

BESSH-16**General Vocabulary in Thai EFL University Students' Writing: A Corpus-Based Lexical Study**Anchalee Veerachaisantikul^{1*}, Sukanya Chootarut²*Rajamangala University of Technology Isan, Thailand*

Abstract

This study questions the use of general vocabulary in Thai EFL University Students' Writing as a part of larger scale task investigating. General vocabulary represents the most frequent words of English, which are useful for English language learners. The corpus-based lexical study had two main purposes: a) to carry out a list of general words used frequently in writing as performed by Thai EFL university students, and b) to compare the word list with the New General Service List (NGSL) (Browne, C., Culligan, B. & Phillips, J., 2013) in order to settle their coverage in the TEFL writing corpus (TEFL Corpus). We compiled and investigated 1,233 writing tasks of Thai EFL university students comprising of 661,596 words. Our analysis acknowledged that, of 2,818 high frequency words in the NGSL, 1,648 appeared frequently in the corpus and the coverage accounted for 1.41% of the token in this corpus.

© 2016 The Authors. Published by Academic Fora. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Scientific & Review committee of BESSH- 2016.

Keywords— General Vocabulary, the New General Service List, Corpus-Based Lexical Study

Introduction

The significance of vocabulary in learning a second or foreign language is very important as claimed by many researches (Schmitt, 2008, Schmitt and McCarthy, 1997 and Read, 2000). Knowledge of vocabulary is a primitive element in language learning as it could be determined the level of a foreign language learner (Al-Desi, 2013). Clouston (2013) recommended that vocabulary is fundamental awareness to English language teaching without adequate vocabulary students cannot comprehend others or express their own ideas. Thus, vocabulary is needed as an aspect of all the four skills for communication (Harris, 1990). One important skill that concerns the vocabulary knowledge is writing as it is the ultimate skill for language learning. When learning a language, writing is an important skill to be accomplished as it acknowledges students find their own voices in their new language. It also allows them to communicate effectively in different contexts and with different audiences (Catramado, 2004). Yet, most of them cannot achieve this skill effectively. This is because writing is an advancing skill of discovering the practical words to write and share thoughts and feelings. Some researchers mentioned that the knowledge of vocabulary precisely affected on ESL or EFL learners' writing ability. As suggested by Bello (1997):

“Writing also enhances language acquisition as learners experiment with words, sentences, and larger chunks of writing to communicate their ideas effectively and to reinforce the grammar and vocabulary they are learning.”

This idea is supported by Al-Dersi (2013) that foreign language learners need the mature vocabulary knowledge in academic lives. One of the problems in English learning is an adequate size of vocabulary since the greater range in teaching higher level involve the greater knowledge of vocabulary as recommended by Nation (2001). Thus, it is essential to teachers to select vocabulary fits the level of learners. Some theoretical and practical researches in EFL guide that general words should teach to EFL learners as it can be highly valuable for them (Kuno, 1999, Ito, 2000, and Saku, 2004). Since research into Thai university students' writing has been poor. This does not mean; however, that there has been no research in this area. There are some linguistic analyses based on verb, linking adverbial, or connectors. However, there is insufficient attention to carry out a general word list. Consequently, the present study aims to examine the most frequently general word list in Thai university students' writing. The results obtained from this study may be developed further to be used as EFL materials for Thai students.

Objectives of the Study

*All correspondence related to this article should be directed to Anchalee Veerachaisantikul, Rajamangala University of Technology Isan, Thailand
Email: veeanchalee@hotmail.com

© 2016 The Authors. Published by Academic Fora. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Scientific & Review committee of BESSH-2016.

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 question:

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)?

Methods

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 have 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 words types, namely 598, 84, 729, 635, 2,788, and 2,970 in task B, C, D, E, F, and H.

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 serve 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. Aside 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 due to 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 is originating word family forms. Then, frequent general words used in the TEFL Corpus 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 order 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, C., Culligan, B. & Phillips, J., 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

To study the frequency and distribution of general words, a corpus of 143,027 words from Thai EFL university students' writing 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	%
1	Be	7,868	6.58	26	like	667	0.56
2	The	5,196	4.34	27	she	664	0.56
3	and	3,986	3.6	28	there	580	0.48
4	a	3,426	3.13	29	name	569	0.48
5	I	3,395	2.84	30	as	560	0.47
6	my	2,474	2.07	31	reason	552	0.46
7	in	2,445	2.04	32	up	550	0.46
8	to	1,947	1.63	33	gas	548	0.46
9	of	1,943	1.62	34	big	501	0.42
10	on	1,371	1.15	35	that	500	0.42
11	woman	1,370	1.15	36	two	493	0.41
12	have	1,367	1.14	37	very	478	0.40
13	for	1,170	0.98	38	we	477	0.40

Table 2
The Top 50 Most Frequently Occurring General Words of TEFL Corpus (cont.)

Rank	Word	F	%	Rank	Word	F	%
14	it	1,080	0.90	39	family	471	0.39
15	at	1,015	0.85	40	you	450	0.38
16	than	962	0.80	41	from	431	0.36
17	man	959	0.80	42	own	430	0.36
18	he	866	0.72	43	food	413	0.35
19	do	768	0.64	44	cost	411	0.34
20	but	752	0.63	45	roommate	411	0.34
21	year	741	0.62	46	sex	411	0.34
22	this	739	0.62	47	smart	411	0.34
23	live	694	0.58	48	weaker	411	0.34
24	shelf	685	0.57	49	well	377	0.32

25	old	672	0.56	50	interesting	344	0.29
----	-----	-----	------	----	-------------	-----	------

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 of 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 appeared 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 the 2nd rank appearing 48,575 times. From the data, “the” in the NGSL was the 1st rank appeared 60,910 times whereas “the” was the 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’ writing by using the Wordlist Tool which provided both alphabetical and frequency order 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 was obtained (Table2). It was found that high frequency 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 type as well since students tried to express their ideas and giving reasons covered by the topic. It can be explained that it may be a larger piece of writing for EFL students when they face 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 word general word most frequently occur 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 1 st June, 1991. Her favorite colors are white, pink,
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
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	studied English for fifteen years but he cannot speak
7	Nakhonratchasima for 6 years. She has	been	studied English for 17 years and she tries to
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	is	not difficult to learn but we have to try hard to
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	be	the next "cool" car to own for the future
15	My name is Gun. I	was	born on 2 nd September, 1992. I am a student at
16	She likes watching movies. She wants	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	stayed 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 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 concordances was used in "infinitive with to" form. For 13 concordances, it also can be seen that "be" was used as main verbs to represent existence, identity, color, mood, and negative sentences. This can be assumed that Thai EFL university students could be used the "verb to be" in appropriate ways.

Conclusion

The importance of vocabulary in foreign language learning in all subject disciplines as it is one of the understanding area 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 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 word 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 used in their daily life.

As many researches stated that corpus-based study have promoted as valuable sources for pointing out the practical information of writing in academic genres (Biber, Conrad & Reppen, 1998, and Flowerdew, 2005). The results of this study would be help teachers to realize the significance of vocabulary in writing and 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 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. In *31st Annual Teachers of English to Speakers of Other Language Convention, Orlando, FL, USA*.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Browne, C., Culligan, B., & Phillips, J. (2013). *The new general service list*. Retrieved from <http://www.newgeneralservicelist.org>.
- Cameron, L. (2001). *Teaching languages to young learners*. Ernst Klett Sprachen.
- 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., & Sipay, E. (1990). *How to increase reading ability* (10th ed.). White Plains, NY: Longman.
- Ito, K. (2000). *Shougakkou eigo: Redii. gou [Elementary school education. Ready? Go!]*. Tokyo, Japan: Gyousei.
- Kuno, Y. (1999). *Konna fuu ni hajimete mitewa? Shougakkou eigo [How about starting elementary school English like this?]*.
- Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Saku, M., & Honda, K. (2004). Shougakkou eigo kyouiku ni okeru goi kyouzai no kaihatsu ni mukete [Toward developing vocabulary teaching material for English education at elementary school]. In *The 30th Japan Society of English Language Education Conference Proceedings* (pp. 598-599).
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.
- Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy* (Vol. 2035). Cambridge: Cambridge university press.
- Scott, M. (1999). *WordSmith Tools version 3*. Oxford: Oxford University Press.
- Scott, M. (2004). *WordSmith Tools version 4*. Oxford: Oxford University Press.
- Scott, M. (2012). *WordSmith Tools version 6*. Stroud: Lexical Analysis Software.