

New Academic Word List in EFL University Students' Research Projects

AnchaleeVeerachaisantikul^{*a}, BoonchaiSrisassawatkul^b, WaraChansin^c

^aEnglish for International Communication Department, the Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan, Nakhonratchasima 30000, Thailand.

^bEnglish for International Communication Department, the Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan, Nakhonratchasima 30000, Thailand.

^cEnglish Department, the Faculty of Humanities, Naresuan University, Phitsanulok, 65000, Thailand.

*Corresponding Author: anchalee.ve@rmuti.ac.th

Abstract

Enhancing academic vocabulary skill is one of the important ways of any language study at university level. In order to become a competent English language learner, the need of plentiful vocabulary to encounter the challenges of applying English in academic contexts. The current study is a corpus-based study that designs to question the use of words in Gardner and Davie's New Academic Word List (AVL) (2013) in research in the field of linguistics and applied linguistics. We gathered and probed 60 research projects of EFL university students composing of 1,071,558 running words that called the Research Project Corpus (RP). Our analysis acknowledged that the top of 500 word families in the AVL, 474 words turned up frequently and the coverage accounted for 14.04% of the token in this corpus. Yet, the current study also classifies the non-academic words that appeared more than 50 times and examined high frequency in both function and content words that related to language education.

Keywords: New Academic Word List, Academic Vocabulary, EFL, Research Project

Introduction

English is a universal language which is acknowledged internationally as of 337 million people speak English as their first language and 235 million people speak English as their second language (Crystal, 2003). It is not exclusively used as a way of expressing thoughts and ideas but also used as a medium in order to forge cultural ties, and diplomatic relations, education, and economy. For language learning, English is very crucial for many purposes such as professional purposes, specific

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

purposes, or academic purposes. Furthermore, one of the most significant skills in learning English is writing as it is a complicated process. Due to it is a common skill that learned by classroom systematic exercises as recommended by many linguists (Hyland, 2003; Kroll, 2003; Matsuda, 2003).

In writing, students are referred to reach their goal with great vocabulary knowledge, grammar and acquaintance of register, genres, and styles in order to write appropriately. With regard to English writing contexts, students could be separated into two groups that are English as a second language (ESL) and English as a foreign language (EFL) context. For ESL context, it is usually in everyday life activities whereas EFL context mostly used in academic environment such as workplaces, universities and colleges. 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 learners.

In Thailand, English is taught as EFL context that is used in academic environment. Thai EFL university students are required to learn how to be able to read and write in English effectively. On the other hand, they still have a number of problems particularly in writing. In addition, university students who are taking degree for their study in colleges or universities that underlines on academic writing, have to write academic theses, independent studies, papers, or research articles in English even their first language is not English.

The academic writing is commonly cited as a basic part and specific language requirement. It is related to academic vocabulary and genres because of its special discourse, has special vocabulary (Hyland, 2007; Coxhead & Byrd, 2007). One genre of academic writing that represents the preferred medium of exchanging and advancing knowledge is research project. Research project is an expansion on the past work in the field of study and it can be applied to promote further knowledge on a topic. Therefore, the ability to write academic writing effectively is not only based on linguistics ability but also on an awareness of the rhetorical features used in academic writing as accepted by the discourse community. Because of the academic words are generally used in academic writing than in our everyday conversational English likewise native speakers who want to succeed in academic contexts, they need to have adequate vocabulary in order to meet the challenges (Beck, Perfetti, & McKeown, 1982; Biemiller, 1999; Biemiller & Slonim, 2001). Thus, EFL students who learn to write academic papers need the knowledge of an advanced linguistics foundation that covers academic vocabulary for accomplishing their goal (Hinkel, 2001). Similarly, Hyland (2007) suggested that academic writing is usually referred a basis component and a special language requirement that are academic vocabulary

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

and academic genres since the special discourses have particular vocabulary (Coxhead & Byrd, 2007). It can be addressed that the competence of using vocabulary is very essential for EFL students to succeed their academic English writing. However, some scholars mentioned that second language students regularly have limited vocabulary knowledge when comparing with native speakers especially speaking and writing (Kaur & Hegelheimer, 2005). This is because the lack of vocabulary could cause to many EFL students feel incapable to use English in academic settings as well as English language teachers regularly do not know what words they should teach their students. Accordingly, the idea that considered as one of the important ways for improving the vocabulary knowledge and assisted EFL students to be achieved in academic English writing is making a wordlist.

Yet, there are little considerations in the academic writing and vocabulary of EFL university students in writing research articles, projects, or articles. Moreover, many researchers in applied linguistic and English language teachers are still having the problem that there is not a core set of academic vocabulary that spans the disciplines (Gardner & Davies, 2014)

For these reasons, the current study aims to explore the most frequently academic and non-academic word list in EFL university students' academic writing. This study will be profitable to writing teachers or course designers to develop their writing teaching materials and improve their students' writing abilities according to the educational and professional requirements for English writing in the future thus the following research questions are addressed in this paper.

Research Questions

1. What are the most frequently used academic words in research projects of EFL university students?
2. How different are they from those in the New Academic Wordlist (AVL)?
3. What are the most frequently used non-academic words in research projects of EFL university students?

Methodology

The Compilation of the Corpus

In this 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 60 research projects produced academic working written by EFL English major students who enrolled in an 'Independent Study' course which is a partial fulfilment of the requirements at their degree. These research projects were carried out in the last

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

semester of their fourth year and selected from 2010 to 2017 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 research questions 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).

	RP Corpus
Running Words	1,071,558
Sentences	51,531
Mean sentence length	18.89
Frequency of academic words	150,557
% of academic words Coverage	14.04

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. Moreover, this software program can create output that serves data as regard to the frequencies and distribution of the academic words in the corpus.

The New Academic Wordlist (AVL)

Academic Vocabulary is a word that is normally used in academic dialogue and text. It is not fundamentally prevalent or generally meets in informal conversation. The most famous of academic word is the Academic Wordlist (AWL) (Coxhead, 2000) that has been used in English for Academic Purposes (EAP) setting for over 15 years (Newman, 2016).

Nevertheless, this study selected the New Academic Wordlist (AVL) of Gardner and Davies (2013) to compare the range, frequency, and distribution with the RP Corpus. The reasons for using the AVL are that the compilation of its was quite different from all its predecessor lists, and the top 500 word families achieved impressively higher lexical coverage than the AWL in academic corpora as the authors claimed. The AVL has been also denoted to be a better list than the AWL by its authors thus it is deserved to explore further.

The New Academic Vocabulary List (AVL) is created by Gardner and Davies in 2013. The AVL developed from a 120-million-token corpus which contains 3,015

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

lemma-headwords and 1,991 word families and covers 14 percent of the Corpus of Contemporary American English (COCA) academic sub-corpus and the British National Corpus (BNC) academic sub-corpus. It covers 9 academic disciplines in education, humanities, history, social science, law and political science, science and technology, medicine and health, business and finance, philosophy, religion, and psychology that take from academic journals, magazines, and newspapers. Additionally, the AVL is claimed that it is suitable for academic English learners. The word selection principle of the AVL is the range, frequency, and dispersion should be at least 50 percent higher in frequency in academic corpus than in non-academic corpus and at least 20 percent of the expected frequency in at least 7 of the 9 academic disciplines.

The ideals of creating the AVL, the authors determined the word by using lemmas not word families then the authors grouped the list into families in order to make sure for the certain instructional and research purposes that make it differ from the AWL. Further, it applied both frequency and dispersion statistics in order to balance the corpus that consists academic and non-academic materials. Gardner and Davies (2014) also suggested that the corpus is required broad enough and the effective statistics for sorting academic core words from general high frequent words and academic technical words.

Procedures and Data Analysis

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

Step 1	Gathering 60 reserach projects
Step 2	Converting reserach projects into plain text (*.txt)
Step 3	Checking the errors and spellings
Step 4	WordSmith Tool Version 6
Step 5	Getting top most frequently word occruing word list
Step 6	Identifying Academic Vocabulary
Step 7	Comparing RP Corpus with the AVL

Figure 1. Research Procedures

According to the purposes of the study, the 60 research projects written by EFL English major students were selected as the most important input for the RP Corpus between 2010 and 2017. 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 serves the criteria in the study covering the frequency and range. The criterion of the study is that the academic words 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 the AVL (2014) Academic Word List manually. This resulted in a list of top 30 most frequently occurring academic words. The second research question is to compare the RP Corpus with the AVL (Gardner & Davies, 2013). In order to answer this question, we adapted the top 500 word families of the AVL as criteria for classifying the differences. Finally, we also explore the non-academic words that used frequently in the RP Corpus. For the non-academic words, we select both function and content words that appeared at least 50 times in the whole

RP Corpus.

Findings

The current study attempted to examine the frequency, coverage, and distribution of academic words in the RP Corpus. Then, a corpus of 1,071,558 running words from EFL university students' research projects has been applied. This section is structured to answer the three posted research questions.

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

Table 1

The Top 30 Most Frequently Academic Words in RP Corpus

Rank	Word	F	Rank	Word	F
1	language	8,646	5	source	3,687
2	meaning	4,068	6	level	3,147
3	research	3,927	7	table	2,832
4	test	3,762	8	example	2,793

Table 1

The Top 30 Most Frequently Academic Words in RP Corpus(Cont.)

Rank	Word	F	Rank	Word	F
9	use	2,772	20	analysis	1,461
10	mean	2,727	21	knowledge	1,431
11	researcher	2,577	22	international	1,425
12	strategy	2,406	23	attitude	1,311
13	data	2,391	24	explanation	1,308
14	result	2,316	25	finding	1,296
15	relate	1,884	26	form	1,254
16	communication	1,851	27	present	1,209
17	university	1,821	28	form	1,026
18	group	1,734	29	information	1,179
19	technology	1,524	30	total	1,131

Table 1 provided the list of the top 30 most frequently academic words in the corpus. The total of frequency of occurrences in this table was calculated from 1st rank to 30th rank. There was a total of 70,896 occurrences that accounted for 6.61% of text coverage, of the whole corpus which was 100%. Still, most of them occurred often to be included in the top 30 most frequently occurring words of the whole RP Corpus. In the list, the top five high frequency words were “language” 8,646 times, “meaning” 4,068 times, “research” 3,927 times, “test” 3,762 times, and “source”

3,687 times.

Research Question 2: How different are they from those in the New Academic Wordlist (AVL)?

Table 2

The Top 10 Most Frequently Academic Words in RP Corpus and the AVL

Rank	Word	F	Word	F
1	language	8,646	study	137,208
2	meaning	4,068	develop	128,974
3	research	3,927	group	125,012
4	test	3,762	system	116,141
5	source	3,687	relate	114,267
6	level	3,147	research	112,649
7	table	2,832	social	103,635
8	example	2,793	result	96,016
9	use	2,772	use	93,271
10	mean	2,727	provide	93,212
Total		21,720		1,120,385

The comparison of our words with the AVL presented that academic words in the RP Corpus were generally similar to the AVL; however, there were some words occurred with high frequency and their frequency order was different from the AVL. For example, the most frequent academic word occurring in the RP Corpus was “research” with 3,927 occurrences while in the AVL “research” was the 6th rank appearing 112,649 times. From the data, “use” in the RP Corpus and the AVL was at the same rank (9th) but there were different frequencies.

Research Question 3: What are the most frequently used non-academic words in research project of EFL university students?

Table 3

The Top 20 Most Frequently Non-Academic Words in RP Corpus

Rank	Word	F	%
1	the	74,451	6.95
2	of	39,579	3.69
3	be	31,740	2.97
4	and	24,654	2.30
5	in	23,328	2.18
6	to	21,825	2.04
7	word	12,816	1.20
8	a	12,789	1.19
9	English	11,022	1.03

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

Rank	Word	F	%
10	that	9,732	0.91
11	for	9,528	0.89
12	student	8,997	0.84
13	by	8,964	0.84
14	translation	8,286	0.77
15	as	5,652	0.53
16	this	5,589	0.52
17	it	4,920	0.46
18	or	4,710	0.44
19	have	4,575	0.43
20	from	4,479	0.42

Total of frequency of occurrences 327,636

Total of % of text coverage 30.6%

From the data shown in Table 3, it can be seen that the top 20 most frequently occurring non-academic words accounted for 30.6% of the text coverage and there were a total of 327,636 occurrences. From the list, the top five high frequency words were “the” 74,451 times, “of” 39,579, “be” 31,740, “and” 24,654, and “in” 23,328. According to the word list, it can be seen that the function words occurred more frequent than the content words.

From the table above, it also can be said that the non-academic words should be as an extensive element of the required knowledge of language learners as well. Since, these words appearing in academic context were employed as a part of academic argument; on the other hand, they were not involved as academic words as suggested by Coxhead (2000).

Conclusion

The particular objectives of the current study concentrated on the analysis of vocabulary found in the RP Corpus. In addition, the results of the study could be used as a principle consideration for material design for EFL course development in the future. Thus, a corpus-based study and concordancing software were used for the investigation of academic and non-academic words. In the current study, we examined a 1,071,558-word corpus of research projects in a specific discipline, i.e. linguistics and applied linguistics. We tried to develop a word list for EFL university students. Additionally, we focused on academic English and aimed to lessen the problems of Thai EFL university students when writing academic research articles, projects, or journals.

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

The first research question of the study was to explore what words from the AVL are found in the RP Corpus. To answer this question, the concordancing software “WordSmith Tools version 6” was used. This concordancing software was used to see how words behave in texts. It was used in order to make the word frequency lists of each research project file by using the Wordlist Tool that offered both alphabetical and frequency order of the words in the text files. Then, the most frequently occurring words of the RP Corpus was obtained. Later, the academic words were screened by the expert and checking the entries with the AVL. The finding found showed that most of the academic words were noun. The top five high frequency words were “language”, “meaning”, “research”, “test”, and “source”.

Next, the second research question was to investigate what AVL words were shared in the RP Corpus. In total, we identified 474 word families used with a reasonable frequency in our corpus. We called these words Research Project Corpus (RP Corpus). A comparison of our words with the AWL displayed that numerous AVL items were used many times in the subject area we investigated. Some words appeared with high frequency; however, their frequency rank was different from the academic words in Gardner and Davies' Wordlist (2013). This could be implied that academic words are used in the same way across disciplines and administrators, teachers, curriculum designers, and English language learners should be considered these words for the academic purposes. The AVL had very often word families (474) and many of these words also occur to be core vocabulary. This could be suggested that in order to accomplish in academic contexts both in reading and writing, mastering the whole AVL may be more advantageous for students. Lastly, the third research question was to discover the most frequently non-academic in this corpus. From the analysis of the study, the top 20 most frequently occurring words of the RP Corpus was obtained. It was found that high frequency words in the corpus were prepositions, conjunctions, articles, and the verb to be. For example, the top five high frequency words were “the”, “of”, “be”, “and”, and “in”. The findings revealed that the non-academic words are highly frequent both function and content words. In addition, the function words appeared more frequent than the content words. This is related to Hyland and Tse (2007) expressed that academic vocabulary regularly follows a general vocabulary.

From the above mentioned, teachers could apply the frequently appearing word lists as a worthy source to guide words to their students who are not acquainted with their studies as the frequency word list is a worthy supplement to various disciplines (Coxhead, 2000). Likewise, the findings of this study would be not only a new finding that is valuable to academic English learners to reach their higher education but also

to teachers or course designers of English for Academic Purpose (EAP) and linguistic scholars who interested in pursuing further study in applied linguistics. For teachers, they could use these results to enhance their classroom teaching for vocabulary and grammar.

References

- Beck, I. L., Perfetti, C., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Biemiller, A. (1999). *Language and reading success* (Vol. 5). Cambridge, MA: Brookline Books.
- Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth innormative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93, 498–520.
- 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
- Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.
- Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, 35, 305–327.
- 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.
- 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.
- 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.
- Newman, J.A. (2016). *A corpus-based comparison of the academic word list and the*

NEW ACADEMIC WORD LIST IN EFL UNIVERSITY STUDENTS'

academic vocabulary list. Brigham Young University.

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

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