

# LEXICAL BUNDLES IN NON-NATIVE LEARNER CORPORA

Engku Haliza Engku Ibrahim

English Language Division, CELPAD, International Islamic University Malaysia  
[ehaliza@iium.edu.my](mailto:ehaliza@iium.edu.my)

Received: 18 March 2020 Revised and Accepted: 19 June 2020

**ABSTRACT:** Research has indicated that learner's successful assimilation of lexical bundles is one of the indicators of writing competency. Additionally, literature has also shown that different types of writing require different types of lexical bundles and it is essential that learners are able to differentiate between personal stance bundles and those of academic prose. As such, this study intends to investigate the grammatical structures of frequently used lexical bundles and as well as find out if there is a difference in frequency of use of lexical bundles in different types of essays. The data were collected from 50 graphic-oriented essays and 50 open-ended essays of upper-intermediate English language learners. An analysis of the data in this study, among others, showed that our learners used less structural classifications in open-ended essays when compared to graphic-oriented essays. Finally this paper presents discussions of the pedagogical implications drawn from the analysis of the data presented.

**KEYWORDS:** Lexical Bundles, Formulaic Sequences, Vocabulary Acquisition

## I. INTRODUCTION

It is a known fact that vocabulary mastery is essential in language acquisition; so much so that according to (1) without vocabulary, nothing could be conveyed. This is even more so when learners need to acquire the language for academic purposes. However, research has shown that the vocabulary of a language does not just consist of individual words. More often than not, these words consist of multi-word items. In fact, Biber and Conrad (2) claim that multi-word sequences make up a significant portion of academic prose. Research has also shown that learners' ability to use appropriate multi-words indicates mastery of the target language; failure to correctly use these multi-words would indicate non-native like writing proficiency. Despite its importance there are few research that focus on the use of multi-words specifically for academic purposes (3). Thus this research intends to fill the knowledge gap. The implication for the learner is that these words need to be learned as multi-word units rather than individual words. As argued by Biber, Conrad (4), corpus study is the best way to study natural use of language which will then allow teachers to specifically address them in the language classroom. Thus the choice of corpus linguistics for the approach of this research, in an attempt to examine our learners choice and range of lexical bundles.

## II. LITERATURE REVIEW

Lexical bundles (2, 4-8) is referred to in the literature using many terms such as phrasal expressions (9), formulaic sequences (10, 11) or recurrent word combinations (7). They consist of 3 or more words that do not fall in the category of idiomatic expressions, nor are they complete structural units (12, 13). The characteristic that sets lexical bundles apart is that they are sequences of words that frequently occur together and this can be evident in both written and spoken corpora. Some examples of lexical bundles include expressions like at the same time, to be able to, as a result of and on the other hand. Being an important feature of academic discourse genre, they perform a variety of discourse functions (14).

### Importance of Lexical Bundles in Language Acquisition

What researchers have agreed on is that learners need to have the ability to correctly use lexical bundles in order for their writing to appear native-like. Specifically, it can also be said that proper use of lexical bundles contribute, to some extent, to discourse coherence and cohesion. Allen (15) contends that mastery of lexical bundles is essential in academic writing as it indicates learners ability to converge to specific discourse registers. However, the fact of the matter is, lexical bundles are difficult to teach as they are not defined as conventional grammar structures (16). Certain features of lexical bundles serve similar functions in showing cohesion as that of Halliday and Hasan (17) conjunctive relations. Furthermore, as highlighted by Wray (18), formulaicity is essential to both the writer or speaker and the listener or reader as it facilitates discourse processing which helps to enhance in the perception of discourse coherence. However, research by Ädel and Erman (19) and Chen and Baker (20) all indicated that writings of non-native speakers contain fewer as well as less varied lexical bundles.

This points to the urgent need for learners to master lexical bundles as it is an indication of a mastery of academic discourse. Undoubtedly there are still ongoing debates as to how lexical bundles are acquired by learners. Wray (18), for instance, is sceptical that the lexical bundles found in learner corpus may not in fact be “stored and retrieved whole from memory” (18). It is however still useful to study learner production so that teachers could address formulaicity as part of their language teaching strategy as it was emphasised by (21) that when learners underuse formulaic expressions in their writings, their writings are often unacceptable.

**Studies on Lexical Bundles**

According to Biber and Conrad (2), corpus study is useful for studying actual language use of learners as they enable researchers to analyse actual distribution in natural texts written by learners. Cortes (6) for instance, carried out a study to compare her student writers’ use of lexical bundles to that of academic writers. Her study revealed her student writers used the target lexical bundles far less when compared to the academic writers. Due to the significance of lexical bundles in the writing of ESL learners, several studies have focussed on investigating learner gains in writing performance after being given instructions of their correct use. Jones and Haywood (21), for instance, gave a 10-week instruction to university students who were non-native speakers of English. His study deduced that his participants had minor gains in the use of formulaic sequences in the writing production, though it was reported that they had high motivation and predisposition towards the use of formulaic expressions. Similarly, Kazemi, Katiraei (22) conducted a study on MA TESL students in Iran. Their findings revealed that the teaching lexical bundles significantly helped their students’ writing ability. All these studies point to the significance of highlighting lexical bundles in the language classroom.

The main objective of this descriptive research is to investigate the use of lexical bundles of Malaysian ESL learners. Specifically, this research aims to compare if there are differences in learners’ use of lexical bundles in graphic-oriented essays and open-ended essays. The research questions of this research are:

1. Which grammatical structure of the lexical bundles was used the most?
2. What is the highest percentage of grammatical structure in graphic-oriented essays?
3. What is the highest percentage of grammatical structure in open-ended essays?

**III. METHODOLOGY**

The corpus of this study was taken from essay scripts of learners who took the English Proficiency Test (EPT) at the International Islamic University Malaysia. The responses were from learners at the high intermediate level of English proficiency. The EPT requires learners to produce two types of written responses; a graphic-oriented essay and an open-ended essay. For the purpose of this study 100 essay scripts (18,805 words) were used; 50 graphic-oriented essays (10,319 words) and 50 open-ended essays (17,753 words). Table 1 summarises the corpus of this present study.

Industrial sector, in this case represented by its primary subsector manufacturing, has played an imperative role on world’s economic development because of its characteristics of having a great capacity for productivity growth, technical and financial externalities, and increasing return to scale so that it can accelerate high economic growth rate (23). An economy dominated by manufacturing sector is expected to have rapid headway. Therefore, many countries compete each other to speed up structural transformation from agricultural sector. Asia itself, one of five world continents that has already been a principal contributor of World’s Gross Domestic Product (GDP) since 2011 based on United Nations Conference on Trade and Development (UNCTAD) data as shown in Figure 1(a), realizes this condition very well. More over, Asia is acclaimed as the biggest contributor of world manufacturing sector output since 2016 as more than 30 percent of world manufacturing sector output has been supported it by Asian Countries

**Table 1:** Frequency of bundles and word count from corpus of open-ended and graphic-oriented essays

TYPE OF ESSAY	NUMBER OF ESSAYS	WORD COUNT	FREQUENCY OF 3 TO 5 WORD BUNDLES
GRAPHIC ORIENTED ESSAYS	50	10,319	530
OPEN-ENDED ESSAYS	50	17,753	1,052
TOTAL	100	28,072	1,582

*The essays were then digitized and inserted into Wordsmith 2.0 (24) for the identification of lexical bundles*

**IV. FINDINGS AND DISCUSSION**

RQ1: The highest frequency of grammatical structure used in the present corpus

Table 2 summarizes the grammatical structure that were found in both types of essays investigated in this study. It can be concluded that the highest grammatical structure was found in both essay types was the noun phrase. This is consistent with the findings of Biber (12) and Dontcheva-Navratilova (25) where it was noted that in academic discourse, almost 70 per cent of the most common bundles consist of noun phrase elements or prepositional phrases which serve the function of discourse organizers and facilitate referential expressions. This is unlike the interactional discourse of conversations, where the majority of lexical bundles are stance expressions that contain a verbal element.

**Table 2:** Percentages of Structural Classification for open-ended essays and graphic oriented essays: a comparison

	<b>GRAPHIC-ORIENTED ESSAYS</b>	<b>OPEN-ENDED ESSAYS</b>	<b>TOTAL</b>
NOUN PHRASE + OF AND OTHER NOUN PHRASE	68.1%	24.4%	92.5%
PREPOSITIONAL PHRASE	8.1%	23.2%	31.3%
PASSIVE + PREPOSITIONAL PHRASE FRAGMENT	7.4%	16.8%	24.2%
ANTICIPATORY IT + VERB/ADJECTIVE	4.5%	6%	5.1%
BE + NOUN/ADJECTIVAL PHRASE	3.2%	0.5%	3.7%
VERB/ADJECTIVE CLAUSE FRAGMENT	1.1%	3.1%	4.2%
PRONOUN PHRASE	0.9%	4.6%	5.5%
ADVERBIAL PHRASE	0.6%	1.5%	2.1%
1ST PERSON PRONOUN + DEPENDENT CLAUSE FRAGMENTS	1%	5.5%	6.5%
2ND PERSON PRONOUN + DEPENDENT CLAUSE FRAGMENTS	-	0.2%	0.2%
3RD PERSON PRONOUN + DEPENDENT CLAUSE FRAGMENTS	-	1.5%	1.5%
YES/NO QUESTION FRAGMENTS	-	0.4%	0.4%
WH-QUESTION FRAGMENTS	-	0.8%	0.8%
OTHERS	0.6%	0.6%	1.2%
TOTAL	100%	100%	

RQ 2: The most used grammatical structure in graphic-oriented essays

Table 3 shows the types of grammatical structure found in the graphic-oriented essays of this study. The findings revealed that in the graphic-oriented essays, the noun phrase (68.1%) was by far the highest type of grammatical structure found in our student writings, followed by prepositional phrases (8.1%). This could be attributed to the nature of the task required in graphic-oriented essays that necessitates students to comprehend a graph and consequently describe the graph, analyse data and interpret results. This activity would require students to use a higher degree of fixed expressions such as in comparison to, the end of the and the average number. The findings of this study revealed that the most used lexical bundle in the graphic-oriented essay was the average number (70%), followed by the number of (65) and the age of (30%). Compared to the percentages of noun phrase constructions found in the open ended essays, there were less variation of the noun phrases used in the graphic oriented essays.

**Table 3:** Percentages of Structural Classification in open-ended essays

	<b>Graphic-Oriented Essays</b>
Noun Phrase + of and Other Noun Phrases	68.1%
Prepositional Phrase	8.1%
Passive + Prepositional Phrase Fragment	7.4%
Anticipatory It + Verb/Adjective	4.5%
Be + Noun/Adjectival Phrase	3.2%
Verb/Adjective Clause Fragment	1.1%
Pronoun Phrase	0.9%
Adverbial Phrase	0.6%
1st Person Pronoun + Dependent Clause Fragments	1%
2nd Person Pronoun + Dependent Clause Fragments	-
3rd Person Pronoun + Dependent Clause Fragments	-
Yes/No Question Fragments	-

Wh-Question Fragments	-
Others	0.6%
TOTAL	100%

RQ 3: The most used grammatical structure in open-ended essays

Table 4 shows the types of grammatical structure in the open-ended essays. The findings revealed that in the open-ended essays, the noun phrase (24.4%) was also the highest type of grammatical structure used followed by prepositional phrases (23.2%). In comparison to graphic-oriented essays, learners used less noun phrases in open-ended essays. Learners also used 2nd person pronoun + dependent clause fragments (0.2%), 3rd person pronoun + dependent clause fragments (1.5%), Yes/No question fragments (0.4%) as well as wh-question fragments (0.8%) in the open-ended essays. These 4 grammatical structures were not used at all in the graphic-oriented essays. What can be seen is that learners used a wider repertoire of lexical bundles in the open-ended essays compared to the graphic-oriented essays. Some examples of the lexical bundles found in our data include that hobbies should be (43%), that there are (15%) and that we can (8%). In comparison to the lexical bundles used by our students in the graphic oriented essay, there were more variation of lexical bundles found in the open-ended essays which is demonstrated by the lower percentages of lexical bundles found.

**Table 4:** Percentages of Structural Classification in graphic-oriented essays

	Open ended essays
Noun Phrase + of and Other Noun Phrases	24.4%
Prepositional Phrase	23.2%
Passive + Prepositional Phrase Fragment	16.8%
Anticipatory It + Verb/Adjective	6%
1st Person Pronoun + Dependent Clause Fragments	5.5%
Pronoun Phrase	4.6%
Verb/Adjective Clause Fragment	3.1%
Adverbial Phrase	1.5%
3rd Person Pronoun + Dependent Clause Fragments	1.5%
Wh-Question Fragments	0.8%
Others	0.6%
Be + Noun/Adjectival Phrase	0.5%
Yes/No Question Fragments	0.4%
2nd Person Pronoun + Dependent Clause Fragments	0.2%
Total	100%

**IV. CONCLUSION**

The objective of this study was to find out if there are different patterns of grammatical structure found in the two type of essays; graphic-oriented-essays and open-ended essays. Our findings reveled that were differences in the number of grammatical structure used, though the noun phrase as well as prepositional phrase were the two most used structural classifications in both essay types. Even though this has been found to be the case across most research findings, previous studies have also shown that non-native learners tend to have a smaller repertoire of lexical bundles (19, 20). Further detailed analysis of the noun phrase + of is needed to find out if there is a tendency for our students to overuse this structure when compared to other grammatical structures. Our findings also revealed that there was a difference in the repertoire of grammatical structures between the two types of essays. Our learners displayed a wider repertoire of grammatical structure types in the open-ended essays. This study further affirms that teachers should explicitly address these fixed forms in order to create noticing and raise learner awareness of their occurrences. This study also emphasises the significance of explicitly addressing lexical bundles in the language curriculum to ensure that learners are able to effectively and successfully produce register convergent discourse in their academic writings.

**V. ACKNOWLEDGEMENT**

This research is supported by a grant from the International Islamic University Malaysia.

**VI. REFERENCES**

[1] Wilkins D. Linguistics in language teaching. London: Arnold. Word [Def 1]. 1972.  
 [2] Biber D, Conrad S. Lexical bundles in conversation and academic prose. Language and Computers. 1999;26:181-90.  
 [3] Omidian T, Beliaeva N, Todd L, Siyanova-Chanturia A. The use of academic words and formulae in L1 and L2 secondary school writing. New Zealand Studies in Applied Linguistics. 2017;23(2):39.

- [4] Biber D, Conrad S, Cortes V. If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*. 2004;25(3):371-405.
- [5] Biber D, Barbieri F. Lexical bundles in university spoken and written registers. *English for specific purposes*. 2007;26(3):263-86.
- [6] Cortes V. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*. 2004;23(4):397-423.
- [7] De Cock S. Repetitive phrasal chunkiness and advanced EFL speech and writing. *Language and Computers*. 2000;33:51-68.
- [8] Hyland K. As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*. 2008;27(1):4-21.
- [9] Martinez R, Schmitt N. A phrasal expressions list. *Applied linguistics*. 2012;33(3):299-320.
- [10] Schmitt N. *Formulaic sequences: Acquisition, processing and use*: John Benjamins Publishing; 2004.
- [11] Wray A. *Formulaic language: Pushing the boundaries*: Oxford University Press; 2008.
- [12] Biber D. *University language: A corpus-based study of spoken and written registers*: John Benjamins Publishing; 2006.
- [13] Biber D, Johansson S, Leech G, Conrad S, Finegan E, Quirk R. *Longman grammar of spoken and written English*: longman London; 1999.
- [14] Nausa RA. Lexical bundles in Academic presentations given by PhD level EAP students: a literature review. *Gist: Education and Learning Research Journal*. 2013(7):259-70.
- [15] Allen D. Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*. 2010;1:105-27.
- [16] Cooper PA. *Academic vocabulary and lexical bundles in the writing of undergraduate psychology students 2016*.
- [17] Halliday MAK, Hasan R. *Cohesion in english*: Routledge; 2014.
- [18] Wray A. Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*. 2000;21(4):463-89.
- [19] Ädel A, Erman B. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes - ENGL SPECIF PURP*. 2012;31:81-92.
- [20] Chen Y-H, Baker P. Lexical bundles in L1 and L2 academic writing. *Language learning & technology*. 2010;14(2):30-49.
- [21] Jones M, Haywood S. Facilitating the acquisition. *Formulaic sequences: Acquisition, processing, and use*. 2004;9:269-92.
- [22] Kazemi M, Katiraei S, Rasekh AE. The impact of teaching lexical bundles on improving Iranian EFL students' writing skill. *Procedia-Social and Behavioral Sciences*. 2014;98:864-9.
- [23] Felipe J. *Asia's Industrial Transformation: The Role of Manufacturing and Global Value Chains (Part 1)*. Asian Development Bank Economics Working Paper Series. 2018(549).
- [24] Scott M. *WordSmith tools manual*: University Press; 1998.
- [25] Dontcheva-Navratilova O. Lexical bundles in academic texts by non-native speakers. *Brno Studies in English*. 2012;38(2):41-62.