

Profiling the Academic Wordlist in Research Project of EFL University Students

Anchalee Veerachaisantikul, Rajamangala University of Technology Isan, Thailand
Boonchai Srisassawatkul, Rajamangala University of Technology Isan, Thailand

The Asian Conference on Arts & Humanities 2017
Official Conference Proceedings

Abstract

The current study is a corpus-based lexical study that purposes to question the use of words in Coxhead's (2000) Academic Word List (AWL) in research in the field of linguistics and applied linguistics. We compiled and investigated 52 research projects of EFL university students comprising of 1,071,558 words, which called the Research Project Corpus (RP). Our analysis acknowledged that, of 570 word families in the AWL, 287 appeared frequently in the corpus and the coverage accounted for 4.08% of the token in this corpus. Furthermore, this study identifies non-academic word that appeared more than 50 times and examined high-frequency content words in the RP Corpus. The non-academic words are mostly nouns and adjectives that related to language education.

Keywords: Academic Word List, EFL, Research Project

iafor

The International Academic Forum
www.iafor.org

Introduction

In the field of English teaching and learning, one of the most important skills is writing which claimed by a number of linguists (Hyland, 2003, Kroll, 2003, and Matsuda, 2003). Writing is a complicate process as it is a skill that commonly learned by classroom practices. Students are referred to acquire good vocabulary knowledge, grammar, and understanding of register, genres, and styles in order to write appropriately. In regard to English writing contexts, there are mostly separated into two groups are English as a second language (ESL) and English as a foreign language (EFL) contexts. The difference between ESL and EFL context is that ESL context is used in everyday life activities while EFL context mainly used in academic environment like colleges, universities, and workplaces. This can be noted that it might directly involve how English writing instruction is taught by writing teachers and how it is learned by L2 writing leaners.

In Thailand, we learn English as EFL context, which is used in academic environment. Moreover, Thai EFL undergraduate students required to learn how to able to read and write in English; however, they still have a number of problems especially in writing. Undergraduate students who are taking degree for their study in an English medium colleges or universities emphasize on academic writing since they have to write academic papers, articles, theses or independent studies in English even their first language is not English.

Academic writing has established a relationship between vocabulary comprehension and academic success and classified characteristics to suggest the L2 word learners over academic tasks, although few people are recognized concerning student perceptions of academic vocabulary and the conscious decision-making process of these learners while they are writing (Nation, 2013). This idea is supported by Hyland (2007) that academic writing is usually referred a basic component and a special language requirement that are academic vocabulary and academic genres as special discourse have particular vocabulary (Coxhead and Byrd, 2007).

Moreover, Laufer & Nation (1995) suggested that academic writing requires lexical richness either the understanding to apply a proper size of high frequency and academic words. Using words in academic writing is hard to do even learners could be remembered the words and recognize them receptively. In spite of that it is more challenging and takes more time for learners in the task that need to apply receptive words to productive words because they have to practice than is often recognized (Laufer & Paribakht, 1998, Lee & Muncie, 2006, Webb, 2005).

Academic vocabulary is specialized vocabulary for second or EFL language learners who intended to do academic study in English. The best example of list of such words is the Academic Word List (AWL) (Coxhead, 2000) that consists of 570 word families that are not in the most frequent 2,000 words of GSL (West, 1953) and technical vocabulary; however, these words occur reasonably frequently over a wide range of academic text. As academic words are commonly used in academic writing than in our everyday conversational English thus EFL students who learn to write academic papers demand the development of an advanced linguistics foundation that covers academic vocabulary (Hinkel, 2001). This could be addressed that the ability to use vocabulary is important for EFL students in order to succeed in academic

English writing. Thus, vocabulary also plays an important role for students in learning English because of knowing sufficient vocabulary can help students to transfer meaning in communication. Although, there are some scholars mentioned that second and foreign language learners normally have finite vocabulary knowledge when comparing with native speakers particularly in speaking and writing (Kaur and Hegelheimer, 2005).

To help EFL students achieve academic writing, making word list are believed to be able to improve the vocabulary knowledge. So far, there are little attention has been paid to the academic writing and vocabulary of EFL English major students' research project. Consequently, the current study aims to investigate the most frequently academic word and non-academic words list in EFL university students' writing. This study will be profitable to writing teachers or course designers to develop their writing programs and improve their students' writing abilities according to the educational and professional requirements for English writing in the future.

Research Purposes

The purpose of this study is to explore the frequency of academic words that used in the field of linguistics and applied linguistics. The results showed in this study derive from an investigation of a greater corpus of research project of EFL university students with more than one million token words. The benefit of larger corpus is that the researchers could design a broader list of academic words, which are oftentimes employed in EFL research projects. Also, the current study analyzes the frequency of non-academic words that is frequently occurred in EFL research projects. The following research questions are addressed in this paper:

1. What are the most frequently used academic words in research project of EFL university students?
2. What non-academic words appeared with high frequency in research project of EFL university students?

Materials and methods

This study combines corpus-based and genre-based approaches, studying the research project to expose specific features of academic vocabulary. This corpus-based study concentrates on frequency, coverage, and distribution of the words from the Academic Word List in Thai EFL university students' research project in the whole project.

The Compilation of the Corpus

For the study, we built a 1,071,558-word corpus of research projects that was formed following the criteria recommended by Sinclair (1990). The corpus contains 52 research projects produced academic working written by EFL English major students who enrolled in an 'Independent Study' course which is a partial fulfillment of the requirements at their degree. These research projects were carried out in the last semester of their fourth year and selected from 2010 to 2014 academic year. The research projects concerned with their major areas in the discipline of linguistics and applied linguistics and the section organization of their works were followed a

prevalent format in seven main sections (Abstract, Introduction, Literature Review, Method, Results, and Discussion and Conclusion, and Reference). In order to answer the purpose of the study, the academic word list of the corpus demonstrated Table 1 gives general information of the data used for the Research Project Corpus (RP).

Table 1: General Information of the RP Corpus

Academic Year	No. of Research Projects	Token words (Running words)
2010	12	222,711
2011	10	211,336
2012	10	212,450
2013	10	212,402
2014	10	212,659
Total	52	1,071,558

The Software for Data Analysis

In the current study, we mainly calculated on the computer software program “WordSmith Tool Version 6” (Scott, 2012) for the lexical analysis and profiling. It is an integrated suite of programs for looking at how words behave in texts and used to find out how words were used in any kind of texts. In addition, this software program can create output that serves data as regard to the frequencies and distribution of the academic words in the corpus.

Procedure and Data Analysis

The process of data analysis can be outlined in Figure 1 that are basically dealt with how data was selected, how selected texts were managed for academic words and non-academic words analyzed.

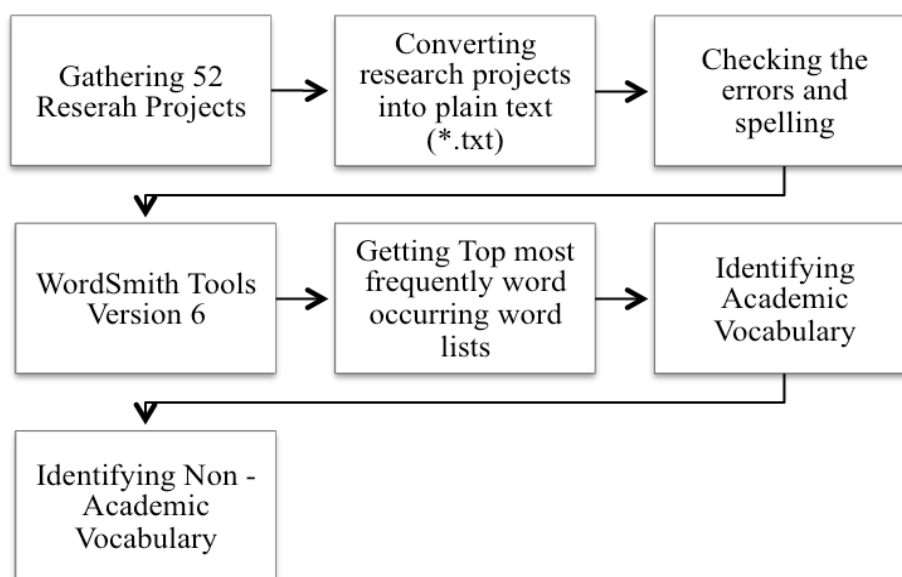


Figure 1: Research Procedures

According to the purposes of the study, the 52 research projects written by EFL English major students were selected as the most important input for the Research Project Corpus (RP) between 2010 and 2014. The first research question aims to probe the frequency and distribution of the academic words. To match this question, the first stage is to acquire a list of academic words that serve the criteria in the study that cover frequency and range. The criteria of the study is that the academic word should appear at least 50 times in the complete RP Corpus. In addition, all text files were saved as a PDF document thus they were converted into plain text (*.txt). Then, the errors and spellings of all the words were examined and revised before using the lexical analysis program. After checking the errors and spellings of all files, the computer software program “WordSmith Tool Version 6” was employed in order to make the word frequency lists of each file by using the Wordlist Tool. The Wordlist Tool offered both alphabetical and frequency order of the words in the text files. Afterwards, the most frequently occurring words were obtained, the frequency word list was screened by the expert in academic field and checking the entries with Coxhead’s (2000) Academic Word List manually. This resulted in a list of top 50 most frequently occurring academic words.

The second research question aims to explore non-academic words that frequently occurred in this corpus excluding in Coxhead’s (2000) Academic Word List (AWL) and General Service List (GSL) (West, 1953). To answer this question, we adapted the similar criteria of the frequency and range and chose only non-academic words include only content which appear at least 50 times in the whole RP Corpus.

Results

The current study concentrated on the frequency, coverage, and distribution of academic words in the Research Project Corpus (RP Corpus). Consequently, a corpus of 1,071,558 running words from EFL English major students’ research projects has been applied. After analyzing the data, following results have been achieved that demonstrated to answer the two posted research questions below.

Research Question 1: What are the most frequently used academic words in research project of EFL university students?

Table 2: The Top 50 Most Frequently Academic Words (AWL) of the RP Corpus

Rank	Word	F	%	Rank	Word	F	%
1	source	3,687	0.34	26	version	476	0.03
2	target	3,552	0.33	27	factor	471	0.04
3	text	2,712	0.26	28	liberal	456	0.04
4	strategy	2,406	0.22	29	topic	447	0.04
5	data	2,373	0.22	30	category	438	0.02
6	research	1,944	0.18	31	reveal	414	0.03
7	technology	1,548	0.14	32	create	408	0.03
8	item	1,479	0.14	33	design	399	0.03
9	chapter	1,428	0.13	34	investigate	396	0.03
10	analyze	1,329	0.13	35	concept	381	0.03
11	attitude	1,311	0.12	36	percent	372	0.03
12	process	1,215	0.11	37	expert	351	0.03

13	found	1,203	0.11	38	publish	336	0.03
14	identify	975	0.09	39	objective	321	0.03

Table 2: The Top 50 Most Frequently Academic Words (AWL) of the RP Corpus (Cont.)

Rank	Word	F	%	Rank	Word	F	%
15	project	816	0.07	40	area	315	0.03
16	structure	791	0.07	41	positive	312	0.03
17	focus	783	0.06	42	significant	306	0.03
18	culture	711	0.07	43	similar	303	0.03
19	final	606	0.06	44	criteria	288	0.03
20	specific	606	0.06	45	tense	282	0.02
21	style	600	0.06	46	obtain	270	0.02
22	major	573	0.04	47	transfer	270	0.01
23	select	537	0.04	48	define	264	0.01
24	aspect	507	0.04	49	technique	264	0.01
25	occur	492	0.03	50	appropriate	255	0.01

Total of frequency of occurrences 42,979
Total of % of text coverage 3.79%

Table 2 provided the list of the top 50 most frequently academic words in the corpus. The total of frequency of occurrences in this table was calculated from 1st rank to 50th rank. There was a total of 42,979 occurrences that accounted for 3.79% of text coverage, of the whole corpus which was 100%. However, most of them occurred often to be included in the top 50 most frequently occurring words of the whole RP Corpus. In the list, the top five high frequency words were “source” 3,687 times, “target” 3,552 times, “text” 2,712 times, “strategy” 2,406 times, and “data” 2,373 times. After comparing with the headwords of Academic Word List (Coxhead, 2000), the findings presented that of 570 headwords of the academic word list, 287 appeared frequent in this corpus and the coverage accounted for 4.08% of the token in the corpus.

Taking source as an example, the corpus has 3,687 hits for this word. The Oxford Learner’s Dictionary has four specialized meanings for this word that are “a place, person or thing that you get something from”, “a person, book or document that provides information, especially for study, a piece of written work or news”, “a person or thing that causes something, especially a problem”, and “the place where a river or stream starts”. Comparing the meanings from five concordances from the RP Corpus with those in different context.

N

1	this strategy, the whole meaning of a	<i>source</i>	source language word could be fully ...
2	of translation relationship between a	<i>source</i>	source system and a target system ...
3	clearly. In the classrooms, listening is a	<i>source</i>	source of English communication such as ...
4	in particular contexts. Vocabulary is also	<i>source</i>	a source of polite term. It is imperative that ...
5	will have a complete command of the	<i>source</i>	source and receptor languages, as a

Research Question 2: What non-academic words appeared with high frequency in research project of EFL university students?

Table 3 The Top 50 Most Frequently Non-Academic Words of the RP Corpus

Rank	Word	F	%	Rank	Word	F	%
1	word	12,816	1.20	26	explanation	1,209	0.11
2	English	11,022	1.03	27	pronunciation	1,197	0.11
3	student	8,997	0.84	28	people	1,188	0.11
4	translation	8,286	0.77	29	information	1,179	0.11
5	language	8,154	0.76	30	loan	1,164	0.11
6	year	3,201	0.30	31	package	1,161	0.11
7	test	3,189	0.30	32	different	1,158	0.11
8	problem	3,120	0.29	33	total	1,131	0.11
9	translator	2,937	0.27	34	number	1,125	0.10
10	level	2,580	0.24	35	general	1,098	0.10
11	Thai	2,502	0.23	36	online	1,080	0.10
12	vocabulary	1,992	0.19	37	important	1,032	0.10
13	example	1,983	0.19	38	neutral	1,026	0.10
14	source	1,944	0.18	39	paraphrase	1,023	0.10
15	communication	1,827	0.17	40	correct	993	0.09
16	university	1,821	0.17	41	learner	954	0.09
17	second	1,752	0.16	42	academic	951	0.09
18	comprehension	1,686	0.16	43	result	939	0.09
19	frequency	1,650	0.15	44	high	933	0.09
20	ability	1,560	0.15	45	skill	933	0.09
21	knowledge	1,431	0.13	46	sentence	837	0.08
22	international	1,425	0.13	47	synonym	831	0.08
23	analysis	1,389	0.13	48	corpus	825	0.08
24	result	1,356	0.13	49	questionnaire	825	0.08
25	group	1,305	0.12	50	sample	822	0.08

Total of frequency of occurrences **115,539**
Total of % of text coverage **10.8%**

From the data shown in Table 3, it can be seen that the top 50 most frequently occurring non-academic words accounted for 10.8% of the text coverage and there were a total of 115,539 occurrences. From the list, the top five high frequency words were “word” 12,816 times, “English” 11,022 times, “student” 8,997 times, “translation” 8,286 times, and “language” 8,154 times.

According to Coxhead (2000), she assumed that general word list (GSL) should be known as a considerable element of the acquired knowledge of language learners. In addition, GSL appearing in academic context were employed as a part of academic argument but on the other hand they were not involved as academic words. Table 4 compares the frequency of the top 10 most frequent academic words and the frequency of the top 10 most frequent non-academic words in the RP Corpus. It can be observed that the words from the non-academic words with academic use in the RP Corpus had a higher frequency than the first 10 words from the academic words in this study. Furthermore, the RP Corpus contains diverse nouns and adjectives that applied in a field that is closely related to language education, for instance, word,

English, translation, and language. These words appear frequently may be resulting from the diversity of researches that were the subject of linguistics.

Table 4 Frequency of the top 10 most frequently academic word and the non-academic word

Words from the AWL				Words from the Non-AWL		
		F	%		F	%
1	source	3,687	0.34	word	12,816	1.20
2	target	3,552	0.33	English	11,022	1.03
3	text	2,712	0.26	student	8,997	0.84
4	strategy	2,406	0.22	translation	8,286	0.77
5	data	2,373	0.22	language	8,154	0.76
6	research	1,944	0.18	year	3,201	0.30
7	technology	1,548	0.14	test	3,189	0.30
8	item	1,479	0.14	problem	3,120	0.29
9	chapter	1,428	0.13	translator	2,937	0.27
10	analyze	1,329	0.13	level	2,580	0.24
Total		22,458		Total	64,302	

Discussion

The current study has provided a list of academic words that are frequently used in research project according to a corpus analysis of Thai EFL university students. Also, it serves a list of non-academic content words which are frequently used in the Thai EFL research projects. 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 the RP Corpus by using the Wordlist Tool which provided both alphabetical and frequency order of the words in the text files.

Our objectives of making a list of academic and non-academic words in Thai EFL research projects is to reveal English teachers or academic English teachers or instructors to realize the importance of academic vocabulary. Furthermore, Ellis (1990) suggested that the courses focus on language feature often cause to better learning than unplanned courses. For this reason, we trust that the results of the current study could be useful for the teaching and learning of frequently occurring academic word and non-academic words as it can support students to promote not only their academic writing but also reading skill. In addition, the frequency-based wordlists can help students of English for Academic Purpose (EAP) who need enlarge their vocabulary size by notifying which words they should master (Cobb, n.d.). Besides, teachers could apply the word lists presented in this study to boost the students’ awareness and guide them to employ in their writing. Moreover, teachers could teach the linguistic features to students such as tenses, parts of speech, or collocations as suggested by Coxhead (2008). Yet, linguist scholars who interested in doing research in academic word list could merge these findings for further studies on other fields and teachers or course designers of English for Academic Purposes (EAP) who wish to develop their teaching material could apply these findings of the study for their teaching AWL as well.

We expect that this study has exhibited the importance of creating academic and non-academic word list to point out learners to get more idea in order to prepare themselves to use better words when they write the academic research and to comprehend the academic texts or articles that they have to read. Moreover, it would be the motivation for many instructors, education programmers, and textbook developers throughout the world.

Conclusion

The significant of English as a largest messenger of international globalization has influenced English to achieve a great role in all subject disciplines in Thailand. The current study concentrates on academic English and goal to decrease the problem of Thai EFL students when writing academic researches and articles. Additionally, the results would be a new findings that could be profitable to learners of academic English in order to achieve their higher education, teachers or course designers of academic English when they write or develop their material for teaching English for Academic Purposes (EAP), and linguistics scholars who interested in pursuing further study in applied linguistics. For teachers, they could use these results not only to improve their classroom teaching for vocabulary but also grammar.

Another important investigation is that the non-academic word that are highly frequent in our corpus. In our EFL context, it could be assumed that EFL learners have good knowledge of non-academic word since they can apply their knowledge to specific topic of linguistics; however, they are often forgetful of grammatical rules. This idea is supported by Hyland and Tse (2007) that an academic vocabulary usually comes after a general vocabulary. With respect to the most frequent occurring non-academic word list, teachers can employ the frequently occurring word lists as a benefit source to instruct words that students are not familiar with their study since the frequently word list is essential supplement to various disciplines as recommended by Coxhead (2000). Still, there is much remains to be done, our work creates valuable findings in the field of academic English. In other words, having accepted that this study has some limitations, which have to be mentioned. The main limitations of this study are suggested as follows: the first limitation concerns a corpus size of this study that were smaller than the Coxhead's (2000) Academic Word List since the RP Corpus contained over one million token words. Consequently, future research might need to broaden the size in order to generalize their findings and to see whether their results would be related to ours. The second limitation is that this study gathers only the frequently occurring academic and non-academic word; however, the linguistic features that are issues concerned with actual word, are outside limits such as collocation, grammatical features, or discourse context.

References

- Cobb, T. (n.d.). Why and how to use frequency lists to learn words. Retrieved May 20, 2016 from [http://www.lex tutor.ca./reserach /](http://www.lex tutor.ca./reserach/).
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Muenier & S. Granger (Eds.), *Phraseology in foreign language and teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Ellis, R. (1990). *Instructed second language acquisition: Learning in the classroom*. Oxford: Basil Blackwell.
- Hinkel, E. (2001). *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. New York, NY: Routledge.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Hyland, K. (2007) Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16, 148-164.
- Hyland, K. & Tse, P. (2007). Is there an “Academic vocabulary.” *TESOL Q.* 41(2), 235-253.
- Kaur, J. & Hegelheimer, V. (2005, October). ESL students’ use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4), 287-310.
- Kroll, B. (Ed.). (2003). *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Laufer, B., & Paribakht, S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48, 365-391.
- Lee, S. H., & Muncie, J. (2006). From receptive to productive: Improving ESL learners’ use of vocabulary in a postreading composition task. *TESOL Quarterly*, 40(2), 295-320.
- Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 15-34). Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge, England: Cambridge University Press. Oxford Advanced Learner's Dictionary (8th ed.). (2010). Oxford, England: Oxford University Press.

Scott, M. (2012). WordSmith Tools version 6. Stroud: Lexical Analysis Software.

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

Webb, S, (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52.