CORPUS-BASED ERROR ANALYSIS OF THAI STUDENTS' LABORATORY SCIENTIFIC ABSTRACT WRITING IN ENGLISH

การวิเคราะห์ข้อผิดพลาดในการเขียนบทคัดย่อรายงานการทดลองทางวิทยาศาสตร์เป็นภาษาอังกฤษของนักเรียนไทยโดยใช้คลังข้อมูลภาษา

Anon Ua-umakul[1] and Jirapa Vittayapirak[2]
อานนท์ เอื้ออุมากุล[1] และจิราภา วิทยาภิรักษ์[2]
[1]Master Degree Student in Applied Linguistics (EST) Program
[2]Associate Professor, the Department of Applied Arts
Faculty of Industrial Education King Mongkut's Institute of Technology Ladkrabang
anon_uaumakul@hotmail.com and kvjirapa@gmail.com

Abstract

The aims of this study were to analyze and compare the main causes of common errors of Thai students using the corpus-based approach. The samples were two groups of upper secondary school students (Grade 11) at Watsuthiwararam School, Bangkok who enrolled Physics Laboratory Technique course in 2014. They were divided into two groups by simple random sampling: (1) the experimental group and (2) the control group. The instruments were (1) a corpus of 540 Laboratory Scientific Abstracts (LSA) written by the participants, (2) an error classification recording scheme, (3) lesson plans for teaching scientific abstract writing in English, and (4) a guided interview for investigating the causes of errors. After the corpus was compiled, the errors and causes of errors were categorized. The frequency and percentage of errors were counted and calculated using Markin 4.2.2 and AntConc 3.2.4 software. Chi–square test (χ^2-test) was used to compare the errors committed by the two groups of students. The results were as follows: (1) The top three morphological errors were the omission of article "the" (9.62%), the omission of article "a" or "an" (8.35%), and the subject-verb agreement (4.65%), respectively. (2) The top three syntactic errors were the misuse of active voice (5.79%), the misplacement of verb (4.51%), and the incorrect form of parallel structure (3.87%). (3) The top three mechanical errors were the use of commas (9.85%), periods (5.04%), and capitalization of words (2.75%). (4) For lexical errors, misspelling (7.82%) was found more than homonym words (0.15%). (5) The top three semantic errors included specific term usage (0.98%), wrong collocation usage (0.97%), and the more general term usage (0.71%). The results indicated that there was statistical difference of errors between the two groups at .01 level of significance. The main causes of error were the L1 interference (33.33%), and the misconception in language rules (31.93%).

Keywords: Learner Corpus; Laboratory Scientific Abstract (LSA); Error analysis; Written error; Causes of error

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์และเปรียบเทียบสาเหตุความผิดพลาดในการเขียนโดยใช้คลังข้อมูลภาษากลุ่มตัวอย่างเป็นนักเรียนไทยระดับชั้นมัธยมศึกษาตอนปลาย (มัธยมศึกษาปีที่ 5) จำนวนสองกลุ่ม ซึ่งเป็นนักเรียนโรงเรียนวัดสุทธิวราราม กรุงเทพมหานคร และลงทะเบียนเรียนในวิชาเทคนิคปฏิบัติการฟิสิกส์ ปีการศึกษา 2557 กลุ่มตัวอย่างคัดเลือกโดยใช้การสุ่มอย่างง่ายและแบ่งเป็นกลุ่มทดลองและกลุ่มควบคุม เครื่องมือวิจัยประกอบด้วย (1) บทคัดย่อรายงานการทดลองทางวิทยาศาสตร์ จำนวน 540 ฉบับ (2) แบบบันทึกและจัดประเภทข้อผิดพลาดในการเขียน (3) แผนการจัดการสอน เรื่อง การเขียนบทคัดย่อรายงานการทดลองทางวิทยาศาสตร์เป็นภาษาอังกฤษ และ (4) แบบสัมภาษณ์สำหรับตรวจสอบสาเหตุความผิดพลาดในการเขียน

หลังจากรวบรวมงานเขียนบทคัดย่อของนักเรียนเสร็จแล้วนำมาจัดสร้างคลังข้อมูลภาษา วิเคราะห์และจัดประเภทข้อผิดพลาดและสาเหตุความผิดพลาดในการเขียน สถิติที่ใช้ ได้แก่ การคำนวณค่าความถี่และร้อยละของข้อผิดพลาดโดยใช้โปรแกรมสำเร็จรูป Markin4.2.2 และ AntConc3.2.4 การเปรียบเทียบความแตกต่างของความถี่ข้อผิดพลาดในการเขียนของนักเรียนสองกลุ่มใช้การทดสอบไคสแควร์ ผลการศึกษาพบว่า (1) ข้อผิดพลาดด้านหน่วยคำที่เกิดขึ้นมากที่สุด 3 อันดับแรก คือ การละคำนำหน้าคำนาม “the” (9.62%) การละคำนำหน้าคำนาม “a”หรือ “an” (8.35%) และความสอดคล้องกันของประธานและกิริยา (4.65%) (2) ข้อผิดพลาดด้านวากยสัมพันธ์ที่เกิดขึ้นมากที่สุด 3 อันดับแรก คือ การใช้กรรตุวาจก (5.79%) การวางตำแหน่งของคำกิริยาในประโยค (4.51%) และการใช้โครงสร้างแบบคู่ขนาน (3.87%). (3) ข้อผิดพลาดเชิงกลไกที่เกิดขึ้นมากที่สุด 3 อันดับแรก คือ การใช้เครื่องหมายจุลภาค (9.85%) การใช้เครื่องหมายมหัพภาค (5.04%) และการใช้ตัวอักษรพิมพ์ใหญ่ (2.75%) (4) ข้อผิดพลาดด้านศัพท์ พบว่ามีการสะกดคำผิดพลาด (8.72%) ซึ่งมากกว่าการใช้คำที่ออกเสียงใกล้เคียงกัน (0.15%) (5) ข้อผิดพลาดด้านความหมายคำที่เกิดขึ้นมากที่สุด 3 อันดับแรก คือ การใช้คำที่มีความหมายเฉพาะเจาะจงเกินไป (0.98%) การใช้คำปรากฏร่วม (0.97%) และการใช้คำที่มีความหมายกว้างเกินไป (0.71%) จากการวิเคราะห์พบว่า ความถี่ของข้อผิดพลาดในการเขียนของนักเรียนสองกลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ **.01** สาเหตุหลักของความผิดพลาดในการเขียน คือ การนำกฎจากภาษาแม่มาใช้ (33.33%) และมโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับกฎเกณฑ์ทางภาษา (31.93%)

คำสำคัญ: คลังข้อมูลภาษาโดยผู้เรียน บทคัดย่อรายงานการทดลองทางวิทยาศาสตร์ การวิเคราะห์ข้อผิดพลาด ข้อผิดพลาดในการเขียน สาเหตุความผิดพลาด

1. INTRODUCTION

The importance of English for Science is receiving increasing attention from educators and policy makers in the field of teaching English as a Foreign Language. As we can see from Thai Basic Education Core Curriculum [1], it appears on Standard F3: Language and Relationship with other learning areas as follows:

‘StandardF3.1: Usage of foreign languages to link knowledge with other learning areas, as foundation for further development and to seek knowledge and widen one’s world view’

According to this standard, there are many ways to use English for linking knowledge with science; for example, the writing and/or presenting a scientific report is one interesting learning activity. The learners receive a good opportunity to use English for communication in this subject area and build up many related vocabularies as well. An abstract is one crucial part in a scientific report because it helps readers understand the main parts of report, i.e. introduction, method, result and discussion, and conclusion [2]. In order to properly write an abstract of a scientific report, the learners need to understand the paragraph organization, the usage of English grammar, the usage of technical terminologies and scientific vocabularies. However, it practically appears that Thai learners have problems with writing. For example, Sattayatham studied the problem about scientific paragraph writing in English made by medical students at Mahidol University [3]. This study was concluded that the two main problems in scientific writing were no transitional words and lack of organization. Nopjirapong studied errors of twenty English essays written by second-year English major students at Srinakharinwirot University [4]. The results revealed that the main error were mother tongue interference and the omission of article “the”. Ratchawicha investigated errors in guided composition of Mattayomsuksa three students [5]. The findings found that the top three common grammatical errors were punctuation (17%), noun phrase (14%), and word choice (13%), respectively. These errors resulted from interference of the mother tongue and lack of grammar rules.

Corder and Ferris suggested that in order to help students to use the English language correctly, one approach that teachers can use for identifying errors and analyze the causes of them was error analysis [6], [7]. The learners’ errors provide evidence of the language system and error analysis can help in effective learning and teaching of English. It is a

useful approach theoretically and practically because the students learn from their mistakes. Furthermore, teachers can understand students' learning process and select materials appropriate to their learning styles. Therefore, this study investigated common errors in writing scientific abstract. The samples were asked to do two laboratory works and write two laboratory scientific abstracts in English. The errors and causes of error were identified, analyzed and implied to teach English for science.

2. RESEARCH OBJECTIVES

The objectives of this study were as follows:

1) To analyze the common errors in laboratory scientific abstracts in English made by Thai upper secondary school students

2) To compare the errors in written scientific English abstracts between the students who were and were not taught scientific abstract writing in English

3) To investigate the main causes of errors in written scientific English abstracts

3. RESEARCH HYPOTHESIS

The hypothesis of this study was defined that there was the statistical difference of errors between the students who were taught how to write scientific abstract and the ones who were not taught at .01 level of significance.

4. RESEARCH METHODOLOGY

4.1 POPULATION AND SAMPLES

The population of the study was the upper secondary school students at Watsuthiwararam School, Bangkok. The samples were selected using the purposive sampling method. They were 270 Thai science and mathematics program students in Mathayomsuksa 5 (grade 11). They were divided into two groups by the simple random sampling method. They were (1) the experimental group (the students who were taught how to write a scientific abstract in English), and (2) the control group (the students who were not taught how to write a scientific abstract in English).

4.2 RESEARCH INSTRUMENTS

The instruments of this study included (1) a learner corpus of 540 Laboratory Scientific Abstracts (LSA) written by the participants, (2) an error classification recording scheme, (3) the lesson plans for teaching scientific abstract writing in English, and (4) a guided interview for investigating the causes of errors.

1. A learner corpus of 540 Laboratory Scientific Abstract (LSA corpus) was compiled. It originated from the laboratory work in class and all students were asked to write abstracts that consisted of four moves: introduction, method, result and discussion, and conclusion [2].

2. An error classification recording scheme was constructed following the related concepts of error analysis. These errors were classified into five categories according to [8], [9]:

2.1 Morphological errors (The error occurred from using the incorrect word structure)

2.2 Syntactic errors (The error occurred from using any incorrect items larger than word, i.e. phrase, clause or sentence)

2.3 Lexical errors (The error occurred from the use of words that look and/or sound similar (misselection), the literal translation from Thai words to English word (misformation), and misspelling (distortion))

2.4 Semantic errors (The error occurred from the confusion of sense relation and use of collocation)

2.5 Mechanical errors (The error occurred from punctuation, number, abbreviation, capitalization, and acronym in sentences)

Each of them was divided into various sub-categories following the tagsets or the error tagging system. It was created by analyzing the errors of the 140 students' scientific abstracts consisting of 17,458 words. According to Díaz-Negrillo and García-Cumbreras, they recommended that the sufficient words suitable for creating tagsets are 15,000 words

[10]. Therefore, the number of words to be studied was sufficient for creating tagsets. From literature review, there were examples of types of error. The samples and major types were found and used to create tagsets as shown in Table 1 below:

Table 1 The Error Tagset System

Errors	Types	Example	Tagset
Morphological	Subject-verb agreement	The data from the table is taken to plot a graph.	MO
Syntactic	Fragment	The refractive index values is.	SY
Lexical	Spelling	It makes small water wafe on water's surface in a beaker.	LE
Semantic	Wrong collocation	The horizontal distance will directly vary as spring contraction.	SE
Mechanical	Capitalization	the error may occur from the friction.	ME

We can see from the table that the tagsets were divided into 5 sets; a morphological error tagset (MO), a syntactic error tagset (SY), a lexical error tagset (LE), a semantic error tagset (SE), and a mechanical error tagset (ME). Each tagset was further analyzed in details of a number of tags as follows: 1) 33 tags of the morphological error, 2) 21 tags of syntactic error, 3) 2 tags of lexical error, 4) 4 tags of semantic error, and 5) 12 tags of mechanical error.

3. Five lesson plans for teaching scientific abstract writing in English were designed to be correlated with the four moves of abstract: introduction, method, result and discussion, and conclusion. The first lesson was about how to write an introduction and the usage of infinitive. The fifth lesson was designed on how to link each move of the abstract with linking words. In short, the course content was divided into 5 topics as shown in Table 2 below:

Table 2 The scientific abstract writing in English course content

Lesson Plan	Topic	Time (min)	Language point
1	Introduction	50	Infinitive with to
2	Method	50	Passive voice
3	Result & Discussion	50	Can & may
4	Conclusion	50	Simple past tense
5	Linking Parts	50	Linking words

In each topic, the teaching approaches were PPP: Presentation, Practice, and Production [11].

4. The guided interview was constructed based on Norrish, Brown, and Richards in order to find out the causes of written errors [12], [13], and [14].

4.3 PROCESS OF THE STUDY

The process of the study was as follows:

1. The concepts of written error analysis, scientific abstract writing, lesson plan design, and corpus-based error analysis were studied from various resources [2], [8], [9], [11], [12], [13] and [14].

2. The five lesson plans for teaching the scientific abstract writing were constructed.

3. The 140 scientific abstracts (approximately 17,000 words) were collected to make a tagset.

4. The samples were divided into two groups based on simple random sampling technique: (1) the experimental group and (2) the control group.

5. The 540 pieces of abstracts were compiled and analyzed based on tagsets.

6. Frequency and percentage of written errors were counted and calculated using Markin 4.2.2 and AntConc 3.2.4 software.

7. The 88 analyzed abstracts [44 abstracts from the control group and 44 abstracts from the experimental group] were randomed to check the validity of error analysis by three experts. All selected abstracts were defined the size based on Yamane Table [15].

8. The χ^2-test was used to compare the statistically significant differences of errors made by the two groups of students.

9. The guided interview of students was made to analyze the causes of written error using content analysis.

10. The results and pedagogic implication were concluded and further studies were recommended.

The process of study showed in Figure 1 below:

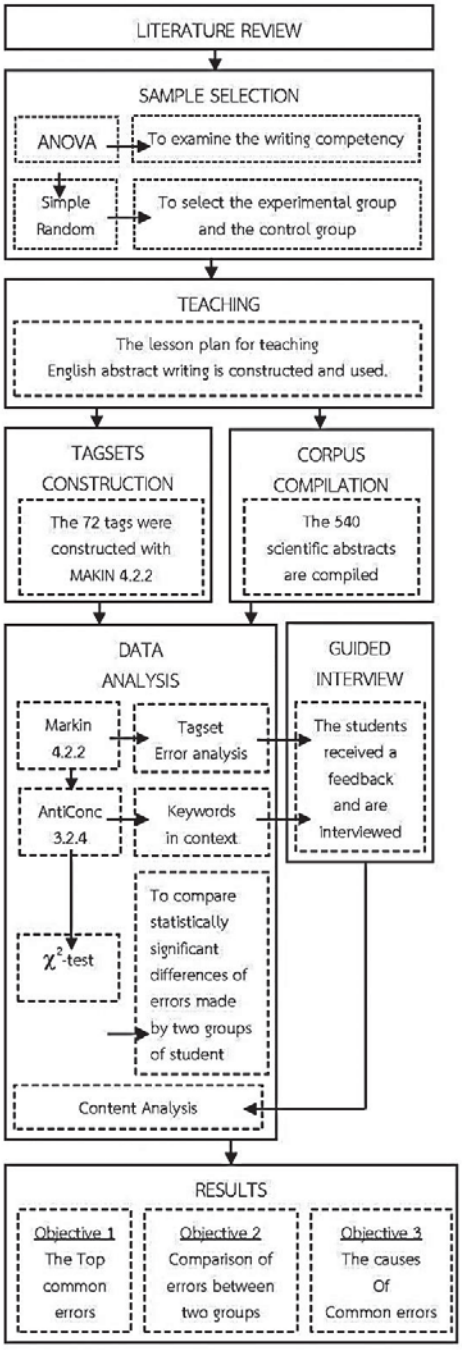

Figure1: The process of the study

5. RESULTS

1. According to the first objective of this research, it aimed to analyze the common errors. It was found that the total number of common errors committed by the participants was 9,143. They were morphological errors (38.99%), syntactic errors (25.82%), mechanical errors (24.30%), lexical errors (7.97%), and semantic errors (2.92%). Table 3 shows the top three of each type of common errors.

Table 3 The analysis of the top three common errors in LSA corpus

Rank	Type of errors	Sub-type of errors	f	%
1	Morphological error	1. The omission of article *the*	962	10.52
		2. The omission of article *a, an*	763	8.35
		3. Subject-verb agreement	426	4.65
		Total	**3,565**	**38.99**
2	Syntactic error	1. Misuse of the active voice	529	5.79
		2. Misplacement of verb	412	4.51
		3. Incorrect form of parallel structure	354	3.87
		Total	**2,361**	**25.82**
3	Mechanical error	1. Commas	901	9.85
		2. Periods	461	5.04
		3. Capitalization	251	2.75
		Total	**2,222**	**24.30**
4	Lexical error*	1. Misspelling	715	7.82
		2. Use of words that look similar sound	14	0.15
		Total	**729**	**7.97**
5	Sematic error	1. The use of specific term	90	0.98
		2. The use of collocation	89	0.97
		3. The use of a general term	65	0.71
		Total	**266**	**2.92**

* There are only two tags in the lexical error.

It can be seen that the top three morphological errors were the omission of article “the” (9.62%), the omission of article “a” or “an” (8.35%), and subject-verb agreement (4.65%), respectively. The top three syntactic errors were misuse of the active

voice (5.79%), misplacement of verb (4.51%), and incorrect parallel structure (3.87%). The top three mechanical errors were the use of commas (9.85%), periods (5.04%), and capitalization of words (2.75%). For lexical error, misspelling (7.82%) was found in writing work more than homonym words (0.15%). Lastly, the top three semantic errors were the use of specific term (0.98%), wrong collocation words (0.97%), and use of general term (0.71%).

2. When the errors made by experimental and control group were compared, it was found that the highest frequency of errors was the morphological errors, followed by syntactic errors, mechanical errors, lexical errors and semantic errors, respectively. The finding was presented in Figure 2 below:

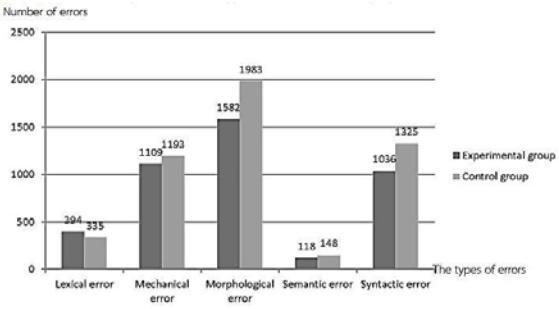

Figure2. The comparison of the five types of errors between two groups of student

It is obvious that the experimental group made less errors than the control group in terms of the mechanical error, the morphological error, the semantic error, and the syntactic error. However, the control group made less in lexical errors.

In order to find out the statistical significant difference in error made between the two groups of students, chi-square test (χ^2-test) was used. The results indicated that difference of errors between the two groups was found at .01 level of statistical significant as shown in Table 4 below:

Table 4 The significant differences in errors between 2 groups

Student Groups	Frequency of error					χ^2 test	P
	LE	ME	MO	SE	SY		
Ex.	394	1,109	1,582	118	1,036		
Con.	335	1,193	1,983	148	1,325	40.5	<.01
Total	729	2,222	3,565	266	2,361		

Ex = Experimental group Con. = Control group
LE = Lexical error ME = Mechanical error
MO = Morphological error SE = Semantic error
SY = Syntactic error

3. The students' interview indicated the main causes of errors in written scientific English abstract. The data were analyzed and represented in Figure 3 below:

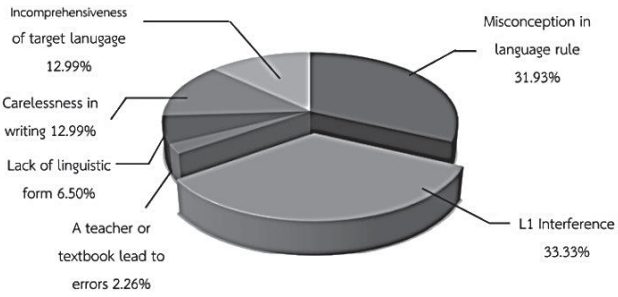

Figure3. The causes of errors occurred in the LSA corpus

According to the students' interview, it was found that the main causes of errors were the L1 interference (33.33%), the misconception in language rules (31.93%), the carelessness in writing (12.99%), the incomprehensiveness of target language (12.99%), the lack of linguistic form (6.50%), and the teacher or textbook leads to errors (2.26%), respectively.

6. DISCUSSION AND CONCLUSION

This research revealed that the morphological errors were most frequently found, especially, (1) the omission of articles "the", and (2) the omission of articles "a" or "an". Most students ignored the use of articles in sentences. This result was corresponding to a previous study conducted by Nopjirapong. The omission of "the" was occurred the most frequently in Thai students' writing of English. The L1 interference may be the main cause of this error because in Thai language, we do not use articles at all in contrast to English [16].

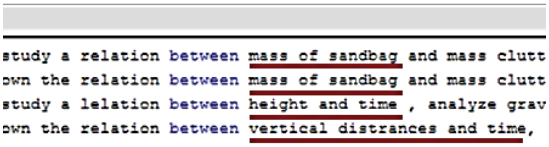

Figure4. The example of morphological errors (use of article)

As for the syntactic error, the top three errors found in the abstract writing were (1) the misuse of the active voice, (2) misplacement of verb, and (3) incorrect parallel structure, respectively. The result conformed to Bumroongthai that error in grammar and sentence structures were most frequently found in Thai student writing [17]. In this study, there are two main reasons supporting why the students made syntactic errors: (1) L1 interference and (2) using the contraction forms in Thai spoken language. From the data, it was apparent that the students did not know how to use passive voice in abstract writing, whereas passive voice was used often by the native speakers, especially the method parts. Besides, most of the students frequently misplace the verb and misuse of parallel structure.

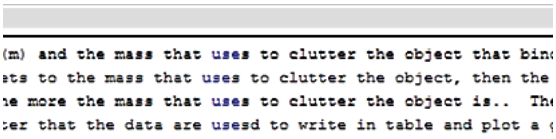

Figure5. The example of syntactic errors (passive voice)

As for the mechanical error, the misuse of the comma, period, and capitalization were found frequently. It corresponded with [5] which found the most written error of Thai student was the use of punctuation (17%). The results indicated that most students omitted comma, period, and capitalization in English writing because of the L1 interference. Thai language does not use these punctuations in general whereas English language does.

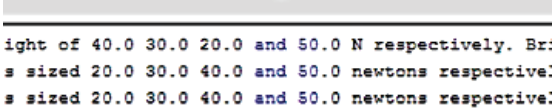

Figure6. The example of mechanical errors (use of comma)

According to the interview, the students made a lot of spelling mistake (7.82%) due to their own carelessness and lack of concentration in writing. For example, they wrote “a sand bag” instead of “a sandbag”, or wrote “proprotional” or “propotional” instead of “proportional”.

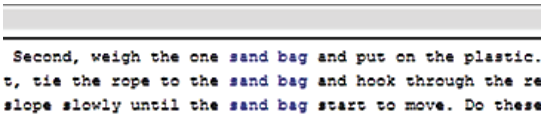

Figure7. The example of lexical errors (spelling)

In terms of the semantic error, it was found that the students frequently used the too specific terms and wrong words. For example, they used the word, “varied” instead of “proportional” in sentences. Normally, the word “varied or variation” was used to mean the changes that occur in populations of living organism over time whereas proportional was used in mathematic or physics for presenting the ratio of quantities. This indicated that some Thai students either could not select and use the appropriate words to write sentences in English or did not know the exact meaning of specific terms and how to use them correctly. Therefore, teachers need to pre-teach the meaning of scientific terminology.

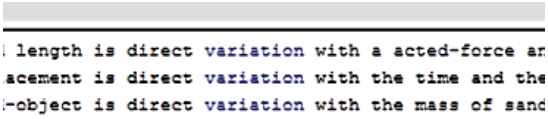

Figure8. The example of semantic errors (use of wrong word)

Another interesting finding in this study was the comparison of error between the two groups of students. It was found that the experimental group made less error than the control group. This finding indicated that the teaching process could help students decrease the errors in writing and the students need to get some inputs such as vocabulary and grammar focus before they perform the writing tasks, because the teaching process made the students aware of the important language point of abstract writing such as the use of modal verbs, and the linking words. However, it is interesting to find out that the experimental group made more spelling mistakes than the control group. This might be because in the process of teaching, the students in experimental group were taught how to write the abstract and new vocabulary and/or terminology. That means this group got some chances to use the new vocabulary

in their written work. Therefore, they have more chances to make misspelling than a control group.

In conclusion, this study has shown the common errors in the laboratory scientific abstract writing of Thai students by using corpus-based analysis. It was found that five main important errors were morphological, syntactic, lexical, semantic, and mechanical errors, respectively. The pedagogical implication of this study is that teachers should use top common errors to design a syllabus to help students avoid common errors in writing such as the omission of articles "the", the use of commas, misspelling, and misuse of the active and passive voice. The results may be used as a guideline for English teachers to improve their writing teaching method for Thai students or the EFL students.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Assoc. Prof. Dr. Jirapa Vitayapirak, an expert of corpus linguistics at Faculty of Industrial Education, King Mongkut's Institute of Technology Ladkrabang, and to thank all thesis committees, all thesis committees, Assist. Prof. Dr. Sirirat Petsangsri, Assoc. Prof. Sukhuman Nilrat, Assoc. Prof. Chalemsri Prijapanij and Assist. Prof. Dr. Pattama Patpong for excellent guidance.

REFERENCES

[1] Ministry of Education. (2008). **The Basic Education Core Curriculum B.E. 2551 (A.D.2008).** Bangkok: The Agricultural Co-operative Federation of Thailand., Limited.

[2] Bazerman, C. (1995). **The Informed Writer**. 5th ed. Boston: Houghton Mifflin.

[3] Sattayatham, A. and Ratanapinyowong, P. (2008). Analysis of Errors in Paragraph Writing in English by First Year Medical Students from The Four Medical Schools, Mahidol University. **Silapakorn University International Journal**, 8, 17-38.

[4] Nopjirapong, R. (2010). **An Analysis of Article Error in Thai university student's Composition**. Master's project,M.A. Bangkok: Graduate school, Srinakharinwirot University.

[5] Ratchawicha, S. (2012). An Error Analysis of Mattayom 3 Student's Guided Compositions a Saint Anthony School, Chachoengsao Province. **Research articles Graduate School of Language and Communication**, 1, 12-21.

[6] Corder, S.P. (1981). **Error Analysis and Interlanguage**. Oxford: Oxford University.

[7] Ferris, D. (2002). **Treatment of Error in Second Language Student Writing**. Ann Arbor: University of Michigan Press.

[8] James, C. (1998). **Errors in Language Learning and Use**. England: Addison Wesley Longman.

[9] Bryan, P. (2010). **Thesis Manual**. Virginia: University of Virginia.

[10] Díaz-Negrillo, A. and García-Cumbreras, A. M. (2007). A tagging tool for error analysis on learner corpora. **ICAME Journal**, 31, 197-203.

[11] Harmer, J. (2007). **The Practice of English Language Teaching**. 4th ed. London: Pearson Longman.

[12] Norrish, J. (1983). **Language Learning and Their Errors**. London: Macmillan Publisher

[13] Brown, H.D. (1980). **Principles of Language Learning and Teaching.** New Jersey: Prentice-Hall Inc.

[14] Richards, J.C. (1971). A Noncontrastive Approach to Error Analysis. **English Language Teaching Journal**, 25, 204-219

[15] Yamane, Taro. 1967. **Statistics, An Introductory Analysis**. 2nd Ed. New York: Harper and Row

[16] Nopjirapong, R. (2010). **An Analysis of Article Errors in Thai university student's Composition**. Master's project, M.A. Bangkok: Graduate school, Srinakharinwirot University.

[17] Bumroongthai, G. (2010). **An Error Analysis in English Paragraph Writing by Students of the Faculty of Liberal Arts Rajamangala University of Technology PhraNakhon**. Bangkok: Rajamangala University of Technology PhraNakhon.