INTERNATIONAL JOURNAL OF LANGUAGE EDUCATION AND APPLIED LINGUISTICS
ISSN: 2289-7208 E-ISSN: 2289-9294
VOLUME 14, ISSUE 2, SEPTEMBER 2024, pp. 20 – 30
DOI: https://doi.org/10.15282/ijleal.v14i2.10897

RESEARCH ARTICLE

A Comparative Analysis of Four-word Lexical Bundles Used by Postgraduate Students in China and America

Min Chen[1,2], Roslina Abdul Aziz[3*] and Syamimi Turiman[1]

[1]Akademi Pengajian Bahasa, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia.
[2]School of Social Management, Jiangxi College of Applied Technology, No.9 Wenfeng Road, Ganzhou, Jiangxi Province, 314000,China.
[3]Akademi Pengajian Bahasa, Universiti Teknologi MARA Cawangan Pahang, Kampus Jengka, 26400 Bandar Jengka, Pahang, Malaysia.

ABSTRACT - The ability to write is widely acknowledged as a crucial skill in higher education. As a key component of fluent linguistic production, lexical bundles (LBs) are an important distinguishing feature in academic writing. However, many postgraduate students face challenges using lexical bundles in academic writing. This study focuses on the similarities and differences of various forms, structural, and functional patterns of four-word lexical bundle usage between Chinese postgraduate EFL learners and American native postgraduate students in Applied Linguistics. It adopts a corpus-based methodology based on two self-built learner corpora and uses quantitative and qualitative methods to analyse the data. The results showed that Chinese postgraduate EFL learners rely more on four-word LBs in constructing academic writing. In terms of structural types, Chinese postgraduate EFL learners tend to have a balance when using the three different types of bundles. In contrast, native American postgraduate students tend to use more prepositional-based bundles. As for the functional types, Chinese postgraduate EFL learners are more inclined to use research-oriented texts to provide descriptions to organise the writer's actions, whilst native American postgraduate students are more inclined to use text-oriented bundles to organise the text. The findings also offer implications for improving teaching lexical bundles in academic writing curricula in China.

ARTICLE HISTORY
Received: 20 June 2024
Revised: 13 September 2024
Accepted: 27 September 2024
Published: 2 December 2024

KEYWORDS
Academic writing
Chinese EFL learners
Comparative analysis
Lexical bundles
Native speaker learners

1.0 INTRODUCTION

Academic writing is broadly recognised as an essential competency in higher education. Many students in universities or colleges require this necessary skill for their academic studies. It is widely accepted that academic writing is one of the most complex and challenging tasks for English as a Foreign Language (EFL) learners, as they must be able to articulate their thoughts and opinions and organise their writing logically (Aldabbus & Almansouri, 2022). According to Day (2023), students must write their dissertations, essays, and examinations critically, formally, and precisely. Therefore, a high level of expertise in scholarly writing is essential for EFL students in achieving academic success.

To develop competency in communication and attain dominance in a specific field, the students must master certain word sequences used among expert users of a given language. Lexical bundle (LB) is a term used to refer to word sequences characterised as sequences that occur most frequently in writing. These word sequences are also referred to as clusters in some dictionaries. Wray (2000) pointed out that formulaic language is a prefabricated sequence, which is stored and retrieved whole from memory at the time of use. Many studies have highlighted the crucial importance of formulaic language in enhancing language proficiency (Yu, 2022). Proficient speakers make use of great numbers of fixed expressions that go far beyond classic idioms. Adequate application of LBs can enhance language users' understanding and fluency in output (Nasrabady et al., 2020). A significant correlation has been established between high-level language proficiency and the proficient employment of lexical bundles (Kim & Kessler, 2022).

Postgraduate EFL learners often find themselves facing challenges in academic writing. They frequently need help with their academic vocabulary when it comes to academic writing, particularly in using appropriate lexical expressions coherently and effectively utilising academic language (Cui & Kim, 2021; Jiang & Hyland, 2020). Zhang et al. (2021) pointed out that Chinese novice master students with little experience in academic writing face difficulty in proficient use of disciplinary language and have poor language logic. Pan (2024) reported that Chinese students generally feel that academic English writing is complex and challenging to improve. The study found that the students cannot use academic sentences to convey information accurately, and English expressions could be more rigorous and standardised. The studies provide empirical evidence of the struggle faced by Chinese EFL learners with academic writing due to limited lexical precision, abundance, and accurate word choices.

This study aims to analyse the LB in MA theses written by Chinese EFL learners and compare them to those used by native English speakers, specifically American learners, to determine the severity of these vocabulary challenges. The rationale for comparing with native learner writings is to identify specific areas where Chinese EFL learners struggle and highlight the differences in lexical

*CORRESPONDING AUTHOR | Roslina Abdul Aziz | ✉ leenaziz@uitm.edu.my

bundle usage. The investigation can provide insights into non-native speakers' unique difficulties and inform targeted teaching strategies to improve their academic writing skills. Specifically, this study aims to find the similarities and differences between using LBs in forms of structural and functional patterns in the Chinese MA theses and compare them to those in the MA theses of American learners.

Following Hyland's (2008b) classification of LBs in structure and function, this study aims to focus on various uses of four-word LBs between postgraduate EFL learners and native American postgraduate students in the field of Applied Linguistics, based on two self-built learner corpora, namely the Chinese postgraduate EFL learners' MA theses (PEL) corpus and the native American postgraduate students' MA theses (NPS) corpus. The research questions of the study are as follows:

1. What are the most frequent four-word lexical bundles (LBs) in the MA theses of Chinese EFL and native American postgraduates?
2. What are the similarities and differences in the structures of the four-word LBs used by Chinese EFL and native American postgraduates in their MA theses?
3. What are the similarities and differences between the four-word LB functions used by Chinese EFL and native American postgraduates in their MA theses?

2.0 METHODOLOGY

2.1 Research Design

This research adopts a corpus-based methodology that uses quantitative and qualitative methods to identify and analyse the frequency, structures and functions, as well as compare the characteristics of LBs in the Master's (MA) theses of Chinese EFL learners and native American students' MA theses. The study focuses only on MA theses as they follow a specific academic genre, making comparing easier. This consistency is crucial when analysing language structures and functions, as it minimises variability due to differences in writing purposes or styles. Secondly, MA theses often reflect the early stages of students' research and academic writing skills. This focus can be beneficial for understanding the development of these skills.

The quantitative analysis involved quantifying the frequency and percentage of the different types of LBs and the utilisation patterns of LBs in EFL learners' MA theses. Subsequently, these were compared with the outcomes of the LBs derived from the NPS corpus. To discern the differences in the use of LBs between the two corpora, a chi-square test was conducted using Statistical Package for the Social Sciences (SPSS) (Version 29). The qualitative analysis involved conducting concordance analysis to ascertain the unique characteristics of LBs utilised across the corpora while examining how they are used in context.

2.2 Corpora

2.2.1 Postgraduate EFL Learner Corpus

The non-native learner corpus built for this study contains 25 theses written by Chinese MA postgraduates who majored in Applied Linguistics. The rationale for choosing Applied Linguistics theses is that in China, only students in Applied Linguistics programs who major in English must write their theses in English. In contrast, students in other fields, such as engineering or medicine, typically write in their native language. This focus ensures the study is relevant and accurate, as it exclusively examines the written English proficiency and use of lexical bundles by learners who are mandated to write in English.

The data were collected from China National Knowledge Infrastructure (CNKI), a database which collected most of China's knowledge resources. This study extracted 25 MA theses written by Chinese postgraduate students from 8 different universities from different cities in China from 2022 to 2024 to ensure that the data are representative. In maintaining consistent structural integrity, the MA theses adhere to the broad MA structure: Introduction, Literature Review, Methodology, Results and Discussion. The total number of words in the corpus is 509,905; an average size of 20,396 words per dissertation is presented in Table 1.

Table 1

The details of the corpus of postgraduate EFL learners (PEL)

Corpus	Total Size	Average Size	Number of Texts
PEL	509,905	20,396	25

2.2.2 Native American Postgraduate Students Corpus

The native MA postgraduate students (NPS) corpus contains M.A. theses retrieved from the ProQuest database. Based on the standards proposed by Wood et al. (2001), the theses were authored by individuals with English and surnames affiliated with institutions or universities in countries where English was their first language and were selected for this corpus. To standardise the sizes of the two corpora, 33 theses from Applied Linguistics were chosen from 18 universities of different provinces according to geographical location, spanning the years 2022 to 2024, in building the NPS. The MA theses also adhere to the broad MA structure: Introduction, Literature Review, Methodology, Results and Discussion (Table 2).

Table 2

The details of the corpus of native American Postgraduate Students (NPS)

Corpus	Total Size	Average Size	Number of Texts
NPS	500,350	15,162	33

2.3 Computational Tool

Considering the demands of this research and its free software license, cross-platform application, straightforward manual, and ability to analyse extensive data, AntConc 4.2.4 (Anthony, 2023) was employed to analyse the data in this study. The N-gram function was used to generate the four-word LBs. Then, the LB list can be utilised to analyse the concordance lines of each lexical bundle. File view was used to uncover the file where the LBs took place.

2.4 Data Analysis

2.4.1 Criteria for Lexical Bundles Identification and Selection

Some identifying criteria have been set to ensure that the extracted LBs have a certain representativeness and that the number is reasonable and controllable. Hyland contends that four-word bundles offer more distinct structures and functions than three-word bundles, which are considerably more prevalent than five-word strings (Hyland, 2008b). Deng and Liu (2023) pointed out that four-word bundles play a vital role in realising the communicative purpose of research articles. Hence, opting for four-word bundles is the most appropriate decision when considering the size of lexical bundles.

The normalised frequency is usually around 20 to 40 occurrences per million words (Biber et al., 2004). Nasrabady et al. (2020) adopted 50, 30, and 15 for different lengths of lexical bundles, while Guan (2022) set out a higher criterion in his study (40 times per million words). This study adopted 30 per million words as a criterion in both corpora, considering the standard in most recent LBs research (Pan et al., 2016).

Dispersion is the third key factor in identifying LBs, as it is crucial to protect against the idiosyncratic uses by individual speakers or authors (Biber et al., 2004). This study adopted Hyland's (2008a, 2008b) criteria of 10 per cent of all corpus texts, and a dynamic dispersion criterion was established in three texts for PEL and four texts for NPS.

2.4.2 Categories of Lexical Bundles

Hyland's (2008a, 2008b) structural classification was employed to direct the structural examination of both corpora. It involves three categories and ten sub-categories, as shown in Table 3.

Table 3

Structural classification of LBs by Hyland (2008b)

Structural Types	Subtypes	Examples
Verb-based	a. Anticipatory it + verb/adjective phrase	*it is necessary to*
	b. Copula be + noun/adjective phrase	*is in line with*
	c. Pronoun/NP + be	*that is to say*
	d. First-person pronoun + dependent clause	*we can see that*
	e. (verb/adjective +) to-clause	*can be used to*
Noun-based	a. NP with of-phrase	*the content of the*
	b. NP with other post-modifier/ Other noun phrases	*an important role in*
Preposition-based	a. Prepositional phrase with embedded of-phrase	*in the process of*
	b. Other prepositional phrase expressions	*at the same time*

The functional taxonomy proposed by Hyland (2008a) was used to guide the functional analysis of all the data in both corpora. It puts forward three key classifications and eleven sub-sections for the retrieved bundles, as summarised in Table 4.

Table 4

Functional classification of LB distribution (Hyland, 2008a)

Functional types	Subtypes	Examples
Research-oriented	Location	*the end of the*
	Procedure	*in the process of*
	Quantification	*a large number of*
	Description	*the similarities and differences*
	Topic	
Text-oriented	Transition signals	*on the other hand*
	Resultative signals	*the results of the*
	Structuring signals	*in the current study*
	Framing signals	*on the basis of*
Participant-oriented	Stance features	*it is clear that*
	Engagement features	*can be seen in*

2.5 Inter-coder Reliability

The concept of reliability concerns whether the study's outcome can be consistently reproduced when conducted in other settings or contexts (Bryman, 2016). Considering the coding of the data, enhancing reliability could be achieved by conducting cross-examinations with other raters or coders to address any inconsistencies. Two coders (an Associate Professor and a senior English lecturer, both of whom have had more than ten years of experience in teaching English at the tertiary level) were selected to code the LBs according to the structural and functional classifications used in this study (refer to Table 3 and Table 4). This procedure was undertaken to reduce subjectivity and improve reliability. Subsequently, the inter-coder reliability between coders for each group was evaluated using Cohen's Kappa in SPSS. After two rounds of calculation, the result is between 0.8-1.00, which signifies a near-ideal level of dependability in structural and functional categorisation in PEL and NPS, based on the level of Kappa statistics (Landis & Koch, 1977).

3.0 RESULTS AND DISCUSSION

3.1 Lexical Bundles Frequency of Chinese EFL Postgraduates and Native American Postgraduates

According to the criteria set out, Antconc 4.2.4 (Anthony, 2023) was used to extract all the four-word LBs. After the refinement process, including removing specific discipline bundles, context-dependent bundles, overlapping bundles, and problematic bundles, the number of LB types was 110 LBs in PEL and 54 in NPS. The total extracted four-word LBs used in PEL is 110, more than that of NPS, which recorded 54 LB types. The figures indicate that Chinese postgraduate EFL learners have covered many LBs. The result is in line with the findings of Guan (2022). His research revealed that L1-Chinese students used significantly more types and tokens of lexical bundles than L1-English students. The result is also consistent with the finding of Yakut et al. (2021), which suggests that non-native speakers (native Turkish speakers) use LBs more frequently in their theses than native English students.

The finding also indicates that Chinese MA students rely more on LBs and use some more repeatedly to complete specific communicative functions. These LBs allow them to express their ideas and thoughts in academic writing. In addition, for some LBs, the frequency of Chinese MA theses is very high, such as the first ranking bundle with the highest frequency recorded 350 occurrences, almost five times that of the first ranking bundle in the NPS corpus. Some Chinese learners excessively reuse or overuse specific word bundles, leading to the high frequency of four-word lexical bundles in their MA theses. The reason may also be that Chinese students overuse some "safe" word bundles, especially the fixed collocations similar to Chinese expressions, such as *from the perspective of*, *in the process of* and *at the same time*.

3.2 The Most Frequently Used Lexical Bundles by EFL Postgraduates and Native American Postgraduates

In Table 5, the frequency of the top 20 most frequent four-word LBs ranged between 45-350/100,000 in PEL, while the frequency is much less in NPS, with the range of 24-104/100,000 frequencies. The most frequent four-word bundle in PEL was *from the perspective of*, which accounted for 350/100,000, followed by the bundle *in the process of* and *at the same time*, which accounted for 123/100,000 and 120/100,000, respectively. There were three bundles which occurred more than 100 per million words and 17 bundles that occurred more than 50 per million words in PEL. The most frequent four-word bundle in NPS was *it is important to*, which accounted for 71/100,000, less than 100 per million words. There was not a single bundle that occurred more than 100 per million words.

Table 5

The 20 most frequent four-word bundles in PEL and NPS

Chinese EFL Postgraduates (PEL)			Native American Postgraduates (NPS)		
Rank	**Bundles**	**Freq./mil**	**Rank**	**Bundles**	**Freq./mil**
1	*from the perspective of*	350	1	*it is important to*	71
2	*in the process of*	123	2	*in the case of*	63
3	*at the same time*	120	3	*in the field of*	54
4	*it can be seen*	87	4	*on the other hand*	51
5	*on the other hand*	85	5	*as well as the*	50
6	*the similarities and differences*	76	6	*in the context of*	49
7	*in the use of*	68	7	*the end of the*	42
8	*can be divided into*	65	8	*participants in this study*	41
9	*can be seen that*	65	9	*when it comes to*	41
10	*on the basis of*	65	10	*the results of the*	36
11	*as well as the*	33	11	*the beginning of the*	35
12	*on the one hand*	57	12	*the ways in which*	32
13	*as is shown in*	54	13	*in the current study*	31
14	*in the field of*	51	14	*as a means to*	28
15	*with the help of*	51	15	*the extent to which*	28
16	*it is necessary to*	50	16	*the rest of the*	28
17	*the content of the*	50	17	*a wide range of*	25
18	*is one of the*	48	18	*in the form of*	25
19	*a large number of*	46	19	*in the present study*	25
20	*the results of the*	45	20	*one of the most*	25

3.3 Shared Lexical Bundles among the Chinese EFL Postgraduates and Native American Postgraduates

The findings for shared LBs reveal that both corpora share 18 bundles, as presented in Table 7**.** The differences in frequencies among these bundles are also shown in Table 6. The frequencies of 7 LBs are higher in PEL than in NPS. The frequency *at the same time* is more than five times in PEL than in NPS, while *it is necessary to* be almost three times in PEL than in NPS. The result suggests that, to some extent, Chinese postgraduate EFL learners have recognised the importance of LBs and have tried to use more LBs to organise sentences and articles in their writing. However, the result also illustrates that Chinese MA postgraduates are more likely to use limited four-word LBs repeatedly and need more flexibility. In addition, the frequency of 11 LBs is lower in PEL than in NPS, among which some bundles are significantly underused among Chinese MA postgraduate EFL learners; for example, the bundles *in the context of*, and *in the case of* are two times less in PEL than in NPS, indicating lack variety in the use of particular LBs.

Table 6

List of shared LBs in PEL and NPS

Rank	Bundles	Structural Type	Frequency		Chi-square	P-value
			PEL	NPS		
1	*at the same time*	Preposition-based	120#	23	63.972	0.000**
2	*on the other hand*	Preposition-based	85#	51	7.869	0.005*
3	*as well as the*	Preposition-based	33	50	3.811	0.050
4	*in the field of*	Preposition-based	51	54	0.152	0.697
5	*it is necessary to*	Verb-based	50#	17	15.636	0.000**

Rank	Bundles	Structural Type	Frequency		Chi-square	P-value
			PEL	NPS		
6	*is one of the*	Verb-based	48#	22	9.172	0.002*
7	*the results of the*	Noun-based	45#	36	0.837	0.360
8	*the end of the*	Noun-based	32	42	1.547	0.213
9	*when it comes to*	Verb-based	26	41	3.648	0.055
10	*in addition to the*	Preposition-based	24#	23	0.007	0.935
11	*in the context of*	Preposition-based	22	49	10.785	0.001*
12	*in the form of*	Preposition-based	19	25	0.936	0.333
13	*that there is a*	Verb-based	18	23	0.708	0.400
14	*are more likely to*	Verb-based	17#	16	0.014	0.905
15	*as a result of*	Preposition-based	17	20	0.303	0.582
16	*in the case of*	Preposition-based	16	63	28.858	0.000**

Note. *=p＜0.05,**=p=0.000, #= higher frequency

The last column of Table 7 shows the results of chi-square tests to ascertain the significance of the frequency variance in each shared four-word LB across the two corpora. The table reveals that the P-values of six four-word LBs (*at the same time, on the other hand, it is necessary to, is one of the, in the context of, in the case of*) are below 0.05 (p<0.05), indicating significant disparities between these bundles in both corpora. Within the group of six four-word LBs, the P-values of 3 LBs (*at the same time, it is necessary to, in the case of*) are 000***, signifying their substantial difference.

3.4 Structural Characteristics of Lexical Bundles among EFL Postgraduates and Native American Postgraduates

Based on Hyland's (2008a) structural classification of LBs, the 110 four-word bundles in PEL and 54 four-word LBs in NPS are classified into three categories (verb-based bundles, noun-based bundles, and preposition-based bundles) and ten sub-categories respectively, with different percentage distributions as shown in Table 7.

Table 7

Structural categories distributions in PEL and NPS

Structure	PEL (Type)		NPS (Type)		Chi-squared p-value
	No.	%	No.	%	
Verb-based	**39**	**35.45%**	**13**	**24.07%**	0.0000
Anticipatory it + verb/adj. phrase	5	4.55%	5	9.26%	0.9760
Copula be + noun/adj. phrase	12	10.91%	2	3.70%	0.0050
Pronoun/NP + be	4	3.64%	2	3.70%	0.4230
First-person pronoun + dependent clause	1	0.91%	0	0.00%	0.2420
(verb/adj.) + to-clause	13	11.82%	4	7.41%	0.0280
Other VP expressions	4	3.64%	0	0.00%	0.0171
Noun-based	**32**	**29.09%**	**14**	**25.93%**	**0.0090**
NP + of-phrase	15	13.64%	9	16.67%	0.2360
NP + post-modifier/Other NP	17	15.45%	5	9.26%	0.0100
Preposition-based	**39**	**35.45%**	**27**	**50.00%**	**0.0160**
PP + of-phrase	20	18.18%	15	27.78%	0.4290
Other PP expressions	19	17.27%	12	22.22%	0.2260
Total	**110**	**100.00%**	**54**	**100.00%**	**0.000**

3.4.1 Verb-based Bundles

As shown in Table 8, verb-based bundles were used in PEL the most, which reached 39 bundles and accounted for 35.45% of the whole 110 four-word LBs compared to only 13 verb-based bundles in NPS, which take up a proportion of 24.07% of the entire 54 four-word LBs. The findings aligned with Zhang's (2021) study, which indicates that Chinese EFL writers use verb phrase-based bundles more frequently. According to the Chi-square test results, there is a significant difference in the use of verb-based bundles between the two corpora ($p=0.000<0.05$). There are also substantial differences between the three subcategories of copula be + noun/adj phrase ($p=0.005<0.05$), verb/adj + to-clause ($p=0.028<0.05$) and other verb phrase expressions ($p=0.017<0.05$).

The primary usage of anticipatory it + verb/adj phrase by Chinese MA postgraduate EFL learners is as follows: anticipatory it + adj. + infinitive-to/that clause (such as it is necessary to, it is clear that); anticipatory-it + be + verb in passive voice infinitive-to/that clause (such as it is found that). Two shared bundles of this subcategory are found in both corpora: *it is necessary to,* and *when it comes to.*

These two sub-categories bundles (copula be + noun/adj phrase and verb/adj + to-clause) are much more used by Chinese MA postgraduate EFL learners. As for copula be + noun/adj phrase, Chinese MA postgraduate EFL learners use 12 bundles of this sub-type, while native MA postgraduate students only use two bundles. Chinese postgraduate EFL learners use many passive phrases, such as can be seen that (65 times), shown in figure (54 times), is related to (21 times), is found (17 times), which is regarded as (17 times). The reason may be associated with the influence of Chinese culture, which is that people try to adopt a conciliatory and non-interventionist stance. The matter is in line with Bao's (2024) study. He pointed out that Chinese students tend to employ passive voice, which deflects focus from the researcher to the research outcomes, which aligned with the traditional, structured academic style frequently adopted by Chinese students, focusing on impartiality and reducing the author's prominence.

As for verb/adj + to-clause type LBs, Chinese learners used 13 bundles, while native American MA postgraduate students only used four bundles. Besides, they do not have a single shared bundle. The commonly used bundles by Chinese MA postgraduate EFL learners can be divided into the following: *used to describe the, can be used to, to a certain extent, and to find out the*.

3.4.2 Noun-based Bundles

According to the chi-square test results, there is a significant difference in using noun-based bundles between the two corpora ($p=0.00<0.05$). There are 32 noun-based bundles, accounting for 29.09%, which ranks third in PEL, while only 14 in NPS, accounting for 25.93%. Noun-based bundles rank second in both corpora. The sub-structural type of NP + of-phrase recorded 15 occurrences in PEL and 9 in NPS, with almost similar percentages of 13.64% and 16.67%, respectively. In the sub-structural type of NP + other post-modifier fragments, there are 17 bundles in PEL and five bundles in NPS, with a percentage of 14.66% and 9.26%, respectively. The structure of the NP + of-phrase fragment used by both students is the + X+of +X, such *as the content, the use of the, the application of the,* and *the nature of the*. However, there are only four noun-based bundle types (the results of the, the end of the, the beginning of the, and the use of the) shared by both students. The significant differences between them are in the core word Noun in the phrase, such as *nature, purpose*, and *ways*, which did not occur in the list of nouns + of phrase in PEL.

It is identified that two NP + post-modifier bundles are part of a relative clause in the PEL corpus, e.g., *the results show that* (19 times) and *the fact that the* (17 times). Two similar bundles are part of the relative clause in NPS, e.g., *the extent to which* (32 times) and *how* (28 times), with higher frequencies. However, no bundle occurred as part of the relative clause introduced by the relative pronoun "which" in PEL. It is evident that Chinese MA postgraduate EFL learners did not use these types of relative clauses as frequently as native experts did, which is in line with Guan's (2022) study; the researcher found that only L1 English writers used NP-based bundles as part of a relative clause, but not the Chinese writers.

3.4.3 Preposition-based Bundles

According to the chi-square test results, there is a significant difference in the use of preposition-based bundles between the two corpora ($p=0.016<0.05$). Among the three structural types, the number of preposition-based bundles reached 26, which accounts for exactly half of the total 54 LB types in NPS, while preposition-based bundles only account for 33.62% in PEL. This result is consistent with Lyu and Gee's (2020) finding that prepositional type phrases make up the highest percentage of students from the United States. Geluso (2022) stated that the *prepositional + of phrases* is the standard structure in academic writing.

As for the sub-categories, there were 20 *PP + of-phrase* bundles in PEL and 15 in NPS, accounting for 17.24% and 27.78% respectively. The structure in *the + noun + of prepositional phrases* occurred the most in PEL, with 9 out of 20 *prepositional + of phrases* structure, such as *in the process of*, *in the use of*, *in the field of*, *in the study of* and so on. It is similar in NPS, with six bundles in the whole 15 *prepositional + of phrases*, such *as in the context of*, *in terms of the*, *in the case of*, and *in the form of*. The number of LBs in the sub-type of other PP expressions is 19 in PEL and 12 in NPS, accounting for 16.38% and 22.22%, respectively. Under the subcategory other PP expressions, the structure *preposition + noun* (e.g., *in other words, the*, *in this way the*, *in this study the* in PEL and *as a tool for*, *in a way that*, *as a means to* an NPS) occurred most frequently in PEL and NPS. The preposition-based bundles function as research-oriented and text-oriented devices in both corpora. Using such LBs can realise the functions of structural organisation and information reorganisation, enhance the rigour and systematism of the overall logic of the content of academic papers, and make the text more readable.

Of the 16 shared LBs shown in Table 6, 9 bundles are preposition-based bundles, which accounted for 56.25%. The Chinese and American MA postgraduate students have more commonalities in using prepositions than other bundles. However, after close

examination, there are many differences between the use of preposition-based bundles and the use of preposition-based bundles. In the list of 39 preposition-based bundles in PEL, 30 did not occur in NPS, such as *from the perspective of*, and *in the process of*.

Meanwhile, almost 70% of preposition-based NPS bundles are unavailable in PEL. The finding suggests that Chinese and American students prefer preposition-based LBs. Specifically, they used preposition-based bundles with nouns, namely *present*, *current*, *fact*, *terms*, and *course,* significantly more frequently, while Chinese students used *perspective*, *process, use,* and *purpose* more regularly. This result suggests that the two groups used this PP-frame for different purposes. Chinese EFL learners used this structure to explain the research field or subject. In contrast, native-speaker students utilised it to define the research backdrop, background details, or pertinent factors (Lu & Deng, 2019).

3.5 Functional Characteristics of Lexical Bundles in Chinese EFL Postgraduates and Native American Postgraduates

Based on Hyland's functional classification of LBs, the 110 types of four-word bundles in PEL and 54 in NPS are classified into three categories (research-oriented bundles, text-oriented bundles, and participant-oriented bundles) and ten sub-categories, respectively, with different percentage distribution (Table 8). During the refinement process, the bundles related to the topic (specific discipline bundles, context-dependent bundles) were deleted, so there was no bundle belonging to this subcategory.

Table 8

Functional categories distributions in PEL and NPS

Function	PEL (Type)		NPS (Type)		Chi-squared p-value
	No.	%	No.	%	
Research-oriented	59	53.64%	21	38.89%	0.0000
Location	11	10.00%	5	9.26%	0.1930
Procedure	10	9.09%	4	7.41%	0.1110
Quantification	10	9.09%	9	16.67%	0.6660
Description	28	25.45%	3	5.56%	0.0000
Text-oriented	39	35.45%	25	46.30%	0.0930
Transition signals	8	7.27%	4	7.41%	0.2570
Resultative signals	6	5.45%	3	5.56%	0.1320
Structuring signals	13	11.82%	6	11.11%	0.1130
Framing signals	12	10.91%	12	22.22%	0.9630
Participant-oriented	12	10.91%	8	14.81%	0.3920
Stance features	4	3.64%	8	14.81%	0.2310
Engagement features	8	7.27%	0	0.00%	0.0010
Total	110	100.00%	54	100.00%	0

Moreover, in Table 8, in PEL, research-oriented bundles represent nearly 53.64% of total LBs, accounting for the most LB types. Text-oriented bundles take a proportion of more than one-third (35.45%), ranked second, followed by participant-oriented bundles, which accounted for only 10.91%. In NPS, text-oriented bundles account for 46.30% of the LB types, followed by research-oriented bundles, which take up slightly more than one-third, 38.89%, while participant-oriented bundles account for the least, which only account for 14.81%.

3.5.1 Research-oriented Bundles

It can be seen from Table 9 that the proportion of research-oriented bundles in PEL is much higher than that in NPS. It shows that Chinese MA postgraduate EFL learners are more inclined to use research-oriented texts to provide descriptions to organise the writer's actions. According to the chi-square test results, there are significant differences in using research-based bundles between the two corpora ($p=0.000<0.05$). The findings are aligned with the previous study. Hyland (2008a) found that Chinese MA students' dissertations are featured by heavily using research-oriented LBs. He pointed out that Chinese master students rely heavily on the physical practicalities of the investigation to conceptualise their research. They employ a more significant percentage of bundles focused on research, yet a lesser percentage of those centred on text than their American equivalents. Bao (2024) also pointed out that Chinese students employ a more significant percentage of research-oriented bundles yet a lesser percentage of text-oriented bundles than their American counterparts.

In the sub-type of 'procedure' bundles, there were 10 in PEL and 4 in NPS. Chinese learners tend to use more bundles related to the word *use*, such as *in the use of*, *can be used to*, *differences in the use*, *through the use of*, *used to analyse the*, while native American students only have one bundle related to the *use* (*the use of them*). The main difference between research-oriented bundles is the focus on the subcategory of 'description' bundles, with 28 in PEL and 3 in NPS. It takes up 25.45% of all LBs in PEL while accounting for 5.56% in NPS, which shows that Chinese MA postgraduate EFL learners rely heavily on the 'description' bundle to describe or provide detailed information in the writing process. Hyland (2008a) also found that the Chinese masters' dissertations pay great attention to those LBs that describe research objects or contexts.

3.5.2 Text-oriented Bundles

It can be seen from Table 8 that text-oriented texts account for more than one-third of all the LBs (35.45%) in PEL, which is much lower than that in NPS (53.49%). It shows that native American students are likelier to use text-oriented bundles to organise the text. They have a stronger sense of text organisation and pay more attention to the layout of academic texts and the logical structure of propositions. It aligns with Zare and Naseri's study (2020), which found that text-oriented research was dominantly used in English research papers. However, the chi-square test results indicate no significant differences in using text-oriented bundles between the two corpora ($p=0.093 > 0.05$).

Regarding the subcategory structuring signals bundles, the proportion of these bundles is similar for both corpora. However, there are differences in their use. American MA postgraduate students use more bundles to attract readers to focus on what is being discussed to improve the understanding of the current text, such as *in the present study, the current study, and the present study*. As for the Chinese MA postgraduate EFL learners, apart from using similar bundles like *in the study of* and *to the study of*, they pay more attention to attract readers to focus on the specific parts, such as the table or the figure, such as *is shown in*, *is shown in figure*, and *as shown in figure*, with much higher frequencies.

In terms of framing signal bundles, there are 12 bundles in both corpora, but they do not share the same percentage. Framing signals bundles account for 10.91% of PEL and 22.22% of NPS, indicating that it is more dominant in NPS. Zare and Naseri (2020) also reported that framing signals were more dominant text-oriented expressions in English research papers. There are four shared bundles (i.e., *when it comes to, in terms of the, in the form of, in the context of*). However, Chinese MA postgraduate EFL learners rely heavily on limited specific bundles, such as *from the perspective of* a frequency of 350 times.

3.5.3 Participant-oriented Bundles

Participant-oriented bundles account for the least in both PEL and NPS, 10.91% and 14.81% respectively. Compared to research-oriented LBs and text-oriented LBs, Chinese MA postgraduate EFL learners' use of participant-oriented LBs in academic texts is more similar to that of native American MA postgraduate students. They share some common features when using the sub-type of stance feature bundles. Stance features bundles, a linguistic means by which the writer expresses his attitude towards a specific proposition or point of view, makes a judgment, and establishes a proper relationship with the readers (Hyland, 2005). There are two shared bundles in this sub-type: *it is necessary to* and is *more likely to*. The authors use the bundle, so *it is essential* to express their stance in the text, which seems more objective, while the bundle *is more likely to* be used in conclusions with some uncertainties. American students have more flexibility to express their point of view or stance by using more bundles, such as *it is important to note that*, and *there was no significant*. Engagement feature bundles can help the author to form academic interaction with the reader and guide the reader to understand the text according to the author's expectations. Chinese MA postgraduate EFL learners are found to use more engagement feature bundles in quantities and frequencies, as *can be seen*, *can be seen that*, *can be seen* compared to their American counterparts.

4.0 CONCLUSION

Based on the self-built learner corpora, this study systematically examines the overall usage characteristics, structural types and function categories of four-word LBs in MA theses by Chinese postgraduate EFL learners and American postgraduate students in Applied Linguistics. There are some similarities between the two groups; however, significant differences were also evident. It is found that compared to native American MA postgraduate students, Chinese MA postgraduate EFL students rely more on four-word LBs in the construction of academic writing, and the degree of LB diversity is slightly higher, which suggests a greater reliance on formulaic expressions by less confident or low proficiency L1 students in constructing their texts (Hyland, 2008a; Li et al., 2023).

Regarding structural types, Chinese MA postgraduate EFL learners tend to have a balanced use of the three different bundles, with verb-based bundles used most frequently, followed by preposition-based bundles and noun-based bundles. Meanwhile, native American MA postgraduate students tend to use more prepositional-based bundles, which comprise half the usage. These differences suggest that Chinese MA postgraduate EFL learners still need to fully acquire the conventions of expert academic writing (Bychkovska & Lee, 2017; Chen & Baker, 2010; Pan et al., 2016). Moreover, Chinese MA postgraduate EFL learners tend to use more passive phrases, which could be linked to the influence of Chinese writing conventions. Adopting an impersonal passive voice is essential to ensuring objectivity. In terms of functional types, Chinese MA postgraduate EFL learners are more inclined to use research-oriented texts to provide descriptions in organising the writer's actions. In contrast, native American MA postgraduate students are more inclined to use text-oriented texts to organise the text.

The findings of this research carry significant implications for academic writing education. Firstly, the NPS corpus can provide realistic and authentic models for Chinese EFL learners. It can be a source reference in the native speaker's use of LBs in their academic writing, resulting in specialised English for Academic Purposes (EAP) resources that more precisely cater to the unique requirements of certain L2 English writer groups (Appel & Murray, 2020). Secondly, the findings offer valuable insights for material

designers, shedding light on various aspects of teaching methods in academic writing involving LBs and cohesive devices, ranging from the design of individual assignments to the development of the entire curriculum.

Furthermore, this study has implications for future research. Future studies should examine data from EFL and ESL learners from various language backgrounds to understand how different linguistic contexts influence lexical bundles in academic writing.

ACKNOWLEDGEMENTS

This study was not supported by any grants from funding bodies in the public, private or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR(S) CONTRIBUTION

Min Chen (Methodology, Writing – original draft, Resources)

Roslina Abdul Aziz (Conceptualization, Formal analysis, Review, Visualisation, Supervision (Corresponding author)

Syamimi Turiman (Review, Visualisation, Supervision)

REFERENCES

Aldabbus, S., & Almansouri, E. (2022). Academic writing difficulties encountered by university EFL learners. *British Journal of English Linguistics*, *10*(3), 1–11. https://doi.org/10.37745/bjel.2013/vol10n3111

Appel, R., & Murray, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes*, *46*(July 2020, 100873). https://doi.org/10.1016/j.jeap.2020.100873

Bao, K. (2024). Comparative analysis of lexical bundles in dissertation abstracts: Insights for teaching Academic English to Chinese students. *English Linguistics Research*, *13*(1), 1–8. https://doi.org/10.5430/elr.v13n1p8

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Bryman, A. (2016). *Social research methods*. Oxford University Press.

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*(November 2017), 38-52. https://doi.org/10.1016/j.jeap.2017.10.008

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14*(2), 30–49.

Cui, X. J., & Kim, Y. J. (2021). A contrastive corpus-based study on chunk use between Korean and Chinese writers in linguistics PhD dissertations. *영어교육연구*, *33*(4), 25-44. https://doi.org/10.17936/pkelt.2021.33.4.2

Day, T. (2023). *Success in academic writing*. Bloomsbury Publishing.

Deng, L., & Liu, J. (2023). Move–bundle connection in conclusion sections of research articles across disciplines. *Applied Linguistics*, *44*(3), 527-554. https://doi.org/10.1093/applin/amac040

Geluso, J. (2022). Grammatical and functional characteristics of preposition-based phrase frames in English argumentative essays by L1 English and Spanish speakers. *Journal of English for Academic Purposes, 55*(January 2022, 101072). https://doi.org/10.1016/j.jeap.2021.101072

Guan, J. (2022). A contrastive study on lexical bundles in argumentative writing by L1-Chinese and L1-English undergraduates. *English Language Teaching*, *15*(12), 1–43. https://doi.org/10.5539/elt.v15n12p43

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–192. https://doi.org/10.1177/14614456050503

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21. https://doi.org/10.1016/j.esp.2007.06.001

Jiang, F. K., & Hyland, K. (2020). "There are significant differences…": The secret life of existential there in academic writing. *Lingua*, *233*(January 2020, 102758). https://doi.org/10.1016/j.lingua.2019.102758

Kim, S., & Kessler, M. (2022). Examining L2 English university students' uses of lexical bundles and their relationship to writing quality. *Assessing Writing*, *51*(2022, 100589). https://doi.org/10.1016/j.asw.2021.100589

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Li, M., Zhang, X., & Reynolds, B. L. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: An ETS corpus study. *Applied Linguistics Review*, *14*(4), 847–873. https://doi.org/10.1515/applirev-2020-0129

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*(May 2019), 21-36. https://doi.org/10.1016/j.jeap.2019.03.008

Lyu, M., & Gee, R. W. (2020). Lexical bundles in thesis abstracts by L1 Chinese learners of English and US students. *English Language Teaching*, *13*(1), 141-155. https://doi.org/10.5539/elt.v13n1p141

Nasrabady, P., Elahi Shirvan, M., & Ehsan Golparvar, S. (2020). Exploring lexical bundles in recent published papers in the field of Applied Linguistics. *Journal of World Languages*, *6*(3), 175–197. https://doi.org/10.1080/21698252.2020.1797992

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*(March 2016), 60-71. https://doi.org/10.1016/j.jeap.2015.11.003

Pan, Y. (2024). An analysis of common problems in academic English and content-based teaching of academic English writing. *Journal of Higher Education*, *8*, 101–104. https://doi.org/10.19980/j.CN23-1593/G4.2024.08.024

Wood, A., Flowerdew, J., & Peacock, M. (2001). International scientific English: The language of research scientists around the world. In *Research Perspectives on English for Academic Purposes* (pp. 71-83). Cambridge University Press. https://doi.org/10.1017/cbo9781139524766.008

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463–489. https://doi.org/10.1093/applin/21.4.463

Yakut, I., Yuvayapan, F., & Bada, E. (2021). Lexical bundles in L1 and L2 English doctoral dissertations. *Journal of Teaching English for Specific and Academic Purposes*, *9*(3), 475-493. https://doi.org/10.22190/JTESAP2103475Y

Yu, Y. (2022). The role of psycholinguistics for language learning in teaching based on formulaic sequence use and oral fluency. *Frontiers in Psychology*, *13*(2022). https://doi.org/10.3389/fpsyg.2022.1012225

Zare, J., & Naseri, Z. S. (2020). Lexical bundles in English review articles. *Iranian Journal of English for Academic Purposes*, *9*(1), 41-56. https://journalscmu.sinaweb.net/article_105523_ac3067af3b257c5f1bc2a6c10116bf5c.pdf

Zhang, S., Yu, H., & Zhang, L. J. (2021). Understanding the sustainable Growth of EFL students' writing skills: Differences between novice and expert writers in their use of lexical bundles in academic writing. *Sustainability*, *13*(10), Article 5553. https://doi.org/10.3390/su13105553