A Corpus-Based Study of English Collocations Found in the Abstracts of Research Articles Written by Thai EFL Writers

Suthathip Thongvitit and Narathip Thumawongsa

Abstract—**This study aims to examine the types and count frequency of grammatical and lexical English collocations used in the abstracts of research articles in the field of liberal arts and humanities, which were written by Thai EFL writers from 2010 to 2015, and to identify misused English collocations produced by the Thai EFL writers. Software, named Antconc and TagAnt, were used to analyze data. The Oxford Collocations Dictionary for Students of English, the second edition published in 2009, was used when identifying misused collocations. It was found that noun + preposition and adjective + noun collocations were grammatical and lexical collocations that were used the most often. The most often misused grammatical and lexical collocations found in the study were noun + preposition and verb + noun collocations.**

Index Terms—**Abstracts, collocations, corpus-based study, Thai EFL writers.**

I. INTRODUCTION

Many Thai graduate students engage in research article writing; and, they are also required to write abstracts in English, even for research articles written in Thai, in order to promote internationalization and meet journal quality evaluation [1]. In a research article, an abstract may help draw attention from both local and international readers as well as help the readers save time because an abstract condenses all important points in a study [2]. Nonetheless, a poorly written abstract may also fail a good work [3]. Therefore, it is possible that an abstract could play an essential role in attaining approval for publishing or attending a research conference.

Rather than content, the use of language is also important when writing an effective English abstract, which may not be an easy task for Thai EFL writers because English is not their native language. Regarding the use of English language, English collocations could be an interesting topic for EFL writers, who need to write in a native-like style.

As in [4], collocations can be explained as pairs or groups of words that co-occur but may not be adjacent. The knowledge of collocations or an ability to know which words usually co-occur is an intuitive ability of native speakers and

is difficult for EFL learners due to its arbitrariness [5], [6].

Thai students learn English as a foreign language and may not use English collocations as well as native-speakers do.

Therefore, this study aims to examine the types and count frequency of English collocations, and to identify misused English collocations produced by Thai EFL writers. Research results may raise awareness of misused English collocations, which may contribute to improvement in English abstract writing.

The study seeks to answer following research questions:
1) What are the types and frequency of English collocations used by the Thai EFL writers in the abstracts of research articles in the field of liberal arts and humanities?
2) What are the misused English collocations found in the abstracts of research articles in the field of liberal arts and humanities written by Thai EFL writers?

II. LITERATURE REVIEW

A. Abstracts

Abstracts are brief descriptions, usually one paragraph, of a whole work. The length of abstracts varies from 50 to 300 or 500 words depending on the requirement of an institution or an instructor [7], [8]. Abstracts appear in academic papers such as journals, research articles, proceedings, and theses, etc. They present information such as the research design, the objective, findings and conclusions of studies [7]. There are many different types of abstracts, which are classified based on how the abstracts present information [9]. The two well-known types of abstracts are descriptive and informative abstracts [10]. Descriptive abstracts describe the aims, scope and methodology of the studies but not the conclusion and discussion. Informative abstracts report the important information from each chapter of the studies as well as findings and conclusions [7], [10].

B. Collocations

Many scholars have defined a collocation according to the pattern of its occurrence. Reference [11] viewed a collocation as "a group of words that occurs repeatedly" or "recurrent phrases". Reference [4] explained that those words may not be adjacent; the words may occur "within a short space of each other in a text". Reference [12] said that "a collocation is a predictable combination of words". Reference [13] mentioned that a collocation is the way in which words naturally co-occur rather than how they are put together. Reference [14] argued that the definitions of collocations are

overlapping as many of them share the same idea about the "co-occurrence of words". In conclusion, the definitions of collocations proposed by the scholars agree on the idea that a collocation is a group of combined words that frequently and naturally occur together.

Reference [11] considered the components of combination and then classified collocations into two types: grammatical and lexical collocation.

Grammatical collocation

As in [11], the grammatical collocation is a repeated combination that consists of a dominant word and a grammatical word. The dominant word can be a verb, noun or adjective, and the grammatical word is usually a preposition. Hence, the major types of grammatical collocations are 1) a verb + preposition collocation, 2) a noun + preposition collocation, and 3) an adjective + preposition collocation. The meaning of the grammatical collocation is inferred from the dominant word.

Lexical collocation

Reference [11] pointed out that what makes the lexical collocation different from the grammatical collocation is the equality of their elements. The lexical collocation is made up from two lexical elements which are equal; none of which is more dominant. In the opposite, the grammatical collocation often contains prepositions, which have no lexical meaning. The main types of lexical collocations are 1) an adjective + noun collocation, 2) a noun + verb collocation, and 3) a verb + noun collocation.

C. Corpus-Based Approach

A corpus, or corpora in plural form, is a collection where language texts are gathered in. It provides a great number of examples of language, which vary in purposes. The texts gathered in the corpus are either written language or the transcription of spoken language [4]. The types of corpus are divided into two types, which are sample corpus and monitor corpus. The sample corpus contains a limited amount of texts. After it is created, it will not be modified [4]. That is because the sample corpus is created to present a specific type of language at a particular period of time [15]. Reference [15] also called this type of corpus a balanced corpus. In contrast to the sample corpus, the monitor corpus is dynamic as its size continues to increase.

D. AntConc and TagAnt

AntConc and TagAnt are software developed by Dr. Laurence Anthony from Waseda University. The AntConc contains seven tools that function to manipulate texts such as sorting, listing, displaying data, and counting words. The seven tools are namely Concordance, Concordance Plot, File View, Cluster/N-Grams, Collocates, Word List, and Keyword List [16]. The TagAnt is a simple tool that can effectively detect the parts of speech of texts, which is the only one main function of the program. Users can either type texts or input files. Then, the software will proceed to tag the parts of speech found in texts after a command is made by users [17].

III. METHODOLOGY

A. Research Design

This study is a non-experimental research that employed corpus-based techniques to obtain and analyze data. The study used the following grammatical and lexical collocation patterns proposed by [11] as a framework due to their clearly observable structure.

Grammatical collocation

Verb + preposition

Noun + preposition

Adjective + preposition

Lexical collocation

Adjective + noun

Noun + verb

Verb + noun

B. Instrumentations

The instruments used in this study were AntConc version 3.4.3, TagAnt version 1.2.0, and the Oxford Collocations Dictionary for Students of English, the second edition published in 2009, all of which were the latest version when the study was conducted.

AntConc and TagAnt were selected because their features were suitable for the study. They work with plain text files; therefore, electronic data collected could be conveniently manipulated. AntConc was also used in previously studies as in [18], [19].

The Oxford Collocations Dictionary was used to confirm correct usage of English collocations when identifying misused English collocations. As in [20], using collocation dictionaries allows learners to be aware of collocations, determine usage, and use collocations effectively. The Oxford Collocations Dictionary was selected due to the reliability of its publisher as it is reputable and is used worldwide. It was also used in previous studies about collocations as in [21], [22].

C. Population

Research articles, in the field of liberal arts and humanities, written in Thai language by Thai writers from 2010 to 2015, were selected. Only works written in this recent period were selected in order to discover the authentic usage at the present time. Then, only English abstracts of the selected research articles were retrieved to build a corpus of English abstracts. As a result, the amount of works that met the selection criteria was somewhat small. Finally, the population of this study was forty English abstracts.

D. Data Collection

The research articles were randomly selected from academic online resources that provide the electronic collection of research articles so that the data could be retrieved in electronic format and be manipulated by the selected software.

E. Data Analysis

Firstly, elements that could make up collocations needed to be recognized by detecting the parts of speech. Therefore, the data, kept in plaintext format, were uploaded into TagAnt. After receiving a command, the software processed the data. Then, all the parts of speech found were tagged and shown on a screen. Finally, the results found were manually noted.

To answer the first research question, all prepositions tagged by TagAnt were used as search items when using

AntConc. AntConc was used to find nouns, verbs, and adjectives, that might co-occur with the prepositions found and make up grammatical collocations. It was also used to count frequency of the potential grammatical collocations.

All the three types of grammatical collocations share one same component that is a preposition, which is located on the right side of a noun, verb, and adjective. Therefore, the prepositions found by TagAnt could be used as search items when using AntConc to find nouns, verbs, and adjectives that occurred on their left span to find potential collocations.

However, components of the three types of lexical collocations are not located in the same fixed location. In addition, components of lexical collocations are two lexical words, including nouns, adjectives, and verbs, which are parts of speech normally required to compose a sentence. As a result, the data contained too many nouns, adjectives, and verbs. Therefore, components of lexical collocations, which were nouns, verbs, and adjectives, could not be effectively used as search items when using AntConc. As a result, potential lexical collocations were searched and counted manually instead of using the software.

To answer the second research question, all the potential grammatical and lexical collocations found were examined. In this process, the Oxford Collocations Dictionary was used to determine whether the potential collocations were common collocations suggested by the Dictionary.

IV. RESULTS

After an investigation, the researchers found that both grammatical and lexical collocations were used by the Thai EFL writers.

Grammatical collocations found were noun + preposition, verb + preposition, and adjective + preposition collocation. The grammatical collocations found were the combination of nouns, verbs, and adjectives with prepositions namely *at, about, as, among, between, by, for, from, in, into, of, on, to, toward/towards* and *with*.

The answer to the first research question was concluded in the following Table I.

TABLE I: TYPE AND FREQUENCY OF ENGLISH COLLOCATIONS

Type of grammatical collocation found	Number of uses	Most often used collocation	Number of uses
Noun + preposition	250	Usage of	13
Verb + preposition	126	Divide into	14
Adjective + preposition	34	Related to	14
Type of lexical collocation found	Number of uses	Most often used collocation	Number of uses
Adjective + noun	337	Age group	14
Verb + noun	161	Use word	12
Noun + verb	39	Results show	9

According to the Table I, among the three types of grammatical collocations, noun + preposition collocation was used the most often at 250 times, followed by verb + preposition collocation used 126 times, and adjective + preposition collocation used 34 times. The most often used noun + preposition collocation was *usage of,* which was used

13 times. The most often used verb + preposition collocation was *divide into*, which was used 14 times. The most often used adjective + preposition collocation was *related to*, which was used 14 times.

Next, lexical collocations found were adjective + noun, verb + noun, and noun + verb collocation. Among the three types of lexical collocations, adjective + noun collocation was used the most often at 337 times, followed by verb + noun collocation used 161 times, and noun + verb collocation used 39 times. The most often used adjective + noun collocation was *age group*, which was used 14 times. The most often used verb + noun collocation was *use word*, which was used 12 times. The most often used noun + verb collocation was *results show*, which was used 9 times.

Furthermore, the researchers also found misused grammatical and lexical collocations.

The answer to the second research question was concluded in the following Table II.

TABLE II: TYPE AND FREQUENCY OF MISUSED ENGLISH COLLOCATIONS

Type of misused grammatical collocation found	Number of misuses	Most often misused collocation	Number of misuses
Noun + preposition	43	Knowledge about Words of	4 4
Verb + preposition	35	Emphasize on	6
Adjective + preposition	1	Important of	1
Types of misused lexical collocation found	Number of misuses	Most often misused collocation	Number of misuses
Verb + noun	35	Use dialect	7
Adjective + noun	18	Basic exercises	4
Noun + verb	7	Results found	4

According to the Table II, all three types of grammatical collocations were misused. Among the three types, noun + preposition collocation was misused the most often at 43 times, followed by verb + preposition collocation misused 35 times, and adjective + preposition was misused only 1 time. The most often misused noun + preposition collocations were *knowledge about* and words *of*, both of which were misused 4 times. The Oxford Collocations Dictionary suggested to use *knowledge of* and *words in* instead. Next, the most often misused verb + preposition collocation was *emphasize on*, which was misused 6 times. The Dictionary did not suggest to collocate *emphasize* with a preposition. Lastly, the only misused adjective + preposition found was *important of*. The Dictionary suggested to use *important to.*

All three types of lexical collocations were also misused. Among the three types, verb + noun collocation was misused the most often at 35 times, followed by adjective + noun collocation misused 18 times, and noun + verb collocation misused 7 times. The most often misused verb + noun collocations was *use dialect,* which was misused 7 times. The Dictionary suggested to use, for example, *enhance knowledge* and *speak dialect* instead. Next, the most often misused adjective + noun collocation was *basic exercises*, which misused 4 times. The Dictionary suggested to use *simple exercise* when referring to a set of questions. Finally, the most often misused noun + verb collocation was *results found*, which was misused 4 times. The Dictionary suggested to use,

for example, *reveal* or *show* to collocate with *result*.

V. Conclusion

A. Discussion

First, according to the research results, it could be implied that the Thai EFL writers have problems with using English collocations and may need improvement, especially with noun + preposition and verb + noun collocation, which were the most often misused grammatical and lexical collocations in the study. Particularly, difficulties with using verb + noun collocation of EFL learners were also recognized in previous studies. For instance, reference [23] made a claim that two types of collocations were problematic for L2 learners; and, verb + noun collocation was one of them. Reference [22] reported that an ability to use verb + noun collocation of Thai EFL students was at a very low level. Reference [24] also found that Thai EFL writers could not produce verb + noun collocation themselves as well as recognizing it.

Second, according to some misused collocations found, problems with using English collocations of the Thai EFL writers might be caused by the interference of a first language. The Thai EFL writers might have applied a direct translation method when producing the collocations; they might have used knowledge from Thai language to support when writing in English. Since words in Thai and English may not be equivalent, using the direct translation method could result in selecting incorrect words and finally producing misused collocations. For example, *knowledge about* was used instead of *knowledge of.* With the direct translation method, *about* could be translated by the Thai EFL writers into เกี่ยวกับ (kiaokab) in Thai language. In Thai language, เกี่ยวกับ (kiaokab) is often used to refer to or introduce a related subject. Therefore, the Thai EFL writers could incorrectly use *knowledge* about when trying to introduce a subject of the knowledge, in English.

Third, some misused collocations could have been resulted from lack of efficient collocation knowledge. Especially, inefficient collocation knowledge could cause difficulties when producing lexical collocation, which is a combination of two lexical words. Several English words could have the same meaning, but each word may be used in different contexts. As a result, the Thai EFL writers made up a collocation by using words that native-speakers do not commonly use. In this study, the researchers found that *basic exercises* was used. *Basic exercises* superficially looks like an adjective + noun collocation. However, a common collocation suggested by the Dictionary was *simple exercise.*

Literally, *basic* and *simple* may have similar meaning; but, *simple* is commonly used to collocate with *exercise* rather than *basic* when referring to a set of questions.

A similar problem was found in EFL learners in Japan. Reference [6] gave an example of misused collocation as "There are many high buildings in Tokyo". This sample sentence is grammatical and understandable. However, "high buildings" may sound uncommon for native-speakers because the native-speakers commonly say *tall buildings*. In Japanese language, たかい (takai) is used when saying both *high* and *tall*. Therefore, some Japanese EFL learners may not be aware that in English not both *high* and *tall* can be used to collocate with *buildings.* All in all, it could be possible that producing proper lexical collocation primarily relies on collocation knowledge of EFL learners.

Finally, it is suggested that EFL teachers focus on teaching English collocations because even Thai EFL writers at an advanced level seemed to have difficulties in using English collocations. The data of this study were retrieved from graduate writing; yet, numbers of misused collocations were found. EFL teachers may review the misused collocations found in this study and determine whether their EFL students may have similar problems with collocations to help the students cope with problems they may have and reduce using a direct translation method. In addition, high attention may be paid to verb + noun collocation because the findings of this study agreed with previous studies, as in [22]-[24], that this type of collocation could be somewhat problematic.

B. Limitations

Some limitations in this study may be considered when conducting further studies in similar area. Firstly, the size of the data was quite small due to the selection criteria. Thus, the results may not be used to generalize all Thai EFL writers. Secondly, a source of data was limited to the field of liberal arts and humanities, which seemed to affect research results. For instance, *language usage* was adjective + noun collocation that was found repeatedly used 13 times in the study. This collocation is commonly used in language and linguistic genre. Therefore, research results may change when collecting data from a different field. Lastly, this study used collocation patterns proposed by [11] as a framework only. Research results could be different if using other collocation frameworks.

References

[1] Thai-Journal Citations Index Centre (n.d.). [Online]. Available: http://www.kmutt.ac.th/jif/public_html/quality_TCI57.html

[2] P. Kanjantra and M. Barraclough. (January-March 2008). English-Language Abstracts in Thai Dental Journals. *Journal of the Dental Association of Thailand*. [Online]. 58(1). pp. 56-65. Available: http://www.jdat.org/dentaljournal/en/journal/view/2008OR1006

[3] R. Wilkins, R. (May 1958). Editorial: How to write an abstract. *Circulation.* [Online]. 17(5). pp. 841. Available: http://circ.ahajournals.org/content/17/5/841.citation

[4] J. Sinclair, *Corpus, Concordance, Collocation*; Oxford: Oxford University Press, 1991, ch. 8, pp. 115, 170.

[5] F. E. Bazzaz and A. A. Samad. (September 2011). The use of verb noun collocations in writing stories among Iranian EFL learners. *CCSE. English Language Teaching.* [Online]. 4(3). pp. 158-163. Available: https://eric.ed.gov/?id=EJ1080737

[6] T. Nagono and K. Kitao. (March 2007). Using simple computational linguistic techniques for teaching collocations. *Journal of culture*

and information science. [Online]. *2*(1). pp. 1-15. Available: https://doors.doshisha.ac.jp/duar/repository/ir/13591/039000020001. pdf

[7] R. Labaree. (January 2016). Research guides: Organizing your social sciences research paper: 3. the abstract. *USCLibraries*. [Online]. Available: http://libguides.usc.edu/writingguide/abstract

[8] T. Koltay, *Abstracts and Abstracting: A Genre and Set of Skills for the Twenty-First Century*; Oxford: Chandos, 2010.

[9] G. Singh, *Information Sources, Services and Systems*; New Delhi: PHI Learning, 2013.

[10] S. Gould. (January 2011). Study guides: Writing. *Centre for academic success*. [Online]. Available: http://library.bcu.ac.uk/learner/writingguides/1.28.htm

[11] M. Benson, "Collocations and Idioms," in *Dictionaries, Lexicography and Language Learning*, R. Ilson, Ed. Oxford: Pergamon Press, 1985, pp. 61-68.

[12] J. Hill, "Revising priorities: from grammatical failure to collocational success," in *Teaching Collocation: Further Development in the Lexical Approach*, M. Lewis, Ed. Hove: Language Teaching Publication, 2000, ch. 3, pp. 51.

[13] M. Lewis, "Language in the lexical approach," in *Teaching Collocation: Further Development in the Lexical Approach*, M. Lewis, Ed. Hove: Language Teaching Publication, 2000, ch. 7, pp. 132.

[14] G. Woolard, "Collocation-encouraging learner independence," in *Teaching Collocation: Further Development in the Lexical Approach*, M. Lewis, Ed. Hove: Language Teaching Publication, 2000, ch. 2, pp. 29.

[15] T. Mcenery and A. Hardie. (2012). Corpus linguistics: Method, theory and practice. [Online]. Available: http://assets.cambridge.org/97805218/38511/excerpt/9780521838511 _excerpt.pdf

[16] L. Anthony. (2014). Antconc version 3.4.3, computer software. [Online]. Available: http://www.laurenceanthony.net/

[17] L. Anthony. (2015). TagAnt version 1.2.0, computer software. [Online]. Available: http://www.laurenceanthony.net/

[18] T. Khamphairoh, "Collocations of keywords found in insurance research articles: A corpus-based analysis," M.A. thesis, Dept. Foreign Languages, Kasetsart Univ., Bangkok, Thailand, 2012.

[19] T. Sae-lim, "A comparative corpus study of transitional words in research articles of Thai and native-like English writers," M.A. thesis, Dept. Language Studies. King Mongkut's Univ. of Technology Thonburi, Bankok, Thailand, 2011.

[20] H. Lien, "The effects of collocation instruction on the reading comprehension of Taiwanese college students," Ph.D. dissertation, Dept. English, Indiana Univ. of Pennsylvania, Indiana, PA, 2003.

[21] C. Yumanee and S. Phoocharoensil. (July 2013). Analysis of collocational errors of Thai EFL students. *Language Institute, Thammasat Univ. LEARN Journal.* [Online]. 6(1). pp. 88-98. Available:https://social.tci-thaijo.org/index.php/learn/article/view/100 0126

[22] N. Wangsirisombat, "A study of collocational errors in Thai EFL learners' writing," M.A. thesis, Language Institute, Thammasat Univ., Bangkok, Thailand, 2011.

[23] U. Nurmukhamedov, "An evaluation of collocation tools for second language writers," Ph.D. dissertation, Dept. English. Northern Arizona Univ., Flagstaff, AZ, 2015.

[24] R. Khittikote, "The ability to use collocations for business purposes by Thai EFL learners," M.A. thesis, Language Institute, Thammasat Univ., Bangkok, Thailand, 2011.

Suthathip Thongvitit was born in Kalasin, Thailand. She completed her bachelor's degree in English from Mahasarakham University, Maha Sarakham, Thailand in 2009. Currently, she is doing her master's degree in English at Faculty of Humanities, Srinakharinwirot University, Bangkok, Thailand.

Narathip Thumawongsa was born in Bangkok, Thailand. She completed her bachelor's in business English from Assumption University and master's and doctoral degree in language and communication from National Institute of Development Administration. She is currently a lecturer and secretary of the doctor of philosophy (English) program committee, Faculty of Humanities, Srinakharinwirot University, Bangkok, Thailand.