

Academic Words in Engineering Research Article: A Corpus Study

Anchalee Veerchaisantikul, Sukanya Chootarut, and Assist. Prof. Ladawan Juajamsai

Abstract—The present study is a corpus based lexical analysis which attempts to explore the word frequency and text coverage of Coxhead's (2000) Academic Word List (AWL) in research articles of engineering. In order to achieve the aims of the study, 451,236 words corpus was compiled comprising of 30 research articles written by EFL graduate students. The analysis acknowledged that 212 of 570 word families in the AWL occurred frequently in the corpus and coverage accounted for 4.33% of the token words. The present study was also made to identify the most frequent word families of AWL in the Engineering Research Articles Corpus (ERA Corpus) as well as to gain a list of non-AWL words which is not included in General Service List (West, 1953).

Index Terms—Academic Word List, Research Article, Corpus Study

I. INTRODUCTION

English is a universal language that 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 solely 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. In language learning, English is very important for many purposes such as academic purposes, specific purposes, or professional purposes. Moreover, one of the most significant skills in learning English is writing. Since, writing is a difficult process because it is a common skill that learned by classroom practices as suggested by many linguists (Hyland, 2003, Kroll, 2003, and Matsuda, 2003), In writing, students are addressed to achieve great vocabulary knowledge, grammar and acquaintance of register, genres, and styles in order to write properly. With regard to English writing contexts, they could be divided into two groups that are English as a second language (ESL) and English as a foreign language (EFL) context. ESL context usually used in everyday life activities whereas EFL context mostly used in academic environment such as workplaces, universities and colleges.

In Thailand, English is taught as EFL context that is used in academic environment. Thai EFL undergraduate and graduate are required to learn how to be able to read and write in English effectively; on the other hand, they still have many problems particularly in writing.

For graduate students who are taking degree for their study in colleges or universities that highlights 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, which are academic vocabulary and genres because of its special discourse, has special vocabulary (Hyland, 2007, Coxhead, and Byrd, 2007). One genre of academic writing which represent the preferred medium of exchanging and advancing knowledge is research articles (Flowerdew, 1999). 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. Since the academic words are generally used in academic writing than in our everyday conversational English thus EFL students who learn to write academic papers need the knowledge of an advanced linguistics foundation that covers academic vocabulary (Hinkel, 2001). It can be noted that the competence of using vocabulary is very essential for EFL students to succeed their academic English writing. However, some scholars addressed that second language students regularly have limited vocabulary knowledge when comparing with native speakers especially speaking and writing (Kaur and Hegelheimer, 2005).

Thus, in order to help EFL students to be achieved academic English writing, making a wordlist is considered as one of important ways for improving the vocabulary knowledge. Yet, there are little considerations in the academic writing and vocabulary of EFL graduate students in writing research articles. Therefore, the present study aims to explore the most frequently academic word and non-academic words list in EFL graduate students' 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 present study is aimed to reach the purposes as follow:

1. To profile the most frequently used academic words in research articles written by EFL graduate students in mechanical engineering field.

2. To explore the most frequently used non-academic words in research articles written by EFL graduate students in mechanical engineering field.

Rajamangala University of Technology Isan, Thailand

II. PROCEDURES

Corpora

The compilation of the present study was gathered from scientific articles belonging to one principal area of scientific knowledge: the field of engineering. The corpora consisted of a corpus of 30 research articles written by EFL graduate engineer students. The amount of items selected in the corpus was 451,236 running words. The 30 research articles written by EFL graduate students were all compiled from the international journal and the proceeding of international conference in the specific field of Mechanic Engineering. The section organization of their research articles were followed a general format in all seven main sections (Abstract, Introduction, Literature Review, Method, Results, and Discussion and Conclusion, and Reference). In order to meet the research purposes, the academic word and non-academic word list of the corpus represented the Abstract, Introduction, Methodology, Results, and Discussion section followed by Swales (1999). Table 1 gives an overview of the data used for the EFL graduate students.

TABLE I:
AN OVERVIEW OF THE DATA

	RAs Corpus
Running Words	451,236
Sentences	3,431
Mean sentence length	19.11
Frequency of academic words	4,598
% of academic words Coverage	1.01

Table 1 contains information on an overview of the ERA Corpus. The number of token words (running words) were 451,236 words and the mean of sentence length was 19.11. As it is seen, there was a total of academic words found in this corpus were 4,598 occurrences that for 1.01% of the whole corpus.

The Software for Data Analysis

In the present 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.

Procedures 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.

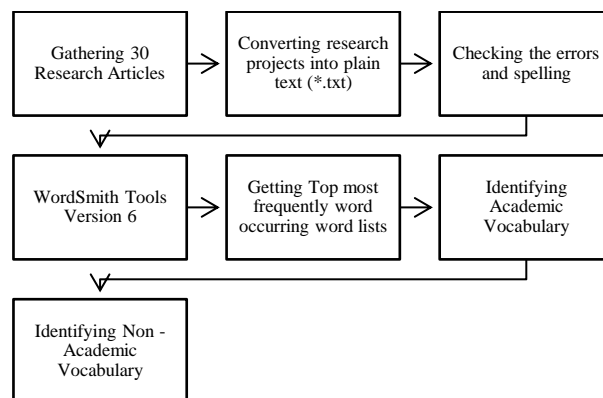


Fig. 1. Research Procedures

To reach the purposes of the study, the 30 research articles written by EFL graduate students were selected as the most important input for ERA Corpus. The first research purpose aims to probe the frequency and distribution of the academic words. To match this purpose, the first stage is to acquire a list of academic words that serve the criteria in the study which covers frequency and range. The criteria of the study is that the academic word should appear at least 50 times in the complete ERA Corpus. In addition, all text files were saved as a PDF and Word documents 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 30 most frequently occurring academic words.

The second research purpose 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 purpose, we adapted the similar criteria of the frequency and range and chose only non-academic words which appear at least 50 times in the whole RA Corpus.

III. RESULTS

The present study concentrated on the frequency, coverage, and distribution of academic words in the RA Corpus. Consequently, a corpus of 451,236 running words from EFL graduate students’ research articles has been applied. This section is structured to answer the two posted research purposes.

Research Purpose 1: To profile the most frequently used academic words in research articles written by EFL graduate students in mechanical engineering field.

TABLE II:
THE TOP 30 MOST FREQUENTLY ACADEMIC WORDS IN ERA CORPUS

Rank	Word	F	Rank	Word	F
1	PHRASE	243	16	OBTAIN	84
2	TRANSFER	235	17	LAYER	74
3	ENERGY	192	18	INPUT	72
4	CONDUCT	148	19	CONSTANT	68
5	PREDICT	133	20	MECHANISM	62
6	MEDIA	122	21	ACHIEVE	61
7	RATIO	120	22	OCCUR	60
8	TECHNOLOGY	109	23	REACTION	60
9	SECTION	108	24	DATA	59
10	REGION	103	25	SIMILAR	57
11	RESEARCH	96	26	FOUND	56
12	FUNCTION	94	27	AUTHOR	55
13	METHOD	90	28	DESIGN	54
14	DIMENSION	89	29	INVESTIGATE	48
15	RANGE	86	30	FACTOR	46

Table 2 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 2,844 occurrences that accounted for 0.63% of text coverage, of the whole corpus which was 100%. However, most of them occurred often to be included in the top 30 most frequently occurring words of the whole ERA Corpus. In the list, the top five high frequency words were “phrase” 243 times, “transfer” 235 times, “energy” 192 times, “conduct” 148 times, and “predict” 133 times. After comparing with the headwords of Academic Word List (Coxhead, 2000), the findings presented that of 212 out of 570 headwords of the academic word list, appeared frequent in this corpus and the coverage accounted for 1.01% of the token in the corpus.

Research Purpose 2: To explore the most frequently used non-academic words in research articles written by EFL graduate students in mechanical engineering field.

TABLE III:
THE TOP 30 MOST FREQUENTLY NON-ACADEMIC WORDS
IN ERA CORPUS

Rank	Word	F	Rank	Word	F
1	THE	4,913	16	HEAT	417
2	OF	3,312	17	GAS	403
3	AND	2,395	18	FROM	375
4	BE	2,375	19	AIR	358
5	IN	1,160	20	WITH	352
6	A	1,138	21	PRESENT	332
7	TO	859	22	FLAME	312
8	POROUS	816	23	FUEL	259
9	COMBUSTION	592	24	AT	232
10	BURNER	553	25	THIS	224
11	FOR	542	26	TRANSFER	223
12	TEMPERATURE	507	27	RADIATION	221
13	BY	471	28	EXPERIMENTAL	217
14	AS	465	29	COMBUSTOR	202
15	ON	458	30	MODEL	197

From the data shown in Table 3, it can be seen that the top 30 most frequently occurring non-academic words accounted for 5.5 % of the text coverage and there were a total of 24,880 occurrences. From the list, the top five high frequency words were “the” 4,913 times, “of” 3,312 times, “and” 2,395 times, “be” 2,375 times, and “in” 1,160 times.

The non-academic words should be known as a considerable element of the acquired knowledge of language learners. In addition, these words appearing in academic context were employed as a part of academic argument but on the other hand they were not involved as academic words as suggested by Coxhead (2000).

TABLE VI:
THE TOP 10 MOST FREQUENTLY ACADEMIC WORDS AND NON-ACADEMIC
WORDS IN ERA CORPUS

Words from the AWL		F	Words from the Non-AWL		F
1	PHRASE	243	THE		4,913
2	TRANSFER	235	OF		3,312
3	ENERGY	192	AND		2,395
4	CONDUCT	148	BE		2,375
5	PREDICT	133	IN		1,160
6	MEDIA	122	A		1,138
7	RATIO	120	TO		859
8	TECHNOLOGY	109	POROUS		816
9	SECTION	108	COMBUSTION		592
10	REGION	103	BURNER		553
Total		1,513			18,113

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 ERA Corpus. It can be observed that the words from the non-academic words with academic use in the ERA Corpus had a higher frequency than the first 10 words from the academic words in this study. In addition, the results showed that the ERA Corpus contained diverse function and some content words.

IV. CONCLUSION

In the present study, we examined a 451,236-word corpus of research articles in a specific discipline, i.e. mechanical engineering. The aim was to identify frequently used academic words in engineering research articles and develop a word list for EFL graduate students. We focuses on academic English and goal to lessen the problems of Thai EFL graduate students when writing academic research articles.

In total, we identified 212 word families used with a reasonable frequently in our corpus. We called these words Engineering Research Articles Corpus (ERA Corpus). A comparison of our words with the AWL displayed that numerous AWL items were not used many times in the subject area we investigated. Some AWL words appeared with high frequency; however, their frequency rank was different from the academic words in Coxhead’s Academic Wordlist. This could be implied that academic words are not used in the same way across disciplines.

Furthermore, the findings revealed that the non-academic words are highly frequent in our corpus. It could be assumed that EFL learners have good knowledge of non-academic words since they could use their knowledge to specific topic of

linguistics; however, they are often forgetful of grammatical rules. As Hyland and Tse (2007) suggested that academic vocabulary commonly comes after a general vocabulary. For the most frequent occurring non-academic word list, teachers can apply the frequently occurring word lists as a valuable source to instruct words to their students who are not familiar with their studies because of the frequently word list is a worthy supplement to various disciplines (Coxhead, 2000). In addition, the findings of the present study would be not only new findings that are beneficial to academic English learners to achieve their higher education, but also for teachers or course designers of academic English when they create their teaching materials for 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.

V. ACKNOWLEDGEMENT

The authors wish to express thank RMUTI graduate students who have given a permission to enable to collect their research articles as the needed data for the study. We are very thankful to the Faculty of Sciences and Liberal Arts, Rajamangala University of Technology Isan for their financial support for this study. We also would like to extend our thanks to the editors and reviewers for their insights and suggestions.

VI. REFERENCES

- [1] Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- [2] 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
- [3] Crystal, D. (2003). *English as a global language*. 2nd ed Cambridge: Cambridge University
- [4] Flowerdew, J. (1999). Writing for scholarly publication in English: the case of Hong Kong. *Journal of Second Language Writing*. Vol. 8 (2), pp. 123-145.
- [5] Hinkel, E. (2001). *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. New York, NY: Routledge.
- [6] Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- [7] Hyland, K. (2007). Genre pedagogy: language, literacy, and L2 writing instruction. *Journal of Second Language Writing*. Vol. 16, pp. 148-164.
- [8] Hyland, K. & Tse, P. (2007). Is there an "Academic vocabulary." *TESOL Q*. Vol. 41(2) pp. 235-253.
- [9] Kaur, J. & Hegelheimer, V. (2005, October). ESLstudents' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*. Vol. 18(4), pp. 287-310.
- [10] Kroll, B. (Ed.). (2003). *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press.
- [11] 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.
- [12] Scott, M. (2012). *WordSmith Tools version 6*. Stroud: Lexical Analysis Software.
- [13] Swales, J. (1999). Informal elements in English academic writing: threats or opportunities for advanced non-native speakers? in Candlin, C. and Hyland, K. (Eds.) *Writing: texts, processes and practices*. London, Longman