HiLL PUBLISHING

The Educational Review, USA, 2022, 6(10), 566-572
https://www.hillpublisher.com/journals/er/
ISSN Online: 2575-7946
ISSN Print: 2575-7938

A Comparative Study of Four-Word Lexical Chunks in English Linguistic Thesis by Chinese and English Writers

Jinjin Zhang

Independent researcher, Kuiwen neighborhood, Tengzhou, Shandong Province, 277500, China.

How to cite this paper: Jinjin Zhang. (2022). A Comparative Study of Four-Word Lexical Chunks in English Linguistic Thesis by Chinese and English Writers. *The Educational Review, USA*, 6(10), 566-572.
DOI: 10.26855/er.2022.10.010

Received: September 23, 2022
Accepted: October 19, 2022
Published: November 14, 2022

Corresponding author: Jinjin Zhang, Independent researcher, Kuiwen neighborhood, Tengzhou, Shandong Province, 277500, China.

Abstract

The lexical chunk, a combination of two or more continuous or incontiguous words, is a language phenomenon that can express a certain meaning and have a complete structure. Meanwhile, the lexical chunk is an essential ingredient of academic articles and an important indicator of language ability in academic articles. This study adopts English papers written by English writers and master theses written by Chinese postgraduates as the object. Using Biber et al's structural classification of lexical chunks as the frame, this study analyzes the characteristics of four-word lexical chunks at the structural levels. It was found that (1) Chinese postgraduates in English Linguistics used four-word lexical chunks more frequently than English writers. Moreover, Chinese postgraduates in English Linguistics tended to overuse four-word lexical chunks in writing. (2) At the structural level, Chinese postgraduates in English Linguistics preferred prepositional four-word lexical chunks and verbal four-word lexical chunks. English writers in English Linguistics were inclined to use prepositional four-word lexical chunks and noun four-word lexical chunks. Significant differences were found in the use of passive verbal four-word lexical chunks between master theses written by Chinese postgraduates in English Linguistics and journal papers written by English writers in English Linguistics. This study facilitates students' understanding and ability in lexical chunk use, meanwhile, this study provides available data and practical implications for lexical chunk learning and teaching.

Keywords

Four-word lexical chunks, corpus, MA theses, English Linguistics, English teaching, comparative study

1. Introduction

This chapter introduces the background and objectives of the study.

1.1 Background of the Study

In the fields of second language acquisition or corpus linguistics, lexical chunks are the focal points. It is well-acknowledged that lexical chunk contributes significantly to language comprehension and expression. Hyland argued that lexical chunks are not only central to the creation of academic discourse, but that they offer an important means of differentiating written texts by discipline (Hyland, 2008). Wary also indicates that second language learners need to get familiar with some common collocations and phrases in the target language before they truly acquire that

language (Wary, 2002). Furthermore, numerous researchers have proven the great importance of applying lexical chunks in practical teaching. Lexical chunks have attracted much attention, and they have been testified to be an appropriate way to promote foreign language learning. Therefore, lexical chunk competence has been regarded as one of the most significant indicators on language learners' language proficiency and a standard that tells an expert from a green hand.

1.2 Objectives of the Study

This study aims to conduct a comparative analysis of four-word lexical chunk use in master theses written by Chinese postgraduates in English linguistics and papers from English linguistics journals written by native writers. The objectives of the study are to answer the following questions:

1) What characteristics of four-word lexical chunks that are extracted from master theses written by Chinese postgraduates in English Linguistics and academic papers written by English writers?

2) Are there any significant differences between four-word lexical chunks extracted from English Linguistic master theses written by Chinese postgraduates and four-word lexical chunks extracted from English Linguistic journal papers written by English writers at the structural level?

3) Are there any significant functional differences between four-word lexical chunks extracted from English Linguistics master theses written by Chinese postgraduates and four-word lexical chunks extracted from English Linguistic journal papers written by English writers at the functional level?

2. Literature Review

This chapter mainly introduces related studies on lexical chunks.

2.1 Studies on Lexical Chunk Abroad and at Home

Lexical chunks refer to numerous word entities from different fields in this study. Various linguists gave dissimilar definitions to lexical chunks from different linguistic perspectives. Firth is the first person to put forward this language phenomenon, and he calls it 'extended collections'. Syder calls it 'form-meaning pairings', and 'lexicalized sentence stems'. Nattinger and DeCarrico define word collections as 'lexical phrases' (Nattinger & DeCarrico, 2000). Some Chinese linguists call it 'lexical chunks', 'lexical bundles', etc. Based on long-term studies and investigations of Corpus Linguistics, Sinclair lodged 'idiomatic principles', which claims the lexical chunk is the development and extension of a comprehensive collection (Sinclair, 1991). Different researchers may only investigate lexical chunks from a solar point of view; therefore, it is hard to give the lexical chunk a fixed definition. Biber et al. is the first person to investigate word collections. He nominates it as 'lexical bundles' in his grammar book (Biber et al., 1998). Altenberg also makes a contribution to the development of lexical chunks. He nominates continuous word collections as 'multi-word sequences', and discontinuous word collections are not included in 'multi-word sequences'. The 'multi-word sequence' is a comprehensive collection with different structures. 'Multi-word sequences' can be collections of sentences, phrases, and half-bakes phrases'. The language acquisition theory is put forward by Skehan. He claims that there are three necessary periods in language learning: lexicalization-grammaticalization-lexicalization (Skehan, 1999). Therefore, lexical chunks have the characteristics of expressing kinds of prefab meanings and are stored in unmodified forms. Mitchell believes that words are different from collocations, which belong to the grammatical category of higher abstractions.

Numerous Chinese linguists also make contributions to the development of lexical chunks. Yang, Y, C. analyzes the pragmatic functions and classifications of language 'plate' structure. He also claims that students should learn lexical chunks in vocabulary learning because lexical chunk learning can benefit students by avoiding some errors in language use, increasing students' confidence in written expressions, and enhancing fluency in language speaking and writing. Xiao, F, S. believes that this word type belongs to a kind of word collocation, a kind of word group, or lexical chunks. He puts forward a view that using a word type approach in English vocabulary teaching can increase students' learning motivation and boost the class involvement of students actively. Wu Jing and Wang, R, D. argue that lexical chunk teaching is suitable for Chinese students' characteristics in language learning, because the lexical chunk is an ideal model for memorizing vocabulary. Pu, J, Z. holds the view that language learners usually make mistakes in lexical chunks connecting and collocating. The lack of lexical chunks in language learners' brains causes weaknesses in lexical chunks connecting and collocating.

3. Theoretical Framework

This chapter mainly introduces classification of lexical chunks from the structural perspective. Reasons for choosing

four-word lexical chunks is also included in this chapter.

3.1 Structural Classification of Lexical Chunk

Biber et al. classified lexical chunks into 12 types, such as 'personal pronoun + verb phrase fragment'; 'it + verb phrases or adjective phrase fragments'; 'passive verbs + pronoun phrases fragments'. This kind of classification has been adopted for many years, and almost every study uses this method to analyze the structure of lexical chunks. Due to the lack of basic theory, researchers found it hard to classify and describe lexical chunks after being distilled from the corpus. Therefore, traditional structure classification has the foundation of language theory and is more convenient in manipulation. A detailed classification of lexical chunks from a structural perspective is shown in Table 3.1.

Table 3.1. Structural Classification of Lexical Chunk

Structure	Examples
Noun phrase with of fragment	The end of the, the nature of the
Noun phrase with other post-modifier fragment	The way in which, the relationship between the
Prepositional phrase with embedded of-phrase fragment	At a result of, in terms of the
Other prepositional phrase fragment	On the other hand, at the same time, by the fact that
Passive verb +prepositional phrase fragment	Is based on the, referred to as the
Anticipatory it + verb phrase/ adjective phrase	It is important to, it is possible to, it can be seen
Copula be + noun/ adjective phrase	Is the same as, is a matter of, is one of the, is due to the
(verb phrase) that- clause fragment	Should be noted that, that there is a, studies have shown that
(verb/ adjective) to- clause fragment	Are likely to be, has been shown to, to ensure that the
Adverbial clause fragment	As shown in figure, as we have seen
Pronoun/ noun phrase +be	This is not the, there was no significant
others	As well as the, may or may not, as shown in figure

3.2 Reasons for Choosing Four-word Lexical Chunk

Cortes points out that most four-word lexical chunks contain three-word lexical chunks, such as 'as a result of' including 'as a result' (Cortes, 2004). Furthermore, the frequency of four-word lexical chunks is higher than the frequency of five-word chunks. Therefore, four-word lexical chunks appear most frequently in writing than other lexical chunks. It is also easy to be extracted from the corpus and can show a more precise structure and function. Therefore, the four-word lexical chunk is the most suitable length for study.

4. Research Design

This chapter introduces methods and purposes of this study. The procedures of data collection and analysis are also expressed in this chapter.

4.1 Research Method

The study employs both quantitative and qualitative methods to conduct a comparative analysis of four-word lexical chunk use. Four-word lexical chunks are extracted from English Linguistic journal papers written by English writers and English Linguistics master theses written by Chinese postgraduates. This study uses four-word lexical chunks extracted from English Linguistic journal papers written by English writers as the standard. It analyzes the similarities and differences in lexical chunk use from frequency of use, functional and structure perspectives.

This study uses the grammatical structure of lexical chunks proposed by Biber et al. and the functional categories of discourse analysis presented by Hyland as the framework to explore the characteristics of four-word lexical chunk use. Four-word lexical chunks are extracted from English Linguistic journal papers written by English writers and master theses written by Chinese postgraduates in English Linguistics.

4.2 Data Collection

This study contains two samples: corpus 1 and corpus 2. Corpus 1 includes 30 papers that randomly is selected from 300 English linguistic journal papers. Corpus 2 consists of theses that randomly selected from master theses written by Chinese postgraduates in English Linguistics. This study selects papers in English linguistic journals indexed by SSCI with a great international influence (impact factor>2). These English Linguistic journals are: Applied Linguistics, Bilingualism-language and Cognition, International Journal of Bilingual Education and Bilingualism, Journal of Applied Linguistics, and Studies of Second Language Acquisition. 300 papers are selected from these five journals, and the publication years of the selected papers are ranged from 2016 to 2021. Furthermore, the first author of these papers must be an English-speaking speaker.

4.3 Research Tool

This study use AntConc3.5.8 to extract four-word lexical chunks. AntConc3.5.8 is a free monolingual corpus analysis and research tool, developed by Laurence Anthony in 2002. There are indexing tools such as “Concordance”, “Concordance Plot,” “File View,” “Clusters” or “N-Gram,” “Collocates,” “Word Lists”, “Keyword List” and other functions. The indexing tool “N-Gram” is adopted in this study to extract four-word lexical chunks in Corpus 1 and Corpus 2.

5. Results and Discussion

This chapter discusses characteristics of four-word lexical chunks from the structural perspective. Firstly, this chapter analyzes characteristics of four-word lexical chunk use between English Linguistic journal papers written by English writers and English Linguistics master theses written by Chinese postgraduates from a whole point of view. Then, similarities and differences of four-word lexical chunks use are discussed at the structural level.

5.1 The Overall Distribution of Lexical Bundles

With the help of the AntConc3.5.8, four-word lexical chunks in English Linguistic journal papers written by English writers and master theses written by Chinese postgraduates in English Linguistics are extracted. The data shows that Chinese postgraduates in English Linguistics use 40 four-word lexical chunks with the frequency of 542; moreover, English writers in English Linguistics use 40 four-word lexical chunks with the frequency of 264. The result shows that a certain number of four-word lexical chunks have been used in these two corpora. However, there still exist differences between the total number and frequencies of four-word lexical chunks used in these two corpora besides, it can be found that both the numbers and frequencies of four-word lexical chunks in corpus 1 are higher than that in corpus 2. This study proves that Chinese students tend to overuse lexical chunks in writing, which corresponds to previous studies of lexical chunks.

There is a difference in the frequency of four-word lexical chunks use between English Linguistic journal papers written by English writers and master theses written by Chinese postgraduates in English Linguistics. However, the main points of this study are to find out whether there are significant differences of four-word lexical chunk use between English Linguistic journal papers written by English writers and master theses written by Chinese postgraduates in English Linguistics from functional, structural perspectives. This thesis continues to find out the characteristics of lexical chunks. This thesis sorted out four-word lexical chunks used by both Chinese postgraduates in English Linguistics and English writers in English linguistics, and the result is shown in Table 5.1.With the help of SPSS, the result of significant differences is shown in Table 5.2.

The result shows the four-word lexical chunks that are used by both Chinese postgraduates and English writers in English linguistics. 13 out of 40 are both used by Chinese postgraduates and English writers in English linguistics. However, the result also indicates that less than 50% of the four-word lexical chunks are used by both Chinese postgraduates and English writers in English Linguistics. Both Chinese postgraduates and English writers prefer to use different four-word lexical chunks in their writings.

Table 5.1 also presents that some four-word lexical chunks are used by both Chinese postgraduates in English Linguistics and English writers in English Linguistics. Although these four-word lexical chunks are used by both Chinese postgraduates in English Linguistics and English writers, these four-word lexical chunks have different frequencies in two corpora. The frequency of most four-word lexical chunks in corpus 1 are higher than that in corpus 2. In addition, there still have some four-word lexical chunks appear more frequently in corpus 2 than that in corpus 1, which means some four-word lexical chunks are frequently used by Chinese postgraduates in English Linguistics, which are not frequently used by English writers in English Linguistics.

Table 5.1. Four-word Lexical Chunks Use

		CP	EW
Total		208	108
1	at the same time	39	7
2	as well as the	36	19
3	on the other hand	25	13
4	on the basis of	16	7
5	in the field of	13	9
6	can be used to	12	10
7	is one of the	12	4
8	in the present study	11	4
9	in terms of the	10	5
10	as a result of	9	9
11	can be seen in	9	10
12	on the one hand	9	5
13	at the end of	7	6

(Notes: CP= Chinese postgraduates EW= English writers)

Table 5.2. Chi-square Test on Overall Distribution of Four-Word Lexical Chunks

	CP	EW	value	df	sig.
number	48	43	7.635	1	.006
frequency	542	264	7.635	1	.006

(Notes: CP= Chinese postgraduates EW= English writers)
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 31.14
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 31.14

Table 5.2 indicates chi-square test of four-word lexical chunks used by Chinese postgraduates in English Linguistics (χ^2=7.635, df=1, p>.05), and English writers in English Linguistics (χ^2=7.635, df=1, p>.05). The result shows that there is a significant difference in the frequency of use between English Linguistic journal papers written by English writers and master theses written by Chinese postgraduates in English Linguistics. It also shows that Chinese postgraduates in English Linguistics use more various four-word lexical chunks than English writers in English Linguistics. Besides, Chinese postgraduates who major in English Linguistics use four-word lexical chunks more frequently in writing. Less than half of four-word lexical chunks are used by both Chinese postgraduates in English Linguistics and English writers in English Linguistics.

This result expresses that those Chinese postgraduates in English Linguistics have realized the importance of lexical chunks in writing. They also learned that using four-word lexical chunks is a necessary condition for a master thesis. Language learners have the awareness of using various kinds of four-word lexical chunks in their master theses.

5.2 The Structure of Four-word Lexical Chunks

Chapter three introduced the structural classification of lexical chunks invented by Biber et al. He classified lexical chunks into 12 types according to their structural differences. His structural classification of lexical chunks is adopted in this study, and the sorted four-word lexical chunks are shown in Table 5.3. To express the differences more clearly, this study also conducted a chi-Square test on the distribution of four-word lexical chunks in structure, which is shown in Table 5.4.

(Notes: n + of=Noun phrase with of fragment; n + other= noun phrase with other post-modifier fragment; p + of= prepositional phrase with embedded of-phrase fragment; p + other= other prepositional phrase fragment; p. v.+ p= passive verb + prepositional phrase fragment; it + v./ adj.= anticipatory it + verb phrase /adjective phrase; be + n./adj/=copula be / noun / adjective phrase; v + that-clause=verb phrase that - clause fragment; v./adj. + to-clause=verb/adjective to-clause fragment; adv.- clause= adverbial clause fragment; p/ n + be= pronoun/ noun phrase + be; Notes: N= number F= frequency CP= Chinese postgraduates EW= English writers)

Table 5.3. Distribution of Four-Word Lexical Chunks in the 12 Structural Types in Structure

		N	CP's percentage	F	N	EW's percentage	F
	Total/types	48		542	43		264
1	n + of	2	3%	16	2	4.50%	12
2	n + other	-	-	-	2	4.50%	12
3	p + of	17	36.90%	200	15	32%	85
4	P + other	8	21%	114	6	14.70%	39
5	p. v.+ p	6	9%	49	8	19.60%	52
6	it + v./ adj.	3	4.20%	23	1	2.60%	7
7	be + n./adj/	1	2.20%	12	2	3%	8
8	v + that-clause	2	4.60%	25	2	4.50%	12
9	v./adj.+ to-clause	4	7.30%	40	-	-	-
10	adv.- clause	1	1.60%	9	-	-	-
11	p/ n + be	-	-	-	-	-	-
12	other expressions	4	10%	54	5	14%	37

From Table 5.3, it can be found that the structural distributions of these four-word lexical chunks and the numbers of four-word lexical chunks in each structural category. Besides, this table shows that these four-word lexical chunks in corpus 1 cover 10 structural types among 12 structural types, and four-word lexical chunks in corpus 2 cover 9 types among 12 structural types. The result implies that Chinese postgraduates adopted a comparatively wide scope of structural types of four-word lexical chunks.

It has no difficulty to conclude that, structurally, Both Chinese postgraduates in English Linguistics and English writers are inclined to use a large number of four-word lexical chunks in their writings. Prepositional four-word lexical chunks have the same proportion between English Linguistics master theses written by Chinese postgraduates and English Linguistic journal papers written by English writers. Additionally, prepositional four-word lexical chunks play an essential role in academic papers. Additionally, English writers in English Linguistics prefer noun four-word lexical chunks in their writing, which accounts for the percentage of 23%. This percentage of noun four-word lexical chunks in native writer's journal papers is larger than its percentage in Chinese postgraduates' master theses. Noun four-word lexical chunks used by Chinese postgraduates who major in English Linguistics show a percentage of 13%. Studies found that verbal lexical chunks frequently appear in conversations, while noun lexical chunks usually appear in formal writings. The result manifests that English writers in English Linguistics tend to use formal four-word lexical chunks in writing. In contrast, Chinese postgraduates who major in English Linguistics use more verbal four-word lexical chunks than English writers in English Linguistics. This phenomenon indicates that four-word lexical chunks extracted from master theses written by Chinese postgraduates are more colloquial than these four-word lexical chunks extracted from English Linguistic journal papers written by English writers.

As shown in Table 5.4, significant differences fail to be found in the use of noun four-word lexical chunks (χ^2=1.344, df=1, p>.05), prepositional four-word lexical chunks (χ^2=4.524, df=1, p>.05), verbal four-word lexical chunks (χ^2=1.256, df=1, p>.05), adjective four-word lexical chunks (χ^2=.489, df=1, p>.05), and clausal four-word lexical chunks (χ^2=.002, df=1, p>.05). However, a significant difference exists in the use of passive verbal four-word lexical chunks (χ^2=18.393, df=1, p<.05). Thus, most structural four-word lexical chunks extracted from master theses written by Chinese postgraduates in English Linguistics are not significantly different from those extracted from English Linguistic journal papers written by English writers. Moreover, passive verbal four-word lexical chunks used by Chinese postgraduates in English Linguistics are significantly different from those used by English writers in English Linguistics. Chinese postgraduates in English Linguistics are inclined to use active four-word lexical chunks instead of passive verbal four-word lexical chunks. On the contrary, English writers have a preference in using passive verbal four-word lexical chunks. Due to this phenomenon, Chinese postgraduates should internalize more passive verbal four-word lexical chunks in language learning and use them in their writing.

Table 5.4. Chi-Square Test on Distribution of Four-Word Lexical Chunks in Structure

	CP	EW	value	df	sig
n + of	16	12	1.344	1	0.246
p + other	114	39	4.524	1	0.042
p. v.+ p	49	52	18.393	1	0
it + v./ adj.	23	7	1.256	1	0.262
be + n./adj/	12	8	0.489	1	0.484
v + that- clause	25	12	0.002	1	0.966

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 0.917
b. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 50.11
c. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 33.08
d. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.83
e. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.55
f. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.12
(Notes: CP= Chinese postgraduates EW= English writers)

To sum up, prepositional four-word lexical chunks appear most frequently in both master theses written by Chinese postgraduates in English Linguistics and English Linguistic journal papers written by English writers. English writers in English Linguistics are not inclined to use prepositional and noun four-word lexical chunks in their writings. While Chinese postgraduates in English Linguistics use verbal lexical chunks more frequently than those used by English writers. These data indicate that both Chinese postgraduates and English writers prefer prepositional four-word lexical chunks in their writings and have slight differences in other four-word lexical chunks' usage. Verbal four-word lexical chunks displays a higher frequency and larger percentage in Chinese postgraduates' master theses. English writers in English Linguistics have a preference for employing noun four-word lexical chunks to verbal four-word lexical chunks in their writing.

6. Conclusion

6.1 Findings of the Present study

In this study, Chinese postgraduates in English Linguistics use various four-word lexical chunks in their academic writings. Furthermore, four-word lexical chunks extracted from master theses written by Chinese postgraduates in English Linguistics have a higher frequency than four-word lexical chunks extracted from English Linguistic journal papers written by English writers. Moreover, 13 out of 40 four-word lexical chunks are used both by Chinese postgraduates in English Linguistics and English writers in English Linguistics, which means less than half of the lexical chunks are used by both Chinese postgraduates in English Linguistics and English writers in English Linguistics. Meanwhile, research data indicates that there is a significant difference in four-word lexical chunk use between master theses written by Chinese postgraduates in English Linguistics and English Linguistic journal papers written by English writers.

References

Biber, D., & S, Concord. (1998). Corpus Linguistics: Investigating Language Structure and Use. *Cambridge, Cambridge University Press.*

Cortes, V. (2004). Teaching lexical bundles in the disciplines: An example from a writing intensive history class [J]. *Linguistic and Education,* 17, 391-406.

Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing [J]. *International Journal of Applied Linguistics,* 18(1), 41-62.

Nattinger, J. & DeCarrico, J. (2000). Lexical Phases and Language Teaching. *Shanghai, Shanghai Foreign Language Teaching Press.*

Sinclair, J. (1991). Corpus, Concordance, Collocation. *Oxford: Oxford University Press.*

Skehan, P. A. (1999). Cognitive Approach to Language Learning. *Shanghai: Shanghai Foreign Language Teaching Press.*

Wray, A. (2002). Formulaic Language and the Lexicon. *Cambridge: CUP.*