

ICH 2019
International Conference on HumanitiesLEXICAL BUNDLES IN ACADEMIC WRITING: THE ISSUE OF
SPECIFICITYAng Leng Hong (a)*
*Corresponding author

(a) School of Humanities, Universiti Sains Malaysia, Penang, Malaysia, lenghong@usm.my

Abstract

The contextual knowledge of a word is closely related to the knowledge of phraseological sequences as words are often used in the phraseological forms. Owing to the importance of phraseological knowledge, much has been done to examine the phraseological sequences for various purposes, including for English for Academic Purposes (EAP). In EAP settings, scholars have argued for the two different approaches to EAP, i.e. discipline-specific and common-core. As such, it is necessary to examine the issue of specificity in EAP with regard to the use of phraseological sequences such as lexical bundles. This study therefore aims to identify lexical bundles in journal articles in the field of International Business Management (IBM). Following corpus-driven approach the corpus analysis software, *Collocate 1.0* was used to extract three- to five-word combinations. These combinations were manually checked to exclude meaningless combinations. To determine to degree of specificity of the lexical bundles, the final lists of lexical bundles were compiled and compared with lexical bundles in Academic Formulas Lists (AFL) using log-likelihood test. The comparison reveals that lexical bundles in the IBM corpus are relatively specific as compared with the lexical bundles in AFL which are derived using common-core approach. A discipline-specific approach to the teaching and learning of lexical bundles in EAP settings is therefore advocated to enhance EAP syllabuses and instruction.

2357-1330 © 2020 Published by European Publisher.

Keywords: English for Academic Purposes, corpus-driven, phraseological sequences, lexical bundles, specificity.

1. Introduction

Academic study and writing require unique demands on language users as constructions and patterns of academic work are very different from the conventions that language users are more familiar with, for example the conversational register. Thus, the process of adapting to a hitherto unfamiliar register may pose difficulties for language learners at tertiary institutions. Research have found that university students have problems with using accurate and effective expository language in the academic register. In the case of non-native students in particular, these problems are compounded by the additional complexities involved in mastering the language itself. In most cases, university students have to write and publish academic research articles despite not having received the necessary training for the task. As a result of this shortcoming, scholars as well as practitioners in the field of English for Academic Purposes (EAP) have started examining the linguistic and textual features of academic writing in various disciplines from the linguistic and pedagogical perspectives.

2. Problem Statement

With the flourishing of corpus-driven phraseological research since the last decade, attention has been shifted to examining and building lists of academic phraseological sequences for EAP curriculum. For instance, in a corpus-driven study of academic discourse, Simpson-Vlach and Ellis (2010) employed a combination of statistically-driven approach and teacher insights to identify and extract a list of the most useful lexical bundles, which they termed Academic Formulas List (henceforth AFL). Simpson-Vlach and Ellis (2010) identified academic lexical bundles common to many academic disciplines that have high frequency and are of general and academic use. They therefore concluded that a general approach to EAP is sufficient to derive lists of common core academic phrases that transcend disciplinary boundaries. Their conclusion followed ideas pioneered by Zamel (1993) who strongly advocated a common-core approach to EAP courses whereby EAP instructors should focus on language forms common to all disciplines. Nevertheless, it was argued that each academic discipline has its own subject-specific conventions (Green & Lambert, 2018). Hyland (2002, 2006), a strong proponent of discipline-specific approach to lexical bundles refuted the idea of core academic clusters by demonstrating variations in the frequencies and functional uses of academic lexical bundles in different academic domains. According to Hyland, there are significant amount of formalities in academic texts, which are characterised by the use of subject-specific vocabulary. The issue of specificity has thus challenged instructors and linguists in the field of EAP to take a stance on how language should be perceived, that is whether language forms and features are transferable across different academic disciplines or specific to particular fields or disciplines. There are differing views with regard to the approaches to phraseology for EAP and this issue is still debatable in the field. It is therefore necessary for researchers to continue exploring phraseological sequences such as lexical bundles in academic discourse for the sake of further enhancing EAP instruction and curricula.

2.1. Definition and Previous Studies on Lexical Bundles

Lexical bundle was first defined and studied in detail by Biber, Johansson, Leech, Conrad, and Finegan in a chapter of the Longman Grammar of Spoken and Written English (henceforth LGSWE)

(1999), their exhaustive and comprehensive corpus study of grammar in English language. This seminal work deserves attention here as most studies on lexical bundles are largely based on the definition and framework proposed by Biber et al. (1999). According to Biber et al. (1999: 989-990), lexical bundles are “bundles of words that show a statistical tendency to co-occur... as recurrent expressions, regardless of their idiomaticity, and regardless of their structural status”. Lexical bundles are seen as sequences of word forms that are found frequently in both written and spoken discourses. They are usually identified empirically and extracted automatically from a corpus using the relevant corpus analysis software. In relation to lexical bundle research in academic genres, numerous studies have been conducted on lexical bundles to examine the use of lexical bundles by native and non-native speakers, and expert and novice writers (e.g., Chen & Baker, 2016; Pan et al., 2016; Bychkovska & Lee, 2017; Esfandiari & Barbary, 2017; Kwary et al., 2017; Hyland & Jiang, 2018; Shin et al., 2018; Lu & Deng, 2019; Shin, 2019; Jeong & Jiang, 2019; Wright, 2019). Nevertheless, little is known with regard to the approaches to lexical bundles in academic settings as the issue of specificity remains largely unexamined.

3. Research Questions

Specifically, this study addresses the following question:

How do lexical bundles in journal articles in the field of International Business Management (IBM) differ from those in AFL?

4. Purpose of the Study

This study compares lists of lexical bundles representing IBM and AFL (Simpson-Vlach & Ellis 2010) to determine the specificity of the lexical bundles in this study. Following common-core approach, AFL is a list of lexical bundles retrieved from a corpus of academic writing sampled across four academic disciplines: Humanities and Arts, Social Sciences, Natural Sciences/Medicine and Technology and Engineering while the lexical bundles identified in this study represent lexical bundles extracted from a specialised corpus which contains only journal articles in the field of IBM.

5. Research Methods

The present study employed corpus-driven approach to identify and extract three- to five- word lexical bundles in a one-million word corpus with 138 original research articles taken from two international peer-reviewed journals relevant to IBM.

5.1. Identification of Lexical Bundles

The corpus analysis software, *Collocate 1.0* (Barlow, 2004) was employed to extract lexical bundles automatically by setting the span options, i.e. three to six words. *Collocate 1.0* extracts lists of word combinations using two statistics: frequency and Mutual Information (MI). Following the literature, three- to five-word combinations that occur at least 20 times per million words in the corpus and achieve MI value of at least 3.0 were extracted. These combinations were then manually inspected to exclude meaningless combinations that were extracted automatically. To determine the degree of specificity of the lexical bundles,

the lists of lexical bundles were compiled and compared with lexical bundles in AFL (Simpson-Vlach & Ellis, 2010) using log-likelihood test.

6. Findings

A total of 1055 lexical bundles of varying lengths remained on the list after the manual filter. These 1055 lexical bundles make up 2.19% of the more than one million words in the current corpus. Table 01 compares the top 50 lexical bundles in IBM corpus with the top 50 core academic lexical bundles proposed by Simpson-Vlach and Ellis (2010). The comparison of the results of the study with those of Simpson-Vlach and Ellis (2010) was to determine the specificity of the lexical bundles in this study. To reiterate, Simpson-Vlach and Ellis's list of academic formulas is a cross-disciplinary list of lexical bundles which uses a common-core approach to compile lexical bundles common in various academic disciplines. In contrast, the list of lexical bundles identified in IBM corpus is a discipline-specific list of lexical bundles, representing phraseological sequences which are seen specific in IBM. As shown in Table 01, there are different types of frequent academic lexical bundles found in IBM and AFL, respectively.

Table 01. Comparison of lexical bundles with AFL (2010)

Rank	Bundle in this Study (IBM)	Bundle in AFL (2010) (Core academic formulas across various disciplines)
1	more likely to	in terms of
2	in order to	the use of
3	as well as	in order to
4	in terms of	as well as
5	the number of	the number of
6	The relationship between	there is a
7	the level of	part of the
8	the impact of	a number of
9	are more likely	the fact that
10	are more likely to	it is not
11	the effect of	there is no
12	the effects of	the case of
13	the importance of	in which the
14	likely to be	in the case
15	the host country	in the case of
16	in this study	based on the
17	as a result	the presence of
18	the results of	due to the
19	based on the	as a result
20	the role of	the role of
21	are likely to	the development of
22	the extent to which	at the same
23	a number of	that there is
24	on the other	likely to be
25	the use of	the effect of
26	the presence of	the basis of

27	on the other hand	such as the
28	the development of	the same time
29	in addition to	with respect to
30	in the host	the effects of
31	the context of	at the same time
32	of this study	it can be
33	related to the	is that the
34	firms in the	on the basis
35	the case of	the importance of
36	consistent with the	in this case
37	is likely to	a variety of
38	of the firm	in relation to
39	is consistent with	can be used
40	the influence of	the context of
41	the likelihood of	in other words
42	the value of	in the same
43	to control for	it may be
44	in other words	a series of
45	we find that	a result of
46	the fact that	is to be
47	in the context of	and in the
48	in line with	the nature of
49	in the same	for example the
50	with respect to	on the basis of

Table 02 presents the list of lexical bundles common in IBM corpus and AFL. Of all the frequent lexical bundles in IBM corpus, 36% of them are seen common in the AFL. Besides, the log-likelihood test performed shows that more than 70% of the shared lexical bundles are more specific to IBM corpus. The results of the comparison indicate that the lexical bundles in IBM corpus are relatively specific as compared with lexical bundles in AFL. A discipline-specific approach to the teaching and learning of lexical bundles for EAP is seen necessary as in this study, more than 60% of the lexical bundles were not found in AFL.

Table 02. Lexical bundles common in IBM corpus and AFL

No.	Lexical Bundle in the Current Study (IBM)	Lexical Bundle in AFL (Core Academic Lexical Bundles across Various Disciplines)	Log-likelihood
1	in order to	in order to	+ 60.79
2	as well as	as well as	+ 60.79
3	in terms of	in terms of	+ 60.79
4	the number of	the number of	+ 33.59
5	the effect of	the effect of	+ 126.31
6	the effects of	the effects of	+ 109.17
7	the importance of	the importance of	+ 104.07
8	likely to be	likely to be	+ 69.84
9	as a result	as a result	+ 33.79
10	the role of	the role of	+ 28.82
11	a number of	a number of	- 3.58

12	the use of	the use of	- 27.59
13	the presence of	the presence of	+ 4.11
14	the development of	the development of	+ 7.11
15	the case of	the case of	- 1.78
16	in other words	in other words	+ 11.37
17	the fact that	the fact that	- 22.23
18	in the same	in the same	+ 9.73

7. Conclusion

The findings of the study indicate that academic lexical bundles are discipline-specific. The findings of this study have implications on how EAP should be informed in language classrooms at tertiary institutions. The outcome of the analysis suggests that EAP instructors should follow a discipline-specific approach, particularly in the teaching of phraseological sequences such as lexical bundles. To sum up, there are two different views on how instructors and researchers approach EAP and this issue remains debatable in the field. It is necessary for scholars in the field to continue examining the various forms of phraseological sequences in academic discourse for enhancing EAP instructions and syllabuses in order to benefit language learners at tertiary institutions.

Acknowledgments

This work was supported by Universiti Sains Malaysia Short Term Grant (304/PHUMANITI/6315044).

References

- Barlow, M. (2004). *Collocate (Version 1.0)*. [computer software]. Available from <http://www.athel.com/colloc.html>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Lexical expressions in speech and writing. In D. Biber, & G. Leech (Eds.), *Longman Grammar of Spoken and Written English* (pp. 988-1036). Longman.
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics, 37*(6), 849-880. <https://doi.org/10.1093/applin/amu065>
- Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21-42. <https://doi.org/10.1016/j.jeap.2017.09.002>
- Green, C., & Lambert, J. (2018). Advancing disciplinary literacy through English for academic purposes: Discipline-specific wordlists, collocations and word families for eight secondary subjects. *Journal of English for Academic Purposes, 35*, 105-115. <https://doi.org/10.1016/j.jeap.2018.07.004>
- Hyland, K. (2002). Specificity revisited: how far should we go now? *English for Specific Purposes, 21*, 385-395. [https://doi.org/10.1016/S0889-4906\(01\)00028-X](https://doi.org/10.1016/S0889-4906(01)00028-X)
- Hyland, K. (2006). *English for Academic Purposes: An Advanced Resource Book*. New York: Routledge.
- Hyland, K., & Jiang, K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics, 23*(4), 383-407. <https://doi.org/10.1075/ijcl.17080.hyl>

- Jeong, H., & Jiang, N. (2019). Representation and Processing of Lexical Bundles: Evidence from Word Monitoring. *System*, 80, 188-198. <https://doi.org/10.1016/j.system.2018.11.009>
- Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, 7(1), 132-140. <https://doi.org/10.17509/ijal.v7i1.6866>
- Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36. <https://doi.org/10.1016/j.jeap.2019.03.008>
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, 40, 1-14. <https://doi.org/10.1016/j.jeap.2019.04.004>
- Shin, Y. K., Cortes, V., & Yoo, I. W. H. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, 39, 29-41. <https://doi.org/10.1016/j.jslw.2017.09.004>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List (AFL). *Applied Linguistics*, 31, 487-512. <https://doi.org/10.1093/applin/amp058>
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1-14. <https://doi.org/10.1016/j.esp.2018.09.001>
- Zamel, V. (1993). Questioning academic discourse. *College ESL*, 3, 28-39.