

Lexical Bundles in Literature Writing: A Corpus-Based Study of Native English Speakers and MCU Students

Ven. Suriya*

Veerakarn Kanokkamalade**

Preecha Khanetnog***

Faculty of Humanities,

Mahachulalongkornrajavidyalaya University, Thailand

E-mail: ven.suriya@mcu.ac.th

Received : June 14, 2022

Revised : August 30, 2022

Accepted : September 04, 2022

Abstract

This research on “Lexical Bundles in Literature Writing: A Corpus-Based Study of Native English Speakers and Mahachulalongkornrajavidyalaya University Students” was aimed to (1) explore the most frequently used four-word lexical bundles in native English speakers’ and Mahachulalongkornrajavidyalaya University students’ writing, (2) identify the structural types of lexical bundles that were employed in native English speakers and Mahachulalongkornrajavidyalaya University students, and (3) discover the variation of functional usage between native English speakers and Mahachulalongkornrajavidyalaya University students writing. Mixed method was used for this study for collecting data. This research paper compiled two corpora of writing from the English international program and the linguistics program to address the research objectives, both representing the writing of native English speakers and Mahachulalongkornrajavidyalaya University students in their respective fields. This study employed AntConc 3.5.9 software to discover four-word lexical bundles and analyses the bundles’ frequency, structure, and function.

* Ven.Suriya. Ph.D Candidate, Linguistic (English Programme), Faculty of Humanities Mahachulalongkornrajavidyalaya University

** Asst.Prof.Dr.Veerakarn Kanokkamalade Faculty of Humanities, Mahachulalongkornrajavidyalaya University. Thesis Advisor.

*** Assoc.Prof.Dr.Preecha Khanetnog Faculty of Humanities, Mahachulalongkornrajavidyalaya University. Thesis Advisor.

The study's findings indicated that native English speakers and Mahachulalongkornrajavidyalaya University students use bundles more frequently, exhibit a greater variety of lexical items to fill structural bundles, and utilize functional bundles differently. These distinctions appear to reflect each discipline's rhetorical requirements. Additionally, Mahachulalongkornrajavidyalaya University students' writing demonstrated a difference in their use of bundles compared to native English speakers. Mahachulalongkornrajavidyalaya University students used more types and tokens of bundles overall, whereas native English speakers used less, although their writing was more advanced. Mahachulalongkornrajavidyalaya University students resorted to repetitive use of specific bundles more than native English speakers did. The study's implications for teaching and curriculum development are examined. The findings can be utilized in instruction by providing students with awareness-raising and guided practice opportunities to observe how bundles are utilized in native English speaker writing and assist them in applying this knowledge to their writing.

Keywords: Lexical Bundle; Four Word Lexical Bundle; Native English Speaker; Nonnative English Speaker

Introduction

Writing can be a hard skill to learn and challenging to master because it's not just cognitive activity but a complex mental production that requires "planning, discipline and attention".¹ The most leading difficulties for second language learners are correctly determining what is written versus spoken language, checking subject-verb agreement, and writing coherent paragraphs.² Getting students to read and then be able to express themselves on paper is another challenge. Consequently, the student frequently encounters grammatical errors. Thus, it can lead to students copying and pasting rather than paraphrasing and summarizing. Although there are some challenges mentioned above, there are some additional factors that lead to language difficulties, including time

¹ Grami, G. M. A. "The Effects of Integrating Peer Feedback into University Level ESL Writing Curriculum: A Comparative Study in a Saudi Context". Unpublished doctoral dissertation, 2010, p. 9.

² Al Fadda, Hind. "Difficulties in academic writing: From the perspective of King Saud university postgraduate students". *English Language Teaching* 5, no. 3 (2012) : 123-130.

pressure because writing needs a lot of time and proper focus to collect and managed information and also develop ideas; furthermore, editing, revising, and proofreading requires a good knowledge of grammar and text coherence.

International students represent an essential part of MCU's students body. More than 1500 international students had joined Mahachulalongkornrajavidyalaya University. While the body of literature on the internationalization of Mahachulalongkornrajavidyalaya University is growing, the experience of these international students remains relatively limited, especially concerning the matter of transfer to the postgraduate level. They are all confronted with the difficulty of transiting into the academic culture, instructional methods, and the requirements for their selected faculties. Most international students are majoring in an international English program. Two thousand additional students are learning English at an Intense English Curriculum to achieve a degree from Mahachulalongkornrajavidyalaya University in Thailand. The university classroom has become an emphasis for writing-based learning both at the undergraduate and postgraduate levels. Therefore, study synthesis, argumentation, and drafting skills are essential for teachers to train non-native students. Writing in a second language is difficult for many students, and some of them are underprepared for the university classroom's linguistic demands after years of planning. When it comes to an academic paper, even native speakers face problems with writing, but non-native speakers' problems can be quite different since most of the nonnative speaker students lack proficiency in the English language.

There have been numerous claims that Mahachulalongkornrajavidyalaya University students sometimes cannot find a correct, precise, and comprehensible expression in writing. Several factors can lead to students' problems with writing. The complexities of the writing challenge may be one of the reasons for the difficulty. Byrne identifies many cognitive, emotional, and linguistic issues that present challenges to writers in both first and second language.³ While, Raimes thinks that: "When students speak about how hard it is to write in a second language, they are referring to the difficulties of not only getting the right words and implementing proper grammar but also finding an expressing idea in

³ Byrne, D., **Teaching Writing Skills**, (London: Longman Press, 1988), p. 4.

a foreign language.”⁴ The standards of academic writing are too difficult to uphold even for fluent English-speaking students. These issues may arise from a lack of understanding of some of these factors, such as academic writing structure, meaning, different forms of academic writing, and writing methods. This study gives a general insight into the research pattern on academic writing that can be influential to those who wish to contribute to this field of research. This research aims to make a meaningful contribution to explain better and increase understanding at vast, contextual, theoretical, and applied levels. The overall objectives of this research are as follows:

Objectives of the Research

1. To explore the most frequently used four-word lexical bundles in native English speakers’ and MCU students’ writing.
2. To identify the structural types of lexical bundles that were employed in native English speakers and MCU students.
3. To discover the variation of functional usage between native English speakers and MCU students writing.

Research Design

Mixed method was used in this study for collecting data. The quantitative approach discovered the number of bundles, the qualitative approach conducted a more in-depth investigation of the environments in which they appeared. Further investigation revealed that academic and lexical words were used in different ways. It is important to note that the term “appropriateness” refers to quantitative analysis, while “density” refers to qualitative analysis.

Corpora and Methods

This research was performed using 186 thesis which were written by MCU students and English native speakers. These collections of the articles were in the fields of Applied Linguistics, English international program, and English Language Teaching and Learning. The data of MCU students’ written articles were taken from one source (MCU E-Library), while the source of English native speaker was collected from two different sources first

⁴ Raimes, A., **Techniques in Teaching Writing**, (Oxford: OUP, 1983), p. 13.

one is British National Corpus (BNC) and second is Open American National Corpus (OANC). The articles of MCU students from 2015 to 2020 were retrieved from (MCU library) in PDF format. Some articles weren't available or inaccessible online from (MCU E-Library); therefore, some of the articles of MCU students were taken from printed papers then manually rescan them into the computer and by using google photo to text all physical paper turn into text which becomes an editable document. Since this study aimed to explore the abstracts, introduction, results, and conclusion, this design corpus will be excluded the chapters or sections that are not a part of the intended parts to explore.

Construction of the Corpus

After gathering the data, the researcher put together all of the abstracts, introductions, results, and conclusions from the papers of MCU students and native English speakers. As was already said, the remaining chapters or parts of the text will be left out. Two hundred eighty-six articles were written by MCU students and native English speakers, will be analyzed. The compilation led to 372 abstracts, introductions, findings, and conclusions, with a total word count of 526126 to 812153. For the analysis, each article was given a number from 1 to 350.

Calculation of the Sample Size

When it came to estimating the sample size, the researcher used the formula developed by Yamane⁵ and Raosoft Calculator⁶. The detailed discussion of the Yamane formula (1967) and Raosoft generator will be shown bellows;

Table 1. Calculation of the Sample Size

$$n = \frac{N}{1+N(e)^2}$$

n = Sample size

N = Population size

e = Sampling error or tolerable error

⁵ Yamane, Taro, **Statistics: An introductory analysis**, (2nd Edition, New York: Harper and Row, 1967).

⁶ Raosoft Calculator, "**Sample size calculator Raosoft**", On online, <http://www.raosoft.com/samplesize.html> (accessed January 5, 2021).

The present study's sample size is calculated as follows;

$$n = \frac{350}{1+300(0.05)^2} = 1.875$$

$$n = \frac{350}{1.875}$$

186

Sampling Process

All the samples for the analysis were collected using a simple random method. Every corpus is given a number from 1 to 350, and all collections of abstracts, introduction, results of findings, and conclusion have numbers assigned to them for convenience. In short, all of the papers had their numbers in the correct series to generate and more accessible when performing the random selection. To standardize the results, the researcher selects the numbers that were generated from google "Random Number Generator", with a total collection of 186, which will then be divided equally for MCU students and English native speakers with the amount of 186 each. This process is randomized selection; thus, bias toward the number or selection is 100% free. In order to ensure fair selection, the researcher will not be carried out by re-rolling or regenerating the number that is already shown once.

Data Preparation For the Software Program

This study used AntConc software to identify the forms of lexical bundles and linguistic features. AntConc is a sophisticated text analysis program that provides information about the text found inside one or more text files if batch processing is chosen. Another built-in AntConc method is "Clusters/N-Grams," which can scan the corpus for N length clusters, effectively detecting distinct but related word patterns. Thus, the samples collected during the sampling process went through a data preparation process before being analyzed further. If collected samples are printed version, then the researcher will scan them manually to a computer, then all the scan papers will process to IMAGE to TEXT generator which is available for free from Google and other online web sites. After all the documents are converted to (*.txt) format, researchers will put them together for analysis in AntConc software.

Lexical Bundle Analysis

For lexical analysis, the study involved assembling a list of the most widely used and efficient vocabulary references from the literature. Target bundles are being studied in this study, based on Cortes' findings⁷. Moreover, this study will also closely follow Lores' identified two methods for defining common word combinations⁸. For functional classification of lexical bundles, the LBs in this study were classified using the same taxonomies that researchers have used for LB classification in the past. This classification is based on Biber et al and other researchers. Each word was then classified according to the taxonomy based on the frequency of use in each corpus in relation to the other words in the taxon to which it is assigned. The functional categorization used the Resource-Based View to denote whether a unit is syntactic, semantic, or pragmatic.

Results of the study

1. Frequency of Lexical Bundle

This section addresses the first part of the research objective, which is to determine the overall frequency of lexical bundle in MCU students' (MS) writing compared to Native English Speaker (NES).

| NES Lexical Bundles | Freq | Range | MS Lexical Bundles | Freq | Range |
|---------------------|------|-------|--------------------|------|-------|
| on the other hand | 126 | 37 | is one of the | 117 | 56 |
| in the case of | 57 | 24 | on the other hand | 89 | 48 |
| the end of the | 52 | 28 | one of the most | 87 | 50 |
| the fact that the | 52 | 24 | the results of the | 48 | 31 |
| the name of the | 49 | 23 | as well as the | 46 | 33 |
| one of the most | 46 | 27 | i would like to | 46 | 17 |
| it is necessary to | 40 | 7 | at the same time | 44 | 24 |
| at the same time | 39 | 27 | as a result of | 38 | 21 |
| it is possible to | 30 | 10 | it was found that | 36 | 16 |

⁷ Cortes, Viviana Solange, **Lexical bundles in academic writing in history and biology**, (Thesis Dissertation, Northern Arizona University, ProQuest Dissertations Publishing, 2002), pp. 1-247.

⁸ Lorés, Rosa, "On RA abstracts: from rhetorical structure to thematic organisation", **English for specific purposes** 23, no. 3 (2004) : 280-302.

| NES Lexical Bundles | Freq | Range | MS Lexical Bundles | Freq | Range |
|---------------------|------|-------|------------------------|------|-------|
| it is difficult to | 29 | 19 | it is important to | 34 | 26 |
| is one of the | 28 | 21 | the purpose of this | 33 | 23 |
| the use of the | 26 | 16 | the study of the | 32 | 21 |
| in the context of | 26 | 13 | the end of the | 32 | 18 |
| on the basis of | 26 | 9 | of the most important | 31 | 23 |
| it is hard to | 23 | 18 | an important role in | 31 | 17 |
| in the form of | 23 | 11 | in the form of | 29 | 20 |
| in the middle of | 22 | 16 | in the process of | 28 | 18 |
| for the first time | 21 | 14 | as one of the | 27 | 20 |
| as a result of | 21 | 10 | the scope of the | 27 | 10 |
| the nature of the | 20 | 12 | it is necessary to | 25 | 20 |
| the use of a | 20 | 10 | to the development of | 25 | 12 |
| the rest of the | 19 | 17 | one of the main | 24 | 15 |
| as well as the | 19 | 14 | for the development of | 23 | 16 |
| it is impossible to | 16 | 12 | was found that the | 22 | 16 |
| in addition to the | 15 | 8 | in the field of | 21 | 18 |
| on the one hand | 14 | 12 | the results of this | 19 | 15 |
| in terms of the | 14 | 6 | in terms of the | 18 | 13 |
| I would like to | 13 | 11 | this study is to | 15 | 13 |
| in the field of | 12 | 10 | The rest of the | 14 | 9 |
| as one of the | 10 | 7 | on the basis of | 10 | 12 |

The investigation of the lexical bundles of native English speakers revealed 30 four-word lexical bundles. In native writing, the most frequently used four-word lexical bundles were “*on the other hand*,” “*in the case of*,” “*the end of the*,” “*the fact that the*,” and “*the name of the*.” The number of four-word lexical bundles, *on the other hand*, was greatly inflated in the native English speakers’ writing, and the analysis produced 30 four-word lexical bundles. “*is one of the*,” “*on the other hand*,” “*on the other hand*,” “*one of the most*,” and “*the results of the*” were the most frequently occurring lexical bundles in the MCU students’ writing. Some past research has shown

that non-native English speakers use fewer and fewer lexical bundles, but this study shows that this is not true. Still, the results back up what other research has found, which is that non-native speakers use a wide range of lexical bundles. Native English speakers use more tokens and bundle types than MCU students. The above table shows the bundles that are used most often by all groups. This shows that bundles are utilized in the same types of texts by both groups. However, native English speakers use them much more often in their writing.

2. Structural Types of Lexical Bundles

This section will attempt to address the second research objective, which is to identify the structural types of lexical bundles and the most commonly used structural types in NES and MS writing.

| Structure | NES | | MS | |
|--|-----------|------------|-----------|------------|
| | Frequency | Percentage | Frequency | Percentage |
| 1st/2nd person pronoun + VP fragment | 1 | 3.1% | 1 | 3.1% |
| Anticipatory it + copula be + adjective phrase | 5 | 16.1% | 3 | 9.4% |
| Anticipatory it + passive verb + that | | | 1 | 3.1% |
| Conjunction | 1 | 3.2% | 1 | 3.1% |
| Copula be + noun phrase/adjective phrase | 1 | 3.2% | 1 | 3.1% |
| Noun phrase + copula be | | | 1 | 3.1% |
| Noun phrase with of-phrase fragment | 8 | 25.8% | 9 | 28.1% |
| Noun phrase with other post-modifier fragm. | | | 1 | 3.1% |
| Other prepositional phrase fragment | 4 | 12.9% | 2 | 6.3% |
| Prepositional phrase with embedded of-phrase | 10 | 32.3% | 11 | 34.4% |
| Verb phrase with passive verb | | | 1 | 3.1% |
| Total | 30 | 100.0% | 31 | 100.0% |

The results indicate that native English speakers and MCU students employed *Prepositional phrase with embedded of-phrase* the most in structural types of lexical bundle, with 32.3 percent founded in native English speakers' writing and 34.4 percent founded in MCU students' writing, respectively. *Noun phrase with of-phrase fragment* was the second most common structure for both groups, accounting for 25.8 percent in

native English speakers' writing and accounting for 28.1 percent in MCU students' writing. *Anticipatory it + copula be + adjective phrase* comprised 16.1 percent in native English speakers' writing and 9.4 percent in MCU students' writing. The fourth most prevalent structure was *Other prepositional phrase fragment*, which accounted for 12.9 percent in native English speakers' writing and 6.3 percent in MCU students' writing. MCU students perform better than native English speakers for similar structure types, and their work has a greater variety of structure types than that of native English speakers.

3. Variation of Function

This section will attempt to address the third research objective, which is to determine the function of a four-word lexical bundle and the most commonly used functional types in NES and MS writing.

| Function | NES | | MS | |
|-----------------------------|------------------------|--------------|------------------------|--------------|
| | Function Subcategories | Percentage | Function Subcategories | Percentage |
| Research-Oriented | 9 | 30% | 7 | 23.3% |
| Description | 3 | 10.0% | 1 | 3.3% |
| Location | 1 | 3.3% | | |
| Quantification | 3 | 10.0% | 3 | 10.0% |
| Procedure | 2 | 6.7% | 3 | 10.0% |
| Text-oriented | 13 | 43.4% | 16 | 53.3% |
| Framing | 6 | 20.0% | 3 | 10.0% |
| Objective | | | 1 | 3.3% |
| Resultative | 2 | 6.7% | 9 | 30.0% |
| Transition | 5 | 16.7% | 3 | 10.0% |
| Participant-oriented | 8 | 26.7% | 7 | 23.3% |
| Stance | 8 | 26.7% | 7 | 23.3% |
| Total | 30 | 100% | 30 | 100% |

As the table above shows, both native English speakers and MCU students mostly used lexical bundles that were focused on *text-oriented*. There were 13 lexical bundles that made up 43.4% of the writing of native English speakers, and 16 lexical bundles that

made up 53.3% of the writing of MCU students in text-oriented. Based on this number, both groups use a lot of “*text-oriented*” lexical bundles. “*Research-Oriented*” was the second most common function. It was responsible for 9 lexical bundles with 30 percent of all bundle types, in the writing of native English speakers and 7 lexical bundles with 23.3 percent of all bundle types, in the writing of MCU students. Third most often used was “*participant-oriented*.” In the writing of native English speakers, there were 8 lexical bundles, which made up 26.7 percent of all bundle types. In the writing of MCU students, there were 7 lexical bundles, which made up 23.3 percent of all bundle types. Since text-oriented meaning is the most common use of lexical bundles, this shows that both native English speakers and MCU students have about the same number of four-word lexical bundles. In Research-Oriented and Participant-Oriented, it was shown that native English speakers used more lexical bundles than MCU students, but overall, MCU students used more function types than native English speakers.

Conclusion

The present study provides a comprehensive investigation into the use of LBs by native English speakers and MCU students in academic writing. The results showed that both native English speakers and MCU students used a greater number of lexical bundles in their writing, and they have utilized structural types and functions in their writing as well.

1) In accordance to the first research objective, which assessed the overall frequency of native English speakers’ and MCU students’ writing. The overall frequency of types and tokens indicates that native English speakers exceeds MCU students; yet, the use of four-word lexical bundles in MCU students’ writing is larger than in native English speakers. The frequency findings corroborate Chen and Baker, Wei and Lei’s, and Cortes’ findings.

2) For the second part of the research objective, which was looking at structural types of bundles, native English speakers displayed differences in two structures; however, in terms of structural types such as Prepositional phrase with embedded of-phrase and Noun phrase with of-phrase fragment, MCU students still use more based on those two structural types.

3) The final research question examined how the functional use of lexical bundles varied between native English speakers and MCU students within disciplines and discovered discrepancies in how native English speakers and MCU students employed these bundles. These distinctions held true across all functional categories, while both native English speakers and MCU students generally employed more text-oriented bundles. The findings indicated that native English speakers and MCU students employed a variety of structures to create lexical bundles. The text-oriented lexical bundles had the highest proportion when comparing to another functions. The findings showed that there was no significant variation of function in both native English speakers' and MCU students' writing.

Discussion

The vast majority of studies have focused on native students studying in English-speaking countries and, with a few exceptions have tended to focus on samples of nonnative speakers at later stages of English language training. This dissertation aims to enrich our understanding of the discourse conventions found in academic writing by investigating the use of English among native and nonnative English language learners at different stages of their acquisition.

In sum, the author argues that: The reader, in particular, should be aware of the limitations of a single study. First, just as this study includes a relatively small sample size of only 286 writers, the number of words used also limits the findings. Second, the data collection consisted solely of participants' written papers and taken from a single university for MCU students. Finally, due to the interconnected nature of a more considerable study across multiple universities, the results and findings reported here represent a very preliminary analysis of the limited data available. Considering the need for additional research, the findings of this study provide the first insight into how graduate students at this university write and how they compare to native English speakers writing within the same corpus. This study adds greatly to our comprehension of second language writing and student demands. MCU international students have received little attention, and evaluating their writing can help us understand what they are capable of with their current level of training. Additionally, though the major objective

of this study was to describe the distinctions between NES and MS writing, the findings had implications for improving the MCU University's writing program. Finally, while this study cannot provide a thorough description of lexical bundles in different fields, it is hoped that future research will build on the proposals made above in order to advance our understanding of lexical bundle disciplinary variation. In other words, researchers would be able to widen their perspectives on linguistic conventions, practices, and necessary linguistic characteristics in order to develop efficient writing methods. Increased awareness of lexical bundles, structure, and function would enable inexperienced writers to convey their research more effectively. The study's conclusions may add to the academic realm's long-term development. MCU students would be encouraged to publish more effective and successful articles by implementing the knowledge gained from this study.

Bibliography

- Al Fadda, Hind. "Difficulties in academic writing: From the perspective of King Saud university postgraduate students". **English Language Teaching** 5, no. 3 (2012) : 123-130.
- Byrne, D. **Teaching Writing Skills**. (London: Longman Press, 1988). p. 4.
- Cortes, Viviana Solange. **Lexical bundles in academic writing in history and biology**, (Thesis Dissertation, Northern Arizona University, ProQuest Dissertations Publishing, 2002). pp. 1-247.
- Grami, G. M. A. "**The Effects of Integrating Peer Feedback into University Level ESL Writing Curriculum: A Comparative Study in a Saudi Context**". Unpublished doctoral dissertation, 2010. p. 9.
- Lorés, Rosa. "On RA abstracts: from rhetorical structure to thematic organisation". **English for specific purposes** 23, no. 3 (2004) : 280-302.
- Raimes, A. **Techniques in Teaching Writing**. (Oxford: OUP, 1983). p. 13.
- Yamane, Taro. **Statistics: An introductory analysis**. 2nd Edition, New York: Harper and Row, 1967.
- Raosoft Calculator. "**Sample size calculator Raosoft**", [online]. Available: <http://www.raosoft.com/samplesize.html> [January 5, 2021].