



## General vocabulary in Thai EFL university students' writing: A corpus-based lexical study

ANCHALEE VEERACHAISANTIKUL\*, SUKANYA CHOOTARUT

Rajamangala University of Technology Isan, Nakhon Ratchasima, Thailand

### Abstract

**Aim:** This study, part of a larger task investigation, questions the use of general vocabulary in Thai EFL university students writing. The most common words in English are represented by general vocabulary, which is helpful for language students. The two main goals of the corpus-based lexical study were (a) to compile a list of frequently used general words in Thai EFL university students' writing and (b) to compare the word list to the New General Service List (NGSL).

**Method:** The authors gathered and analyzed 1,233 writing assignments totaling 661,596 words, all completed by Thai EFL college students. We used WordSmith Tool Version 6 (Scott, 2012) for this research, a concordancing program.

**Findings:** According to our findings, out of the 2,818 NGSL high-frequency words, 1,648 were used frequently throughout the corpus, with this coverage accounting for 1.41 percent of the token total. By comparing our words to the NGSL, we found that most of the core vocabulary in the TEFL Corpus was, in fact, very close to the NGSL.

**Implications/New Contribution:** The study's findings will inform educators about the role of vocabulary in writing and will be useful to students and curriculum developers as they create resources for the instruction of writing in English as a Foreign Language (EFL) sector. Furthermore, these findings could be used by educators to enhance grammar and vocabulary instruction in the classroom. In addition, students can use these findings to better comprehend and appreciate how English is actually used in the real world to develop their writing skills.

*Key Words:* General Vocabulary, The New General Service List, Corpus-Based Lexical Study

**Received:** 02 September 2016 / **Accepted:** 22 October 2016 / **Published:** 20 December 2016

### INTRODUCTION

Numerous studies have argued that vocabulary is crucial to success when studying a second or foreign language (Schmitt 2008; Schmitt and McCarthy 1997; Read 2000). Vocabulary knowledge is foundational to language learning, as it can proxy for a learner's proficiency (Desi-Konsortium 2013). Lessard-Clouston (2013) argued that vocabulary knowledge is crucial when instructing students of English as a second language. Students will only communicate their ideas and understand those of others once they develop a wider vocabulary.

Therefore, vocabulary is an integral part of the four facets of communication (Harris and Sipay 1990). Writing, the capstone competency for any language study is especially relevant to vocabulary knowledge. The ability to express oneself in Writing is a significant milestone for language learners because it validates how they develop their unique style in the target language. This also improves their ability to communicate in front of various audiences and in various settings (Catramado 2004). However, the vast majority of them are unable to master this ability. Because learning to express oneself in written form is an acquired skill that involves expanding one's lexicon. Some studies have found that ESL and EFL students' writing skills improve along with their vocabulary knowledge. To paraphrase Bello (1997):

“Writing helps students learn new vocabulary and grammar rules by providing them with practice in putting those rules into practice as they try out new ways of expressing their thoughts on paper.”

Al-Dersi (2013) agrees with this sentiment, writing that it is essential for students of a foreign language to have a mature vocabulary to succeed in their studies. Ince, as Nation advises, a larger vocabulary is necessary for teaching at higher levels. A lack of vocabulary is a common challenge in English learning (2001). Therefore,

\* Corresponding author: Anchalee Veerachaisantikul

† Email: [veeanchalee@gmail.com](mailto:veeanchalee@gmail.com)

educators must choose vocabulary appropriate for their students' skill levels. Theoretical and empirical studies have recommended general vocabulary instruction for EFL students due to its potential benefits for language learners (Kuno 1999; Ito 2000; Saku and Honda 2004). Here is a need for studies examining the Writing of Thai college students. However, this does not mean that no study has been conducted on the topic. Some linguistic studies focus on individual words, such as verbs, liking, adverbs, and connector studies. However, not enough effort is put into implementing a standard vocabulary list. That's why we're conducting this research: to compile a comprehensive list of Thai college students' most commonly used words. The findings of this research could be refined and used to create EFL resources for Thai students.

### **Objectives of the Study**

1. To collect general words from the Thai EFL University Students' Writing (TEFL) Corpus.
2. To compare the general word list of the TEFL Corpus with the New General Service List (NGSL).

### **Research Questions**

The current study examines the following questions:

1. What are the most frequently used general words in Thai EFL university students' writing (TEFL) Corpus?
2. How different are they from those in the New General Service List (NGSL)?

## **METHODOLOGY**

### **The Corpus**

The selected corpus of this study composed of 1,233 writing tasks written in English by 154 EFL students of Rajamangala University of Technology Isan (RMUTI) as a part of English Writing for Daily Life course for over an academic year. RMUTI students were asked for studying this course as a requisite subject of their Bachelor degree. In addition, they had to pass two fundamental English courses in their first year before attending this course. In this course, they were asked to write essay sets that had eight topics as follows:

Task A: Who I am?

Task B: Writing about your classmate/roommate

Task C: Writing about your family

Task D: Writing about yourself

Task E: The Weaker Sex

Task F: The Shared Refrigerators

Task G: Reason not to own Hammer Car

Task H: The Smart Car

Table 1: General statistics of each task

Task	A	B	C	D	E	F	G	H	Total
No. token	9,598	11,527	20,961	12,657	11,016	28,037	37,865	11,366	143,027
Types	49	598	84	729	635	2,788	3,389	2,970	11,242

According to Table 1, the tokens or running words of each task were 9,598 for task A, 11,257 for task B, 20,961 for task C, 12,657 for task D, 11,016 for task E, 28,037 for task F, 37,865 for task G, and 11,366 for task H. Task G had the highest number of token words while task A had the smallest number of tokens. As for word types, task G also had the highest number of different word types (3,389) whereas task A had the smallest number of different word types (49). The six other tasks of the TEFL Corpus were close to one another in terms of number of word types, namely 598, 84, 729, 635, 2,788, and 2,970 in task B, C, D, E, F, and H respectively.

**The Software for Analysis**

The concordancing software called “WordSmith Tool Version 6” (Scott 2012) was selected for this study. This is an integrated suite of programs for looking at how words behave in texts. It was used to examine how words were used in any kind of texts. The Wordlist tool of concordancing software was used for making a word list of vocabulary. The characteristic capacity of wordlists serves the alphabetical and frequency order of the words and phrases (Scott 1999). It can make a comparison between numbers of words or phrases then the results are completed by the selection and grading of the words and phrases as the sources of dictionaries or teaching materials. Apart from that, it provides to distinguish the category of vocabulary by calculating token (running words) and type (distinct words) (Scott 2004).

**Procedure and Data Collection**

In order to investigate the frequency and range of general words, all writing tasks were typed in Microsoft Word Office 2010 because all data were handwritings. In this study, the researchers comprised the normalization, segmentation, and standardization. These words were changed into their simple forms such as plural nouns were converted into singular and segmentation in originating word family forms. Then, frequent general words used in the TEFL Corpus were classified in order to create a list of general words and compared with the New General Service Wordlist (NGSL) for analyzing the differences between two corpora.

**Data Analysis**

The research question purposed to examine the frequency of general words that are used in the TEFL Corpus. In order to answer these research questions, the first step was to employ the Wordlist tool for making the word frequency lists of each task. The Wordlist tool offered both alphabetical and frequency orders of the general words in the TEFL Corpus. Next, the most frequently occurring general words were obtained. After that, we picked the most frequently occurring general words in the TEFL Corpus then compared with the NGSL (Browne et al. 2013). The purpose of the comparison between general words in the TEFL Corpus and the NGSL is to illustrate the differences of frequency and range of most frequently occurring words.

**RESULTS AND DISCUSSION**

To study the frequency and distribution of general words, a corpus of 143,027 words from Thai EFL university students’ writings has been applied. After analyzing the data, the following results have been achieved that demonstrated to answer the two posted research questions below.

Research Question 1: What are the most frequently used general words in Thai EFL university students’ writing (TEFL) Corpus?

Table 2: The top 50 most frequently occurring general words of TEFL corpus

Rank	Word	F	%	Rank	Word	F	%	Rank	Word	F	%	Rank	Word	F	%
1	be	7,868	6.58	14	it	1,080	0.90	26	like	667	0.56	39	family	471	0.39
2	the	5,196	4.34	15	at	1,015	0.85	27	she	664	0.56	40	you	450	0.38
3	and	3,986	3.6	16	than	962	0.80	28	there	580	0.48	41	from	431	0.36
4	a	3,426	3.13	17	man	959	0.80	29	name	569	0.48	42	own	430	0.36
5	I	3,395	2.84	18	he	866	0.72	30	as	560	0.47	43	food	413	0.35
6	my	2,474	2.07	19	do	768	0.64	31	reason	552	0.46	44	cost	411	0.34
7	in	2,445	2.04	20	but	752	0.63	32	up	550	0.46	45	roommate	411	0.34
8	to	1,947	1.63	21	year	741	0.62	33	gas	548	0.46	46	sex	411	0.34
9	of	1,943	1.62	22	this	739	0.62	34	big	501	0.42	47	smart	411	0.34
10	on	1,371	1.15	23	live	694	0.58	35	that	500	0.42	48	weaker	411	0.34
11	woman	1,370	1.15	24	shelf	685	0.57	36	two	493	0.41	49	well	377	0.32
12	have	1,367	1.14	25	old	672	0.56	37	very	478	0.40	50	interesting	344	0.29
13	for	1,170	0.98					38	we	477	0.40				

Frequency of occurrences 60,001

Total of % of text coverage 50.69%

Table 2 provided the list of the top 50 most frequently occurring words of the TEFL Corpus. The total frequency of occurrences in this table was calculated from 1st rank to 50th rank. There was a total of 60,001 occurrences that accounted for 50.69% of text coverage, of the whole corpus which was 100%. It was to be supposed that high frequency words in this corpus were article, pronouns, and the verb to be. In the list, the top five high frequency words were “be”, “the”, “and”, “a”, and “I”. Also, the verb to be in the TEFL Corpus, “be” appeared 259 times, “am” 660 times, “is” 4,228 times, “are” 2,502 times, “been” 50 times, “was” 151 times, and “were” 18 times. Conversely, content words appeared less frequently than function words. For example, “women” was the highest frequency content word in this corpus which occurred 1,370 times at the 11th rank. The next most frequent content word was “man” with 959 times of occurrences at the 17th rank. The third most frequent content word was “year” appearing 741 times at the 21st rank.

Research Question 2: How different are they from those in the New General Service List (NGSL)?

Table 3: The comparison of 20 most frequently occurring general words of TEFL corpus and the New General Service List (NGSL)

Rank	TEFL Corpus	Freq.	NGSL	Freq.
1	be	7,868	the	60,910
2	the	5,196	be	48,575
3	and	3,986	and	30,789
4	a	3,426	of	30,126
5	I	3,395	to	29,272
6	my	2,474	a	27,872
7	in	2,445	in	21,142
8	to	1,947	have	14,210
9	of	1,943	it	13,772
10	on	1,371	you	12,810
11	women	1,370	he	11,017
17	men	959	on	7,763
18	he	866	with	7,381
19	do	768	this	7,003
20	but	752	I	6,820

The comparison of our words with the NGSL presented that general words in the TEFL Corpus were generally similar to the NGSL; however, there were some general words appearing with high frequency, their frequency order was different from the NGSL. For example, the most frequent general word occurring in the TEFL Corpus was “be” with 7,868 occurrences while in the NGSL “be” was at 2nd rank appearing 48,575 times. From the data, “the” in the NGSL was at 1st rank appearing 60,910 times whereas “the” was at 2nd rank in the TEFL Corpus.

### Discussion

To answer the research questions, the concordancing software WordSmith Tool Version 6 was used. It was used to view how words behave in texts and to create the word frequency lists of Thai EFL university students writings by using the Wordlist Tool which provided both alphabetical and frequency orders of the words in the text files. The entire corpus was 143,027 tokens or running words. Firstly, according to the analysis of this study, the 50 most frequently occurring words of the TEFL Corpus were obtained (Table 2). It was found that high frequency words in the corpus were article, pronoun, conjunction, and the verb to be. For instance, the top five high frequency words were “be”, “the”, “and”, “a”, and “I”. Moreover, the general statistics of each task demonstrated the number of the different word types of the whole corpus to be 11,424 word types. According to the data (Table 1), task G had the highest number of tokens and the highest number of different word types than the other tasks, namely task A, B, C, D, E, F, and H respectively. This is because some word types of the other seven tasks may repeat in more than one task of writing tasks. In addition, task G had the highest number of tokens and the highest number

of different word types as well since students tried to express their ideas giving reasons covered by the topic. It can be explained that it may be a larger piece of writing for EFL students when they are faced with writing such arguing, discussing, or giving reason and opinion topic. This means they tried to use various words to explain or describe their opinion, idea, or feeling more than other writing tasks. For these results, it is possible to improve EFL students/learners' writing by applying in practical situation, arguing, discussing, or giving reason and opinion. Secondly, after obtaining the top 50 most frequently occurring words, the top 20 most frequently occurring words were compared to the NGSL. It was revealed that the general words of the TEFL Corpus were mostly similar to each other that can be assumed that Thai university students have adequate vocabulary knowledge in English learning; however, there were differences in ranking. The reason for this difference might be the limitation of this study such as the total number of tokens or running words. Further, we also investigated the pattern in which high frequency general words most frequently occurred to double-check their vocabulary knowledge and how well the students know the words.

Table 4: The concordance lines of "BE"

N	Concordance		
1	watching Korean drama. Her birthday	is	on 1st June, 1991. Her favorite colors are white, pink, and
2	of a street parking space. Its body	is	modern in design. You can park two
3	season. Even a short time but we	were	very happy. I love my family very much and I want them
4	football team is Liverpool. I want to	be	a programmer in the future and I love to be free, too
5	lined up in rows. The carrots	are	old, salami is brown, the bread is bought yesterday
6	since he was born. He has	been	studying English for fifteen years but he cannot speak well
7	Nakhonratchasima for 6 years. She has	been	studying English for 17 years and she tries to understand
8	don't buy a Hummer. Hummer car	is	gas hogs. The H2 model is very big so it can cause to
9	company employees. I think that English	is	not difficult to learn but we have to try hard to understand
10	I appreciate it. Hey Juanita, I	was	so courteous to remind you again that could you
11	disgusting. My roommates and I	are	different but we get along well when we are together
12	are old salami is brown, the bread	is	green, and the lettuce is so dried out
13	of eggs sits on the shelf. The eggs	are	broken, carrots and salami are fresh because I
14	is easy on the wallet. The Smart car will	be	the next "cool" car to own for the future
15	My name is Gun. I	was	born on 2nd September, 1992. I am a student at
16	She likes watching movies. She wants to	be	an engineer in her near future
17	and I want my family to	be	happy in the future with me
18	at Nakhonratchasima. She has	been	staying in here for 7 years with her sisters
19	My mother and father	are	teachers. I have two younger brothers and older sister
20	My brother and I	were	students at the same high school in our province

The selected keyword was investigated in a concordance line by using Concord Tool of WordSmith Tools Version 6. In concordance lines, the keywords were set at the center with the left and right contexts. The above findings displayed in 20 concordances that seven of them were used as auxiliary verbs. As shown in Table 4, 3 concordances (N 6, 7, and 18) were used in present perfect passive sentences whereas 3 concordances (N 4, 16, and 17) were used for past simple passive events and 1 concordance was used in "infinitive with to" form. For 13 concordances, it also can be seen that "be" was used as main verb to represent existence, identity, color, mood, and negative sentences. This can be assumed that Thai EFL university students could use the "verb to be" in appropriate ways.

### CONCLUSION, RECOMMENDATIONS AND IMPLICATIONS

The importance of vocabulary in foreign language learning in all subject disciplines is one of the understanding areas in language (Cameron 2001). This study aimed to show the general words list of Thai EFL university students' writing. For this aim, a corpus of 143,027 running words of 1,088 writing tasks which were written by

154 EFL students of RMUTI was used. The researchers typed all tasks in the Microsoft Office Word 2010 and used the concordancing software “WordSmith Tool Version 6” (Scott 2012) in order to create a general words list of this study. As results in Table 1, 2, and 3 display, have a great number of the whole corpus. The findings exhibit that general words play an important role for EFL students who used in their daily life. As many researches stated that corpus-based studies have promoted as valuable sources for pointing out the practical information of writing in academic genres (Biber, Conrad, and Reppen 1998; Flowerdew 2005). The results of this study would help teachers to realize the significance of vocabulary in writing and be useful for learners or course designers when they write material for teaching writing in EFL area. Besides, teachers could apply these results in order to develop their classroom teaching for grammar and vocabulary. Also, learners can adapt these results to comprehend and realize how English is used in daily life for achieving their ability to write. In addition, this study reveals that a corpus study and the usage of concordancing software could be a reasonable tool for developing resources as direction to writing material in EFL courses. Presently, it is easier for material writer or course designers to develop materials and serve beneficial resources than in the past with the availability of computers and concordancing software.

## REFERENCES

- Al-Dersi, Z. E. 2013. “The Use of Short-Stories for Developing Vocabulary of EFL Learners.” *International Journal of English Language & Translation Studies* 1(1): 72-86.
- Bello, T. 1997. “*Writing Topics for Adult ESL Students.*” Presented at the 31st annual teachers of English to speakers of other languages convention, Orlando, FL.
- Biber, D., Conrad, S., and Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press.
- Browne, C., Culligan, B. and Phillips, J. 2013. “The New General Service List.” Retrieved November 12, 2016 (<http://www.newgeneralservicelist.org>).
- Cameron, L. 2001. *Teaching Languages to Young Learners*. Cambridge, UK: Cambridge University Press.
- Catramado, M. M. 2004. “Integrating the Six Skills in Every ESL/EFL Class.” Retrieved November 02, 2016 (<https://goo.gl/QgtJRq>).
- Desi-Konsortium. 2013. *Teaching and Skills-Acquisition in German and English*. Weinheim, Germany: Beltz.
- Flowerdew, L. 2005. “An Integration of Corpus-Based and Genre-Based Approaches to Text Analysis in EAP/ESP: Countering Criticisms Against Corpus-Based Methodologies.” *English for Specific Purposes* 24(3): 321-332.
- Harris, A. J., and Sipay, E. 1990. *How to Increase Reading Ability*. 10th ed. White Plains, NY: Longman.
- Ito, K. 2000. *Elementary School Education. Ready? Go!*. Tokyo, Japan: G Yousei.
- Kuno, Y. 1999. *How About Starting Elementary School English Like This?* Tokyo, Japan: Sanseido.
- Lessard-Clouston, M. 2013. *Teaching Vocabulary*. Alexandria, VA: TESOL International Association.
- Nation, P. 2001. *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press.
- Read, J. 2000. *Assessing Vocabulary*. Cambridge, UK: Cambridge University Press.
- Saku, M. and K. Honda. 2004. “*Toward Developing Vocabulary Teaching Material for English Education at Elementary School.*” Presented at the 30th Japan society of English language education.
- Schmitt, N. 2008. “Review Article: Instructed Second Language Vocabulary Learning.” *Language Teaching Research* 12(3): 329-363.
- Schmitt, N., and McCarthy, M. 1997. *Vocabulary: Description, Acquisition and Pedagogy Cambridge*. UK: Cambridge University Press.
- Scott, M. 1999. *Word Smith Tools version 3*. Oxford, UK: Oxford University Press.
- Scott, M. 2004. *Word Smith Tools version 4*. Oxford, UK: Oxford University Press.
- Scott, M. 2012. *Word Smith Tools version 6*. Stroud, UK: Lexical Analysis Software.