (Liberty, 2001). The books were selected from five basic engineering subjects required for all Thai university engineering students to obtain a professional engineering license issued by the Council of Engineers of Thailand (Council of Engineers of Thailand [COE], 2010), namely Engineering Drawing, Engineering Mechanics, Engineering Materials, Computer Programming, and Electrical Engineering. These books are generally well-known and widely published textbooks, recommended by engineering teachers, students, and faculty members.

The data were analyzed using the Range program, a Windows-based computer program created by Heatley, Nation & Coxhead in 2002, AnyBizSOft PDF Converter, Microsoft Excel 2010, and Microsoft Word 2010. The statistical analyses used were percentages and Chi-square statistics.

## Results

Table 1 presents the words in the GSL and the AWL found in the nine English engineering texts.

**Table 1** Number of tokens, types and families in the nine English engineering text-based corpus

Word list	Tokens/%	Types/%	Families
1st 1000	1,038,503/65.4	2,805/8.6	952
2nd 1000	89,343/5.6	1,709/5.2	731
Academic	136,485/8.6	1,913/5.8	541
not in the lists	324,734/20.4	26,340/80.4	...
Total	1,589,065/100	32,767/100	2,224

The computer run yielded a total of 1,589,065 tokens, containing 32,767 types which were from 2,224 word families. The proportions of the text coverage taken by the first 1,000 word families, the second 1,000 word families, and 570 academic word families were 65.4%, 5.6%, and 8.6% respectively. About 324,734 words or 20.4% were not in these three word lists.

The compiled word lists, excluding months, days of the week, numbers, titles, and letters A-Z, contained 895, 738, and 543 word families in the first 1000, the second 1000, and academic word families, as shown in Table 2.

**Table 2** Frequency and the number of word families in the 1,589,065 token corpus

Frequency	Number of word families		
	1st 1000	2nd 1000	Academic
100,000 and above	1	0	0
10,000 and above	10	0	0
1,000-9,999	135	14	32
100-999	369	146	204
10-99	253	252	217
1-9	127	312	90
0	0	14	0
Total	895	738	543

Under the frequency of 100,000 and above, only one family “the” occurred in the first 1000 word families and 14 families occurred in the second 1000 word families below a frequency of 0. There were word families in all three word lists under the frequency of 1,000-9,999, 100-999, 10-99 and 1-9 respectively.

Table 3 shows samples of the 10 most frequent word families from the compiled corpus.

**Table 3** The 10 most frequent function words and word families from the corpus of 1,589,065 running words

1 <sup>st</sup> 1000		2 <sup>nd</sup> 1000		Academic	
Function Word and Word Family	Frequency	Word Family	Frequency	Word Family	Frequency
THE	127989	RESIST	2187	SELECT	3085
BE	69844	TEMPERATURE	2178	CREATE	2995
A	58697	LOAD	2034	INPUT	2867
OF	52854	PLANE	1934	OUTPUT	2699
AND	36331	SOLVE	1928	SECTION	2622
IN	31140	FREQUENT	1867	EQUATE	2547
TO	30581	MOTION	1582	CHAPTER	2479
THIS	25005	BUTTON	1483	METHOD	2126
FOR	14336	PROPERTY	1355	COMPONENT	2027
THAT	12089	MODEL	1297	OPTION	2014

In Table 3, the function word “THE” and word families “RESIST”, and “SELECT” were the most frequent word families in the first 1000, the second 1000, and academic word families respectively.

There were significant differences between the distributions of nouns, verbs, adjectives, adverbs of the content words in the academic word list, as shown in Table 4.

**Table 4** Frequency number of nouns, verbs, adjectives, adverbs of the content words in the academic word list compiled from the nine English engineering text-based corpus of 1,589,065 running words

Part of speech	New	Prin	Elec	Python	Sta	Dyn	Fund	Pro	C#	Total
Nouns	5,452	7,216	14,947	6,244	5,803	5,505	17,153	12,218	7,033	81,571
Verbs	1,661	2,166	3,569	1,522	1,761	2,457	4,697	6,178	2,891	26,902
Adjectives	1,990	2,724	4,848	1,227	2,440	2,640	4,660	1,656	1,201	23,386
Adverbs	297	225	423	193	200	351	1,389	259	294	3,631
Total	9,400	12,331	23,787	9,186	10,204	10,953	27,899	20,311	11,419	135,490

In Table 4, the nine different engineering texts consisted of 81,571 nouns, 26,902 verbs, 23,386 adjectives, and 3,631 adverbs, totaling 135,490 words. Nouns had the largest number of words; verbs the next; the adjectives and adverbs were the smallest number of words.

The results of  $X^2_{(24)}$  tests which investigated whether the distributions of nouns, verbs, adjectives, and adverbs of the content words in the academic word list significantly differ from one another in the nine English engineering texts were much greater than p value:  $X^2_{(24)}$  values equaled 5,404.24 with an alpha level of significance of 0.0001 (p-value was .0000). The results of  $X^2_{(8)}$  equaled 1,593.224 between nouns and verbs; 2327.425 between nouns and adjectives; 827.4817 between nouns and adverbs; 3515.339 between verbs and adjectives; 1143.43 between verbs and adverbs; and 825.6891 between adjectives and adverbs.

A word list created from the chi square results consisting of 539 academic word families, was edited and organized into 10 sublists based on the AWL (Coxhead, 2000). In the list, there were 60 word families in sublist 1, 58 in sublist 2, 57 in sublist 3, 58 in sublist 4, 55 in sublists 5, 6, and 7, 57 in sublist 8, 58 in sublist 9, and 26 word families in sublist 10. Sublist 1 contained the most frequent word families, sublist 2 the next most frequent word families, and sublist 10 the least frequent word families. Table 5 shows a sample entry from the 539 academic word family list.

**Table 5** A sample entry from the 539 high frequency English common core vocabulary word family list

Sublist 1

Headword	Frequency	Members	Part of speech	Range	Frequency
ANALYSE	1082				
		ANALYSE	v	2	2
		ANALYSIS	n	8	758
		ANALYTICAL	adj	6	13
		ANALYTICALLY	adv	3	4

In Table 5, the list was organized into alphabetical order, comprising headwords and family members, giving parts of speech (nouns, verbs, adjectives, and adverbs), frequency numbers (how often the headword and its members occurred in the corpus), and a range (how many of the nine different engineering texts the headword and its members occurred in).

## Conclusions and Discussions

This research investigated English vocabulary, and focused on the first 1000, the second 1000 word families and function words of the GSL, and the 570 word families of the AWL, appearing in the nine English engineering texts, based on four main kinds of vocabulary in English by frequency criteria: high frequency vocabulary, academic vocabulary, technical vocabulary, and low frequency vocabulary.

According to the results, the most common vocabulary in the GSL and the AWL were found in the nine English engineering texts with proportions equivalent to Coxhead (2000): 70%, 5%, and 10% (65.4%, 5.6%, and 8.6%, in fact) in the first 1000, the second 1000, and 570 academic word families respectively.

Nation's (2004) study also showed that the two word lists between the BNC and the GSL plus the AWL overall contained much of the same vocabulary; however, they were not distributed in the same way in each of the lists. The AWL vocabulary occurred across the three BNC lists, although they were developed from quite different corpora and at widely different times.

In the same way, this research also compared the compiled lists with two standard word lists of the GSL and the AWL. The results showed that even though these sets of word lists were developed from quite different corpora, overall they contained much of the same vocabulary. Although the vocabulary was not distributed equally in the same proportion of items in each of the lists, the difference was not very large: 895-738-543 VS 998-988-570 (Heatley, Nation & Coxhead, 2002).

All three compiled word lists contained much of the same vocabulary with 895, 738, and 543 word families in the first 1000, the second 1000, and the 570 academic word family lists, respectively. The word list of 539 high frequency English common core vocabulary contained much of the same vocabulary as the 570 word families of the AWL by Coxhead (2000).

However, there were 27 word families from the AWL that did not appear in the corpus: amend (5), abandon (8), albeit (10), currency (8), debate (4), deny (7), ethnic (4), estate (6), ethic (9), forthcoming (10), gender (6), immigrate (3), incentive (6), ideology (7), infer (7), levy (10), ministry (6), notwithstanding (10), overseas (6), partner (3), paradigm (7), practitioner (8), revenue (5), sex (3), subordinate (9), voluntary (7), and welfare (5). These missing word families were not in sublists 1, and 2 which each contained the 60 most frequent, widest range word families. On the other hand, there were four word families out of the 30 least frequent word families in the sublist 10. This implied that the most frequent words in sublists 1 and 2 widely occurred across the engineering texts.

Thus, the findings confirmed that the compiled list was similar to the standard list of the AWL. From the results, the nine different engineering texts consisted of 81,571 nouns, 26,902 verbs, 23,386 adjectives, and 3,631 adverbs, totaling 135,490 words. Nouns represented the largest number of words; verbs the next; adjectives and adverbs had the smallest number of words. In other words, the percentages of the frequency of words by part of speech in the nine engineering texts were 60% nouns, 20% verbs, 17% adjectives, and 3% adverbs.

Thus, there were significant differences in the distributions of nouns, verbs, adjectives, adverbs and types of engineering texts.

Today there are several standard word lists, such as the General Service List (GSL), the University Word List (UWL), the Academic Word List (AWL), and the BNC word list (BNC). However, these word lists represent words in the broadest sense as a widely available word list for a baseline of general studies. The lists are organized into frequency or alphabetical order, comprising headwords and family members, giving frequency numbers and the rank. However, no parts of speech code are shown for each word. Without parts of speech information, a student's reading comprehension would be inadequate.

Moreover, knowledge of English grammar is essential for students to understand word usage. These important aspects of grammar include sentence elements and parts of speech, especially the distinctions between subject and predicate in a sentence, and between open-class items and closed-system items in parts of speech. Simply put, most of the sentences encountered in the engineering

texts were too long and complicated for students to understand the word form and the word function.

Based on the results of  $X^2_{(24)}$ , and of  $X^2_{(8)}$  mentioned earlier,  $X^2$  values were much greater than the p value with an alpha level of significance of 0.0001 (p-value was 0.000). When comparing the Chi square distribution table with 24 degrees of freedom, the p value of  $X^2_{(24)}$  5,404.24 were not in any row, therefore, the probability was 0.000. This was much smaller than the conventionally accepted significance level of 0.05 or 5%. The hypothesis that there are significant differences between distributions of nouns, verbs, adjectives, adverbs and types of engineering texts was accepted, since the resultant  $X^2_{(24)}$  5,404.24 exceeded the critical value for a 0.05 probability level  $0.0001 < P < 0.0000$ .

The findings imply that the knowledge of parts of speech for English is required for students in understanding word usage, not only in word forms but also in word functions as nouns, verbs, adjectives, and adverbs in sentences as in:

1. For example, if you are creating a purchase (adj) form, you might create one radio button for each possible shipping method, such as UPS, FedEx, and so forth.
2. Suppose, however, that you purchase (v) a List class from a third-party vendor whose capabilities you wish to combine with those specified by Storable.

Moreover, in relation to verbal forms, English verbs have five forms: the base, the -s form, the past, the -ing participle, and the -ed participle (Quirk & Greenbaum, 1976, p. 27). The findings found that there were differences between verbal forms and functions used in the sentences of the -ing participle, and the -ed participle forms as in:

3. In the preceding (adj) examples, both x and y are variables.
4. Before proceeding (n), take a look around.
5. The use of pointers is almost never required (v), and is nearly always discouraged.

The knowledge of the distinction between subject and predicate in a sentence plus parts of speech for English enabled correct identification of the word forms and functions of the words in a sentence.

## Recommendations

1. This research was mainly conducted on general and academic vocabulary and their occurrences in the nine English engineering texts, excluded technical vocabulary. The future study of the occurrence of technical vocabulary in a corpus of other textbooks is advised.

2. The sample texts were selected from five basic engineering subjects for Thai university engineering students. Texts from other subjects should be investigated.

3. The research tools used for data collection and data analysis was the Range program. Ideally, several other computer programs should also be used to discover other possible aspects of the data.

4. The typical proportion of the text coverage taken by the first 1000, the second 1000, and 570 academic word families and function words should be confirmed by other research studies.

5. The Range program could bring genuine benefits in doing a corpus-based research and the standard Microsoft Excel 2010 and Microsoft Word 2010 provided easier access to the reliable application of formulae and the calculation of  $X^2$  values instead of using web-based formulae.

6. The compiled word list should be introduced directly to students, and teachers can design the language tasks by using this list based on the sublists 1-10.

Thus, the compiled word list enables Thai university engineering students to directly use it in reading English language academic engineering texts with the technical vocabulary and proper nouns necessary for their comprehension of the text. English teachers and subject teachers (e.g. engineering teachers) can apply it in setting vocabulary learning goals for academic reading purposes; and for educators to have a resource for setting vocabulary learning goals for Bachelor's degrees in engineering. In addition, the list could provide educators and researchers with information on the vocabulary load faced by Thai university engineering teachers, enabling them to plan and conduct lessons more effectively.

## References

- Association of Southeast Asian Nations. (2012). *ASEAN Communities*. Retrieved June 27, 2013, from <http://www.asean.org/communities/asean-economic-community>.
- British National Corpus. *BNC electronic list*. Retrieved June 27, 2013, from <http://www.natcorp.ox.ac.uk/>, 2012.
- Callister, W. D. (2001). *Fundamentals of Materials Science and Engineering*. (5th ed.). New York: John Wiley & Sons.
- Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15(2), 103-116.
- Chung, T., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251-263.
- Council of Engineers of Thailand (COE). (2010). *Basic Engineering*. Retrieved June 30, 2013, from [http://www.coe.or.th/\\_coe/\\_coenew/mainIndex.php?aMenu=8](http://www.coe.or.th/_coe/_coenew/mainIndex.php?aMenu=8).
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Floyd, T. L. (2005). *Electronic Devices*. (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Floyd, T. L. (2007). *Principles of Electric Circuits*. (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341-363.
- Heatley, A., Nation, P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Retrieved May 28, 2012, from [http://www.vuw.ac.nz/lals/staff/Paul\\_Nation](http://www.vuw.ac.nz/lals/staff/Paul_Nation).
- Hibbeler, R.C. (2010). *Engineering Mechanics- Dynamics*. (12 th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hibbeler, R.C. (2010). *Engineering Mechanics- Statics*. (12 th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hwang, K., & Nation, P. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- Liberty, J. (2001). *Programming C#*. Sebastopol, CA: O'Reilly & Associates.
- Martini, J. (2012). *High Frequency Vocabulary in a Secondary Quebec ESL Textbook Corpus*. MA Thesis, Concordia University, Montreal, Quebec, Canada.
- Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22(1), 56-70.
- Ministry of Education Thailand. (2008). *The Basic Education Core Curriculum 2008*. Bangkok: Kurusapa Ladprao. Retrieved June 30, 2013, from <http://www.curriculum51.net/viewpage.php?tid=95>

- Moiinvaziri, M. (2012). Analyzing Vocabulary Used in Payame Noor University General English Textbook: a Corpus Linguistic Approach. *Frontiers of Language and Teaching*, 3, 204-212.
- Moudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purpose*, 25, 235-256.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. (2004). A study of the most frequent word families in the British National Corpus. In Bogaards, Nation, P., & Laufer, B. (Eds.). *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam: John Benjamins: pp.3-13.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, P. (2012). *Measuring vocabulary size in an uncommonly taught language*. Retrieved June 27, 2013, from <http://www.sti.chula.ac.th/files/conference%20file/doc/paul%20nation.pdf>.
- Nation, P., & Meara, P. (2010). Vocabulary. In Schmitt, N. (Ed.), *An Introduction to Applied Linguistics*, (2 nd ed.), pp. 34-52. London: Hodder Education.
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In *Vocabulary: Description, Acquisition and Pedagogy*. Schmitt, N. and Quirk, R., & Greenbaum, S. (1976). *A University Grammar of English*. London: Longman.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal*, 25(2), 34-50.
- Tickoo, S. (2001). *Pro/Engineer for Designers*. CAD/CIM Technologies.
- Wan-a-rom, U. (2012). Lexical Evaluation of Teacher-made Coursebooks: Thai Case Studies of Foundation English Courses at Tertiary Level. *English Language Teaching*, 5(8),146-156.
- Warne, D. F. (2000). *Newnes Electrical Engineer's Handbook*. Oxford: Butter worth-Heinemann.
- West, M. (1953). *A General Service List of English Words*. London : Longman, Green & Co.
- Xue, G., & Nation, P. (1984). A University Word List. *Language Learning and Communication*, 3(2), 215-229.

- Zechmeister, E.B., Chronis, A.M., Cull, W.L., D'Anna, C.A., & Healy, N.A. (1995). Growth of a functionally Important lexicon. *Journal of Reading Behavior*, 27(2), 201-212.
- Zelle, M.J. (2002). *Python Programming: An Introduction to Computer Science*. Wartburg College.