6 ICLEHI 2017-064 Rinda Warawudhi

Thai EFL Learners' Writing Styles via Corpus Driven Approach

Rinda Warawudhi*, Wipa Praditwiengkham
Department of Western Languages, Burapha University,
Chon Buri, Thailand
*Corresponding Author: lect.rinda@gmail.com

ABSTRACT

Keyword analysis can be used to reveal learners' writing styles and also to observe the behaviors of overuse, underuse and misuse of EFL learners. This study therefore investigates EFL learners' writing style through the writing corpus comparing with British Academic Written English (BAWE) corpus using AntConc program. Learner corpus was gathered from 19 writing project assignments of English-major undergraduate students from Burapha University, Thailand. It was found that there were four overused patterns: the use of words for subjects in the study, 2) the use of "got" instead of "obtain", 3) the use of buy, buyer, buyers, buying and bought, and 4) "who" relative clauses. Also three underused patterns of learners are found as follows: 1) the underuse of hedges and attitude markers, 2) the underuse of contrastive linking words, and 3) the underuse of "its". It was found from that learners's overused and underused patterns (e.g. the use of plural/singular forms, the use of article - a/an/the) that they have tried to avoid using some patterns by overusing that patterns that they were more confident to produce in their writing tasks. In addition, the spoken styles like informal words (e.g. got) were found frequently in their formal written assignment. Consequently, the findings could be implemented to curriculum design in selection of and the emphasis on certain written structure and vocabulary. It could benefit both teachers and learners in order to compare and enhance the awareness in the writing tasks with the reference corpus.

Keywords: learner analysis, EFL learning, writing styles, keyness, corpus

Introduction

Since computer learner corpus (CLC) research was introduced in the late 1980s, lot of studies on CLC have been generated and developed. Contractive Interlanguage Analysis (CIA), one of the major approaches in CLC research, has been conducted to make quantitative and qualitative comparisons between the two corpora: native (NS) and non-native (NNS) data or different varieties of non-native (NNS) data. The comparisons between NS and NNS can highlight a range of features of non-nativeness in learner writing and speech whilst the NNS/NNS comparisons can differentiate features among non-native learners (Granger, 2002). Granger (2004) has reviewed the CIA studies and found that all of these studies have investigated several linguistics features that were overused, underused and misused by learners. A wide range of linguistics features in NS/NNS comparisons included cohesive advices (Ersanli, 2015), connectives (Mo, 2015), features of writer/reader visibility (Paquot, Hasselgård, and Oksefjell Ebeling, 2013), interactional metadiscourse (Hong and Cao, 2014), linking adverbials (Leedham and Cai, 2013), phrasal verbs (Hao-Jan Chen, et al, 2015), and so on. There were also some studies focusing on cross-genre analysis in the NNS/NNS comparisons (Qin and Uccelli, 2016; Huang, 2015) as well as the study applying data-driven approach to investigate learners' writing styles (Lee and Chen, 2009).

However, most of the previous research studies have commonly examined advanced EFL learners (Hong and Cao, 2014) with targeted linguistics features. Since there was no prior study on students' writing performance using corpus in Burapha University, Thailand before, the study focusing on specific linguistics features might not be able to explain students' overall writing performance. This study therefore attempts to investigate writing styles of Thai undergraduate students by using data-driven approach having corpus of native learners as a reference corpus. Also, in order to investigate learners' writing styles and to assist lecturers to develop curriculum and class methodology, data-driven approach with native learner corpus might be an appropriate alternative option.

Research Question

There is one research questions in this study which is:

"How do Thai EFL learners' writing style differ from native speaker learners?"

In order to find answer for this research question, the analysis of positive keyness and negative keyness were examined.

Literature Review

With ability to analyze a large databank of texts, corpus analysis allows scholars and educators to discover patterns of authentic language. It can be a useful source in writing instruction (Ersanli, 2015) and also raise awareness of the use of linguistics features and help avoid making mistakes. Consequently, learners could write more authentic (Ersanli, 2015; Granger, 2002). A considerable amount of literature has been published on corpus analysis. Granger, 2002 states that there are two dimensions of comparisons including NS/NNS comparisons and NNS/NNS comparisons. The first type focuses on a range of features that non native speakers differ from the native one while the second type highlights on the differentiation between speakers from different mother tongue backgrounds. In addition, on the review, the comparison between three or more groups were also found with the attempts to explain the phenomenon of the outstanding linguistic features. Hence, the comparison studies were conducted in several dimensions as follows:

NS/NNS Comparisons

Two studies using NS/NNS comparisons were reviewed in this study. Mo's study (2015) examines the use of causal connectives of Chinese EFL learners with the Louvain Corpus of Native English Essays (LOCNESS) corpus, which is a corpus of argumentative essays written by British and American university students. He found that Chinese learners tended to use connectives with higher density but lesser variety. Initial positions of sentences were popular among Chinese learners. Another study by Ersanli (2015) measures the use of cohesive devices of Turkish university students comparing with British Academic Written English (BAWE) corpus. The findings were similar to Mo (2015) that the learners use cohesive devices with higher density but lesser variety as well as the preference to use devices at initial positions of sentences.

NS /NNS/NNS Comparisons or NS/ NS/ NNS Comparisons

Research studies on three-corpus-or-more comparisons were conducted in large volume as well. In these studies, learner corpus (NNS) was considered targeted corpus whist native speaker (NS) corpus was used as a reference corpus (e.g. BAWE) and another group of corpus can be NS or NNS corpus depending on the objectives of each study.

Hao-Jan Chen, et al, 2015 compares three corpora: Taiwanese master's theses, native speaker master's theses and native speaker journal articles for common phrasal verbs. Eight

phrasal verbs found in both three corpora were analyzed for overuse and underuse. L1 influence and mis-collocation were explained for these phenomena.

Paquot, Hasselgård, and Oksefjell Ebeling (2013) compare writer/reader visibility of French and Norwegian learners corpora (International Corpus of Learner English -ICLE, the Varieties of English for Specific Purpose Database -VESPA)with native speaker corpora (LOCNESS and BAWE). Corpora were analyzed in several dimensions in order to observe the use of learners' writer/reader visibility. Similar to Mo (2015) and Ersanli (2015), EFL learners produced overuse of writer/reader visibility features and less academic-like writing styles.

Leedham and Cai (2013) evaluate the use of linking adverbials among three corpora. They compared Chinese learner corpus with BAWE to find negative and positive key linking adverbials. Corpus of reading lessons and textbooks in mainland China were analyzed for the frequency and variety. The researchers concluded that ELT teaching materials could influence the overuse of linking adverbials in Chinese learners.

In addition, data-driven approach is also one of the analysis method for NS/NNS/NNS corpus studies. Lee and Chen (2009), for instance, applied keyword analysis for overuse and underuse words or phrases by learners (NNS corpus) with two reference corpora (NS learners and NNS journals) in order to finding Chinese students' writing patterns.

NNS/NNS Comparisons or NNS/NNS/NNS Comparisons

The studies among NNS corpus were also reviewed. Qin and Uccelli (2016), for example, have investigated Chinese EFL learners' writing performance whether they used similar or different language and functions in two genres: argumentative essays and narratives. Huang's study (2015) has also applied NNS/NNS comparisons by examining frequency and accuracy of lexical bundles used by two groups of EFL learners: junior year and senior year students. Three learner corpora were also studied in Hong and Cao, 2014 for discussing interactional metadiscourse features including hedges, boosters, attitude markers, self-mentions, and engagement markers. Similarity and difference in the use of these features could assist pedagogical implications.

There were some limitations in the reviewed studies above. First, English proficiency of learners varied. In Qin and Uccelli (2016), their target group was secondary-school learners of English whilst graduate students in Ersanli's study (2015) and (Hao-Jan Chen, et al, 2015) were examined their writing performance. Even other variables were controlled and the findings can discover writing behaviors of targeted learners, it might be implemented only in the specific contexts (each study's setting, for example). Second, the majority of the studies in this review section had already had objectives to evaluate specific linguistics features. This might not describe the whole writing patterns of learners, otherwise, interesting patterns found in the study may be overlooked with these limited objectives. Furthermore, there has been no study on students' writing performance in Burapha University using corpus analysis before. This study therefore investigates the writing patterns of EFL learners using data-driven approach.

Methodology

This section can be divided into three subsections: background information of the learners, corpora, and data analysis as follows:

Background information of the learners

Apart from 31 credits of general courses, undergraduate students majoring in English in Burapha University are required to study 84 credits of core courses and major courses. One of five compulsory writing courses were arranged for students every semester starting from

Year 2 Semester 1. The difficulty and the length of the writing pieces were increased constantly course by course in order to strengthen students' writing proficiency. For the last course, Project Writing in English, students are required to produce an English writing work (called "writing project") in accordance with the learner's interest under the guidance of lecturer. Students' writing project could be considered as an output from what the five writing courses have gradually shaped their performance.

Corpora

There were two corpora in this study including:

BUU corpus. 19 writing projects from 45 fourth-year undergraduate students who had agreed for this study to investigate their writing performance. BUU corpus or non-native learner corpus consisted of 60,805 words. Greenbaum, 1991 (cited in Ersanli, 2015) states that a corpus of 20,000-30,000 words is sufficient to analyze professional texts; BUU corpus size could therefore be relevant to investigate students' writing performance. BUU corpus was excluded tables, figure references, appendix as well as Literature review section because students were relatively new to conduct a research study and they could not paraphrase or crystallize the concepts of the related studies. To avoid plagiarism, quoting might be their alternative technique. In addition, from personal discussion with two out of five lecturers of this course, because of this reason, they suggested to delete literature review section out of the corpus since students tended to use quotation for this part.

BAWE corpus. The BAWE corpus contains 2,761 pieces of proficient assessed student writing, ranging in length from about 500 words to about 5,000 words. Holdings are fairly evenly distributed across four broad disciplinary areas (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences) and across four levels of study (undergraduate and taught masters level). This corpus was developed by the universities of Warwick, Reading and Oxford Brookes. Corpus files are available from the Oxford Text Archive (http://ota.ahds.ac.uk/).

Data Analysis

Corpus analysis program. AntConc (Anthony, 2005 & 2016) is a corpus analysis toolkit designed by Anthony, 2005 for specific use in the classroom that includes concordancers, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot. Since it is free and friendly-user, it is utilized to find positive and negative keyness words as well as other functions. Version 3.4.4w was utilized in this study.

Statistical testing. To answer research questions, keyness words were analyzed using Log-Likelihood ratio. The keyness words in this study included both positive words and negative words. Top positive words could be explained the learners' writing behavior that they overused words that learners from a reference corpus used lesser whist top negative words would be observed over the two corpora in order to examine the words that Thai learners used lesser than writers from a reference corpus with *Log-Likelihood ratio*. Paquot and Bestgen (2009) have compared the three statistical tests for keywords extraction including the Log-Likelihood ratio, the t-test, and the Wilcoxon-Mann-Whitney test through the British National Corpus with a 100 million word collection. Weighing advantages and disadvantage of three test, Log-Likelihood ratio was supported for keyword analysis.
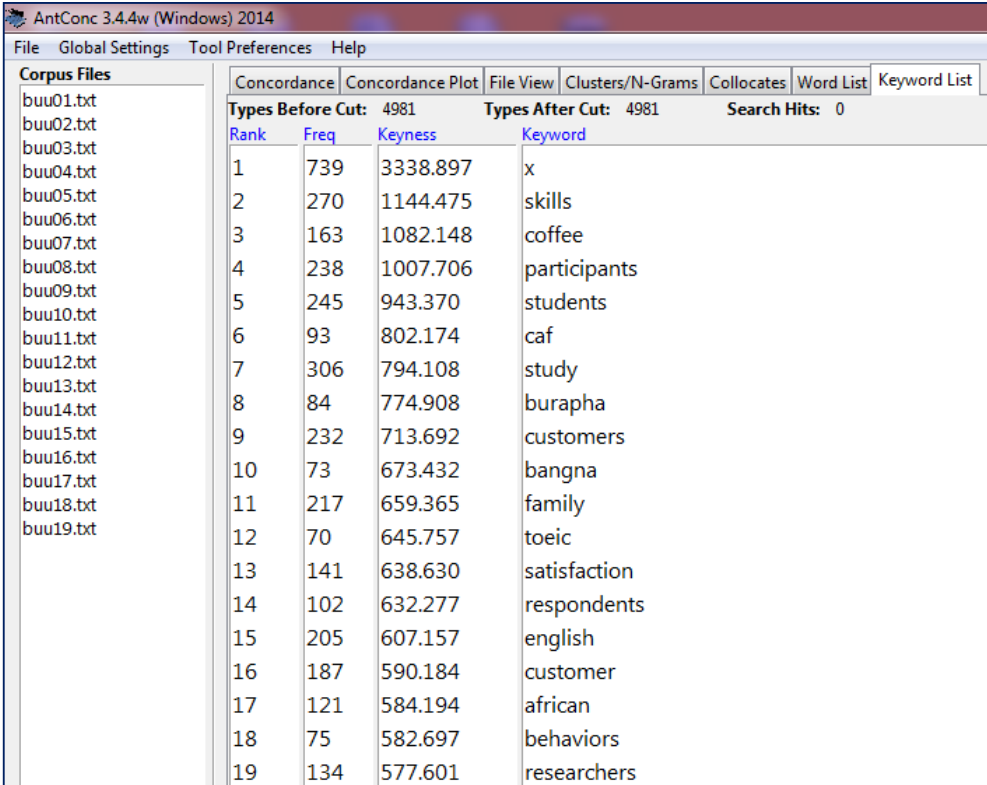
Results

The BUU corpus were collected from students' writing projects. Their topics were various depending on their interests, for instance, Analysis of Gender Identity: Giovanni's room, An analysis of students' lexical skill: A case of Burapha University students, A Case

Study of Racism through The Boy in the Striped Pyjamas, Satisfaction of English Major Students in Burapha University toward Summer Work & Travel in USA program, and so on. Content words related to students' topics therefore were excluded in this analysis. The results will be divided into two parts: positive keywords and negative keywords as follows:

Analysis of Positive Keywords

It was found that 4,981 words were considered positive keywords, which means that Thai learners used them more frequently than learners in BAWE corpus. The findings can be seen in Figure 1.

According to the findings from 500 top positive keywords, the discussion can be divided into four topics: 1) the use of words for subjects in the study, 2) the use of "got" instead of "obtain", 3) the use of buy, buyer, buyers, buying and bought, and 4) "who" relative clauses.

Figure 1. Results of positive keywords in BUU Corpus

The use of words for subjects. As can be seen in Figure 1, the words "participants", "students", "customers", "respondents", and "customer", in rank 4, 5, 9, 14, and 16, respectively are the top nouns represent "subjects" and "researchers" in rank 19 representing the writers. Below are the examples of "participants" from BUU corpus.

"...In addition, with reference to the answer of the participants in an interview, there was unexpected answer. Some of the participants felt out of control themselves when they had to talk with the foreigners. Even the participants studied in English major, sometime they felt lack confident and excited to communicate with the foreigners..." (buu05.txt)

It can obviously be noticed that the words "participants" was repeatedly used several times in the one paragraph or it was used contiguously sentence by sentence. There was a lack of use of pronoun in this case.

The use of "got" instead of "obtain". The use of the word "got" in most of situations as shown in Figure 2 may not be considered academic writing styles. In Hit #7, the word "became" might be more academic whilst in Hit#18, the word "obtain" should be replaced for more academic writing styles. In this aspect, it can assume that writers in BUU corpus seem to use spoken language in their academic writing projects.
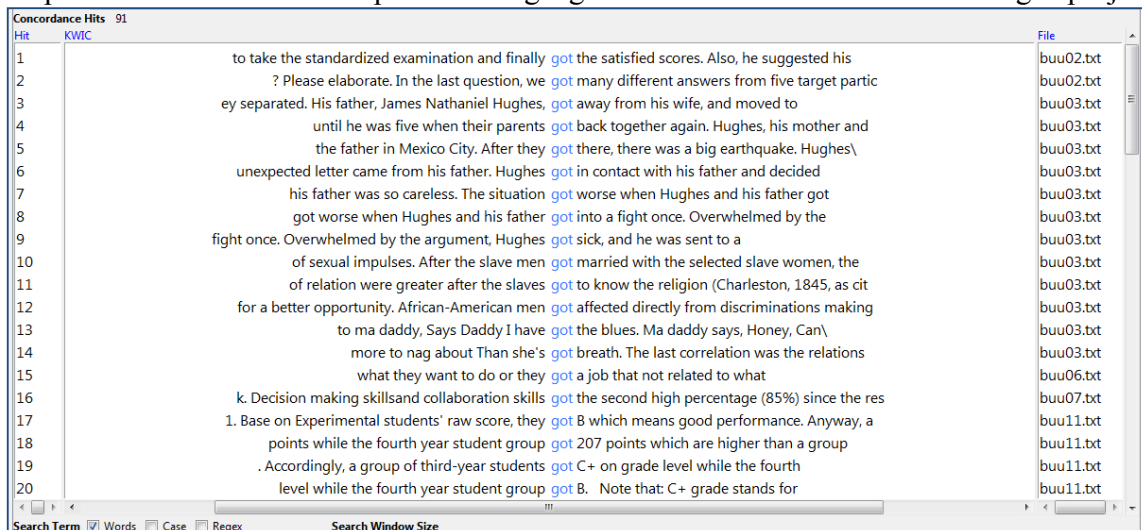

Concordance Hits	91		
Hit	KWIC		File
1	to take the standardized examination and finally got the satisfied scores. Also, he suggested his		buu02.txt
2	? Please elaborate. In the last question, we got many different answers from five target partic		buu02.txt
3	ey separated. His father, James Nathaniel Hughes, got away from his wife, and moved to		buu03.txt
4	until he was five when their parents got back together again. Hughes, his mother and		buu03.txt
5	the father in Mexico City. After they got there, there was a big earthquake. Hughes\		buu03.txt
6	unexpected letter came from his father. Hughes got in contact with his father and decided		buu03.txt
7	his father was so careless. The situation got worse when Hughes and his father got		buu03.txt
8	got worse when Hughes and his father got into a fight once. Overwhelmed by the		buu03.txt
9	fight once. Overwhelmed by the argument, Hughes got sick, and he was sent to a		buu03.txt
10	of sexual impulses. After the slave men got married with the selected slave women, the		buu03.txt
11	of relation were greater after the slaves got to know the religion (Charleston, 1845, as cit		buu03.txt
12	for a better opportunity. African-American men got affected directly from discriminations making		buu03.txt
13	to ma daddy, Says Daddy I have got the blues. Ma daddy says, Honey, Can\		buu03.txt
14	more to nag about Than she's got breath. The last correlation was the relations		buu03.txt
15	what they want to do or they got a job that not related to what		buu06.txt
16	k. Decision making skillsand collaboration skills got the second high percentage (85%) since the res		buu07.txt
17	1. Base on Experimental students' raw score, they got B which means good performance. Anyway, a		buu11.txt
18	points while the fourth year student group got 207 points which are higher than a group		buu11.txt
19	. Accordingly, a group of third-year students got C+ on grade level while the fourth		buu11.txt
20	level while the fourth year student group got B. Note that: C+ grade stands for		buu11.txt

Figure 2. Sentences with "got" in BUU Corpus

The use of buy, buyer, buyers, buying and bought. For the word group of "buy" and "purchase", it was found that the use of "buy" group were considerably high comparing with "purchase" group as displayed in Table 1.

Table 1
Lemma Word Forms of "buy" and "purchase" in BUU Corpus

Rank	Frequency	Keyness	Keywords
57	63	226.425	buy
99	36	148.731	buyers
101	41	145.740	buying
122	27	117.420	buyer
137	31	108.616	bought
329	29	61.605	purchase
408	16	28.976	purchasing
456	6	26.240	purchasers

The word "purchase" was considered business terms for "buy" which was used more frequently as a spoken language. However, results in Table 1 reveal that learners from BUU corpus significantly use word group of "buy" higher than word group of "purchase".

"who" relative clauses. It was found from the positive keyword that "who" relative clauses were utilized to modify nouns. Other relative clauses with other words e.g. where, when, or which were not included in the top 500 positive keywords. When comparing with BAWE corpus, the number "who" adjective clauses in BUU corpus were relatively high in BUU corpus.

Figure 3. Sentences with "who" in BUU Corpus

Analysis of Negative Keywords

1,704 words were reported as negative keywords, which means that Thai learners used them lesser than expert learners in BAWE corpus. The findings can be seen in Figure 4.

Figure 4. Results of Negative Keywords in BUU Corpus

The findings from 500 top negative keywords can be discussed in three topics including: 1) the underuse of hedges (may, would) and attitude markers (should, must), 2) the underuse of contrastive linking words (however, although, and whilst), and 3) the underuse of "its".

The underuse of hedges (may, would) and attitude markers (should, must). Regarding to Hong and Cao (2014), hedges and attitude markers are subtypes of interactional discourse to mark writer's epistemic and affective stances. Hedges can be defined as words to mitigate or tone down propositional strength. Examples of hedges included could, may, maybe. Also, attitude markers, for example should and must, mean

words to express affective stance towards propositions. Table 2 below shows the frequency and keyness values of "may", "would", "should", and "must" from BUU corpus.

Table 2
Keyness Values of "may", "would", "should", and "must" in BUU Corpus

Rank	Frequency	Keyness	Keywords
17	43	54.137	may
19	68	52.947	would
29	29	40.123	should
55	17	27.432	must

Here are the sentences from BUU corpus.

They <u>may</u> be a single parent by choice or by life circumstances. (buu03.txt)

They <u>would</u> escape the place where does not belong to them. (buu09.txt)

The further research <u>should</u> categorize questions before collecting questionnaires. (buu05.txt)

Their answers were surprised because they <u>must</u> have had the awareness of those words, but they <u>must</u> not. (buu11.txt)

According to Table 2, it was obviously seen that BUU learners use these groups of words infrequently, however, they were able to use it accurately in order to mitigate and express their affective stance.

The underuse of contrastive linking words (however, although, and whilst). The findings revealed the negative keyness on the use of contrastive linking words (however, although, and whilst) as shown in Table 3.

Table 3
Keyness Values of "however", "although", and "whilst" in BUU Corpus

Rank	Frequency	Keyness	Keywords
18	51	53.410	however
39	13	32.779	although
29	1	27.545	whilst

Thai learner uses these contractive linking words infrequently. On the other hand, considering additional linking words, it was found that the word "in addition" were found 54 hits and the positive keyness value of the word "addition" was 56.179. Surprisingly, there were no contractive linking words in 500 top positive keywords. It could be assumed that these learners were familiar to use additional linking words rather than contrastive linking words.

Positioning of the linking words was also interesting issue, Thai learners tended to use the linking words in "initial parts of sentence" as shown in Figure 5.

Figure 5. Sentences with "however" in BUU Corpus

The underuse of "its". It was found from concordance plot of the words "its" in two corpora in Table 4 below that in BUU corpus, the word "its" was found in 9 out of 19 files or 47.36% whereas the word "its" in BAWE corpus was found in 2,146 out of 2,716 files or 79.01%. This would signify that Thai EFL learners wrote this word much lesser than expert learners in BAWE corpus. Nearly half of students did not use "its" in their assignments.

When comparing with the word "their", another possessive adjectives, its frequency and the files having the word "their" were different with the appearance of the word "its" as can be seen in Table 4 below.

Table 4
Appearance of "its", and "their" in Two Corpora

	BUU corpus			BAWE corpus	
	Hits	Keyness	Found plots	Hits	Found plots
its	17	-138.061	9 (47.36%)	11,933	2,146 (79.01%)
their	389	+154.904	19 (100%)	19,073	2,375 (87.44%)

As shown it Table 4, in BUU corpus, the word "its" was found in quite low numbers that was 17 hits in 9 out of 19 files (47.36%) while the word "their" was found 389 hits in every file or 100%. In BAWE corpus, there were not much different of these two words in number of hits and plots, in other word, the word "its" was found in 79.01% and the word "their" was hit in 87.44% from all files. This means that the native learners in BAWE corpus use these two words in quite similar proportion and this is different phenomenon with Thai learners. In addition, from the analysis using *Log-Likelihood ratio*, the word "their" was found as a positive keyword in Thai corpus with the keyness value of 154.904 whereas the word "its" was in the negative keyword category with the keyness value of 138.061. They rarely used the word "its" since there were small number of singular nouns in their writing.

Discussion and Conclusion

This study would like to explore Thai EFL learners' writing styles in Burapha University Thailand by comparing with BAWE corpus, a reference corpus. AntConc, a concordance program developed by Antony, 2005 was used in the study for keyword analysis in order to observe distinguish words in both positive and negative keywords of BUU learners. Four features that Thai EFL learners overused were found including: 1) the use of words for subjects in the study, 2) the use of "got" instead of "obtain", 3) the use of buy,

buyer, buyers, buying and bought, and 4) "who" relative clauses. In addition, three issues from negative keywords that Thai writers infrequently used were 1) the underuse of hedges (may, would) and attitude markers (should, must), 2) the underuse of contrastive linking words (however, although, and whilst), and 3) the underuse of "its".

From the overuse of the words for subjects which were mostly plural forms and underuse patterns of "its", it can explain that Thai learners tended to use "plural" in a higher degree than "singular". In fact, when considering the use of "a" and "an", it was found that they fall in negative keyness category with the value of 230.911 and 71.495, respectively. The findings then obviously reflect students' writing styles that they tended to play safe in their writing. The overuse of plural nouns and pronouns can be interpreted that they tried to avoid using singular words which might be more complex.

Word choices can also be a problem in their writing, it was obviously shown in the positive keywords for the word "got" instead of "obtain" or the word "buy" instead of "purchase". This means that a spoken language was applied in their academic writing work. Apart from "who", students also lacked of the use of other relative pronouns e.g. when, where, which, and so on.

In addition, EFL learner underused hedges (may, would) and attitude markers to express their own stance. This may lead from the genre of their writing. The BUU corpus was gathered from study academic writing project. They might use hedges and attitude markers in a greater volume if they are asked to write argumentative essays, for example. However, it was found that Thai learners use contrastive linking words such as "however" or "although" lesser than writers in a reference corpus and also lesser than additional linking words in their own corpus e.g. "moreover" and "in addition". This could reflect that BUU learners could not write in a variety types of sentences. They were also familiar to use "linking words" in the initial part of sentence. To sum up, the underuse and overuse of some linguistic features found in this study show that they willing to keep themselves away from mistakes.

Considering course syllabi of five writing courses students have learned as well as a discussion with another two instructors in previous writing courses, it was found that students have been informed during their previous writing courses to use variety of word choices and sentence patterns. However, these were not written in grading criteria for these actions. To be in the safe zone, students then preferred to use words or patterns that they were very certain for accuracy. It comes to the conclusion among writing teachers that if the variety of lexical and grammar patterns were added in the grading criteria, students then were automatically encouraged to significantly apply various words and forms in their writing assignments.

The implications of this study could be categorized into two areas: curriculum design, and classroom methodology. The findings from this study will be benefits for instructors of writing courses in the selection and structuring of teaching contents. In the field of grammar teaching, the variety use of patterns or variety styles to express writer stances should be focused, for example. In the field of vocabulary, synonyms or similar words for use in various registers, e.g. formal/ informal, spoken/academic, etc. should be emphasized in a class. For classroom methodology, the findings from this study should provide the picture of the writing performance of native speakers and non-native speakers. Acknowledging the difference would enhance students' awareness in their writing performance. In addition, the use of corpus in the classroom should be advantages for students to observe the way native learners compose sentences and to compare themselves to them as well.

Limitations

Only the writing assignments from 19 groups or a-third of all fourth year English major undergraduate students from Burapha University were investigated, the findings therefore could not represent all Thai EFL learners. Also, the topics of writing in BUU corpus

varied depending on students' interest, so it might not match to a reference corpus even the content words were excluded in the discussion parts.

Recommendations

This study focused only on the works of students in their last writing class from the total five compulsory subjects in English-major bachelor degree, hence, it might reflect what students have learned from the first class to the last one. The use of extra exercise reflecting the writing behaviors for these learners to judge their perceptions toward their writing style is suggested to validate the findings from the corpus analysis. Moreover, it might be interested to have longitudinal studies to observe their development of writing performance along these five writing courses.

Acknowledgement

We would like to acknowledge Dr. Sirirat Na Ranong from Thammasat University for granting useful advises and comments to do this research. Our thanks also go to the EFL students studying at Burapha University for participating in this study.

References

Anthony, L. (2016). AntConc Homepage [Software, manual and training videos]. Retrieved from http://www.laurenceanthony.net/software/antconc/

Anthony, L. (2005). AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom. *Proceedings of IEEE International Professional Communication*, 729-737. DOI: 10.1109/IPCC. 2005.1494244

Ersanli, C. (2015). Insights from a Learner Corpus as Opposed to a Native Corpus about Cohesive Devices in an Academic Writing Context. *Journal of Educational Research*, 3(12), 1049-1053.

Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, S. Petch-Tyson, & J. Hulstijn (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (Vol. 6, pp. 3–33). Amsterdam & Philadelphia: John Benjamins. Retrieved from http://books.google.be/books?id=SgEtnIOdC5kC&printsec=frontcover&dq=Computer+learner+corpora,+second+language+acquisition+and+foreign+language+teaching&hl=fr&sa=X&ei=0kbdUauTOcKr0QX92oHQDQ&ved=0CDEQ6AEwAA#v=onepage&q&f=false;http://benjamins.com/#catalog/books/lllt.6/main

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: a multidimensional perspective* (pp. 123–145). Amsterdam & Atlanta: Rodopi.

Hao-Jan Chen, H.; Ting-Yu Yang, C.; Feng-Fan Wei, I.; and Jiang, A. (2015). A Corpus Study on Phrasal Verb Use in the Academic Writing of Published Authors, Native English-Speaking Students, and Taiwanese EFL Learners. *English Teaching & Learning*, 2015 Special Issue (39), 63-91.

Hong, H. and Cao, F. (2014). Interactional metadiscourse in young EFL learner writing A corpus-based study. *International Journal of Corpus Linguistics*, 19(2), 201-224.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System,* 53(2015), 13-23. DOI: 10.1016/j.system.2015.06.011

Lee, D., and Chen, S. X. (2009). Making a bigger deal of the smaller words: function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18(3), 149–165.

Leedham, M. & Cai, G. (2013). Besides . . . on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students' use of linking adverbials. *Journal of Second Language Writing*, 22(2013), 374-389.

Mo, J. (2015). A contrastive study of the use of causal connectives by Chinese EFL learners and English native speakers in writing. *Theory and Practice in Language Studies*, 22(5). 2426-2432.

Paquot, M and Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In *Corpora: Pragmatics and Discourse* (pp247-269). DOI: 10.1163/9789042029101_014

Paquot, M., Hasselgård, H., and Oksefjell Ebeling, S. (2013). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead. *Proceedings of the First Learner Corpus Research Conference (LCR 2011)* (pp. 377–387). Louvain-la-Neuve: Presses Universitaires de Louvain.

Qin, W. and Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, 33(2016), 3-17. DOI: 10.1016/j.jslw.2016.06.001

University of Oxford Text Archive. (2016). *British Academic Written English Corpus* [Data file]. Retrieved from http://ota.ahds.ac.uk/headers/2539.xml