A Contrastive Study on Lexical Bundles in Argumentative Writing by L1-Chinese and L1-English Undergraduates

Jin Guan[1]

[1] School of Foreign Languages, Shenzhen Technology University, Shenzhen, China

Correspondence: Jin Guan, School of Foreign Languages, Shenzhen Technology University, Shenzhen, Guangdong, 518118, China.

Abstract

This study compares the use of lexical bundles in argumentative essays written by L1-Chinese students and L1-English students through a corpus-based approach. The data consist of two corpora: a L1-Chinese corpus with 506 English argumentative essays produced by Chinese undergraduates in disciplines related to science and engineering and a L1-English corpus with 207 argumentative essays written by L1-English undergraduates. The identified lexical bundles were analyzed both structurally and functionally. The findings suggest that L1-Chinese students used significantly more types and tokens of lexical bundle structures than L1-English students, and also employed all the three functional categories of bundles more frequently than L1-English students. In addition, L1-Chinese students' writing was marked by a higher preference for clause-based bundles, which features the academic writing of lower-proficiency writers, and a wide use of conversational bundles (e.g. *a lot of people*), which implies their lack of awareness of academic register. The pedagogical implications are then provided regarding lexical bundles for ESL teachers.

Keywords: academic writing, argumentative writing, lexical bundles, undergraduates, corpus-based

1. Introduction

In recent decades, a growing body of corpus-based phraseology research has added weight to the significance of formulaic language. Studies in second language acquisition (SLA) have also found a close correlation between knowledge of formulaic sequences and L2 proficiency in both spoken and written discourse (Boer et al., 2006; Lewis, 2009). These formulaic sequences were named by Biber et al. (1999) as lexical bundles, a term then commonly used in corpus-based studies, referring to multi-word sequences, retrieved through a frequency-driven approach, that show a statistical tendency to co-occur in a collection of texts (Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004).

The corpus-based approach made it possible for researchers in EAP (English for academic purposes) and SLA to investigate the features of lexical bundles in a myriad of L1 and L2 academic texts. Previous research on lexical bundles used by Chinese ESL students were targeted more on undergraduates majoring in English (Huang, 2015; Hu, Shi, & Ji, 2017), and little on those in disciplines related to science and engineering. Students of science and engineering in Chinese universities usually spend one or two years studying College English as a compulsory course, with different curriculum design, teaching approaches, and learning motives compared to English majors. Therefore, it is imperative to investigate the features of lexical bundles used by science and engineering students in an attempt to provide some pedagogical implications accordingly for their English teachers and textbook developers.

Among all the writing genres targeted by previous researchers, argumentative writing has received far less attention than degree theses or published journal articles. Since argumentative writing is the most common genre in undergraduate writing where students tend to encounter more difficulties, this corpus-based study aims to compare the use of lexical bundles in English argumentative writing produced by L1-Chinese science and engineering students at Chinese mainland universities and L1-English students based in British and American universities, expecting to offer some pedagogical insights for TESOL practitioners.

2. Literature Review

Over the past two decades, lexical bundles have received considerable attention in academic writing research. Researchers examined lexical bundles mainly across genres, professional levels (expert and non-expert), disciplines, L2 proficiency levels and study levels.

Biber and his colleagues at Northern Arizona University (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004) investigated lexical bundles across a variety of genres in both instructional (e.g. classroom teaching, textbooks) and non-instructional settings (e.g. office hours, classroom management talks, written syllabi). Their research found that lexical bundles were common in all spoken university genres, while they were relatively rare in academic written genres (Biber & Barbieri, 2007). The functional analysis revealed that lexical bundles in spoken university registers mostly expressed stance, and in written university registers mostly served referential functions (Biber & Barbieri, 2007).

Some research on lexical bundles in academic prose focused on the comparison between the use of lexical bundles by experts and novice writers. Cortes (2004) studied how the lexical bundles in published research articles were used by university students in two disciplines of history and biology. It was found that students rarely used the lexical bundles in published articles, and when they used certain bundles, their use was not congruous with that of journal authors. Scott & Tribble (2006) compared lexical bundles in Polish advanced apprentice writers' writing and published research articles in the same field and found considerable variations in the frequency of forms, structures and functions across students and expert writing. Hyland's (2008) research into lexical bundles in published research articles, doctoral dissertations and master's theses supported the conclusions by Cortes and Scott and Tribble. For example, master's texts contained a larger number of clusters and less *noun phrase + of* structures than research articles.

More scholars have been drawn to the use of academic lexical bundles by ESL/EFL writers over the past decade. Chen and Baker (2010) compared lexical bundles in English essays written by L1-Chinese students, L1-English students from British universities in BAWE (British Academic Written English) corpus and published academic texts by native experts. The results showed that expert native writers employed the widest range of lexical bundles while L2 students used the smallest range and tended to overuse certain expressions (e.g. *all over the world*) which native academics rarely used. Adel and Erman's (2012) research into lexical bundles in English essays written by L1-Sweidish university freshmen in the discipline of linguistics supported Chen and Baker's study, indicating that L2 students used a smaller variety of lexical bundles than native speakers. Targeting at a more specific genre of academic essays, Bychkovska and Lee (2017) compared lexical bundles in argumentative essays produced by Chinese ESL students and native students from American universities. In contrast to the findings by Chen and Baker (2010) and Adel and Erman (2012), the study showed that L1-Chinese students used a broader range of bundles with more verb (clause) bundles and stance bundles than L1-English students.

In addition to students' essays, a number of studies on ESL/EFL lexical bundles have also focused on research papers. Pan, Reppen, & Biber (2016) studied lexical bundles in published research articles in English-mediated telecommunication journals written by L1-English and L1-Chinese academic professionals. Both structural and functional differences were found in the use of lexical bundles by the two groups. For instance, the bundles used by L2 writers were dominated by verbs and clause fragments (especially passive verb structures), while L1 writers use more bundles consisting of noun and prepositional phrases. The research into journal articles of applied linguistics found that L1-Chinese authors employed more text-oriented bundles (Li & Liu, 2016) and participant-oriented bundles (Pan & Liu, 2019) than L1-English authors.

Some researchers focused on lexical bundle use in academic prose across study levels of ESL/EFL learners. Qin (2014) explored the use of lexical bundles in academic papers by non-native English graduate students of applied linguistics across four study levels. The results showed that PhD-level students made use of a greater number and variety of target bundles than MA-level students. Regarding the structures and functions of lexical bundles, higher-level student writing contained more academic bundle structures such as noun phrases with post-modifier fragments, and more bundles serving the function of text organization and stance than lower-level student writing. Chinese scholar Huang (2015) examined frequency and accuracy of lexical bundles in English essays across junior and senior English majors. The results indicated that senior students produced lexical bundles more frequently and with a wider variety in their essays, but they did not use lexical bundles more accurately than juniors.

Research into use of lexical bundles across L2 proficiency levels also appeals to many researchers. Staples et al. (2013) studied lexical bundles in written responses across three proficiency levels in the TOEFL iBT. The study showed that all students used stance and discourse organizing bundles in a similar way and rarely used

referential bundles. Lower-level learners utilized more bundles overall but also more bundles same as those in the prompts. Appel and Wood (2016) found that low-level L2 English writers tended to use more stance and discourse-organizing expressions in their academic essays, while high-level writers made greater use of referential bundles in their writing.

We have found that most previous studies on learners' lexical bundles focused on the genre of research papers and only a few on argumentative essays. Even fewer targeted the argumentative essays written by novice undergraduates who tend to encounter more difficulties in academic writing. Among the limited number of studies on lexical bundles in English argumentative writing by L1-Chinese students, the majority focused on the essays by Chinese students at British universities (e.g. Chen & Baker, 2010) and American universities (e.g. Bychkovska & Lee, 2017). While a few studies have examined the use of lexical bundles by L1-Chinese students at Chinese universities, the students involved in the studies were predominantly English majors (Huang, 2015; Hu, Shi, & Ji, 2017). Little, however, is known about the lexical bundles used by L1-Chinese science and engineering students based in Chinese mainland universities. Given the different English curriculum designs and learning motivations between English majors and science and engineering students in Chinese universities, it is imperative to investigate the language patterns of the latter in order to facilitate the customized English teaching for them. To provide greater insights into use of lexical bundles by L1-Chinese science and engineering students at Chinese mainland universities, the present corpus-based study aims to compare the use of lexical bundles in English argumentative essays produced by L1-Chinese students at Chinese mainland universities and L1-English students based in British and American universities. The following research questions guided the study:

1) What are the differences in types and tokens of lexical bundles in English argumentative essays produced by L1-Chinese science and engineering undergraduates and L1-English undergraduates?

2) What differences exist in the structural types of lexical bundles used by the two groups?

3) What differences exist in the functional types of lexical bundles used by the two groups?

3. Data and Methodology

3.1 Description of the Corpora

Two corpora are used in the present study: the L1-Chinese ESL argumentative writing corpus (L1-Chinese corpus) and the L1-English argumentative writing corpus (L1-English corpus).

The L1-Chinese corpus contains 506 argumentative essays written by the undergraduate students in 9 science and engineering majors at Shenzhen Technology University, including 1) electronics science and technology, 2) mechanical design, manufacturing and automation, 3) new energy science and technology, 4) biomedical engineering, 5) industrial design engineering, 6) light source and illumination, 7) automotive service engineering, 8) Internet of things, 9) transportation. All the essays were submitted as writing assignments for the course College English.

The native argumentative writing data are a subset of the Louvain Corpus of Native English Essays (LOCNESS), a large collection of essays written by native English students from Britain and America. To ensure comparability, by confining the genre of writing and writers to argumentation and college-level students respectively, the study extracted 207 argumentative essays written by native British and American university students. The size of each finalized corpus for investigation is shown in table 1.

Table 1. Description of the Two Corpora

Corpus	Types	Tokens	No. of texts	Ave. length
L1-Chinese corpus	7263	150543	506	298
L1-English corpus	10754	167686	207	810

Note. Ave. length = average essay length.

3.2 Lexical Bundle Identification

Three criteria were adopted to extract lexical bundles from the two corpora. The first criterion is the length of the bundle. The study targets four-word lexical bundles because four-word combinations are the most manageable for classification and concordance checks (Chen & Baker, 2010), and they are frequently inclusive of three-word bundles (Cortes, 2004) and more common than five-word ones (Cortes, 2004; Hyland, 2008). In identifying lexical bundles, the contractions (e.g. *don't, isn't*) and hyphenated items (e.g. *face-to-face*) are regarded as single words.

The second criterion is the cut-off frequency. The cut-off frequency in lexical bundle research is somewhat arbitrary, ranging from 20 to 40 per million words (e.g. Biber et al., 2004; Hyland, 2008; Bychkovska & Lee, 2017). This study set a relatively high normalized frequency of 40 times per million words to ensure the representativeness of the retrieved lexical bundles. This threshold was translated into 6 raw occurrences in L1-Chinese corpus and 7 raw occurrences in L1-English corpus.

The last criterion is the dispersion. Word combinations are usually required to occur in at least 3-5 texts (e.g. Biber &Barbieri, 2007; Cortes, 2004) to avoid idiosyncrasies from individual writers or speakers. This study set the dispersion threshold at 5 different texts, which is advisable for a 200,000-word corpus (Biber & Barbieri, 2007).

The corpus tool AntConc 3.5.9 (Anthony, 2020) was used to retrieve the four-word bundles. The context-based words (e.g. *in the United States*), the topic-specific words (e.g. *to have an abortion*, *people feel stressful in*) and the given word combinations from the prompts (e.g. *global fight against Covid*, *Hellen Keller said although*) were manually excluded from the bundle lists. It is worth noting that a large number of L1-Chinese students used the exact words or expressions from the writing prompts, leading to a dramatic drop in the number of lexical bundles after refinement (see table 2). To avoid inflation of quantitative results, the bundle overlaps were addressed according to the practice of Chen and Baker (2010) and Bychkovska and Lee (2017). First, the two bundles that were completely overlapped would be combined for counting. For example, *a matter of fact* and *as a matter of* both occur 8 times, as part of the longer sequence *as a matter of fact*. Second, if two or more four-word bundles overlap, the lower-frequency bundles would be subsumed into the higher-frequency bundle. For example, *pay more attention to* occurs 24 times, and *should pay more attention* 14 times. Concordances show that all the 14 *should pay more attention* are followed by *to*, so *should pay more attention* will be combined into *pay more attention to*. There are 33 cases of overlaps found in L1-Chinese corpus in total, and 2 cases in L1-English corpus. The types of lexical bundles decreased significantly after refinement, shown in table 2.

Table 2. The Types of Four-word Lexical Bundles Before and After Refinement in Two Corpora

	Types of Lexical Bundles	
	Before refinement	After refinement
L1-Chinese corpus	490	177
L1-English corpus	77	48

3.3 Lexical Bundle Analysis

The identified lexical bundles are analyzed from both structural and functional perspectives. Biber and colleagues (Biber et al., 1999, 2004) proposed three main structures of lexical bundles: noun phrase-based (NP-based), prepositional phrase-based (PP-based) and verb phrase-based (VP-based) bundles. NP-based bundles include any noun phrases with post-modifier fragments (e.g. *the development of the*, *the way in which*). PP-based bundles are prepositional phrases with a noun-phrase fragment (e.g. *on the other hand*, *in relation to the*). VP-based bundles refer to the word sequences with a verb component (e.g. *pay more attention to*, *is one of the*). To present a clearer and comprehensive picture of the structural patterns of lexical bundles in the present study, we modified Biber et al.'s classification and proposed four major categories: 1) NP-based, 2) PP-based, 3) VP-based, 4) clause-based (see table 7).

This study will adopt the functional taxonomy developed by Biber et al. (2004) who examined various genres of spoken and written discourse in university settings, such as conversations, course notes, textbooks, academic essays, and proposed three functional categories of lexical bundles: stance bundles, discourse organizers and referential bundles. Stance bundles convey a writer's certainty (or uncertainty) or attitude about a contention (e.g. *are more likely to*, *it has to be*). Discourse organizers show relationships between prior and coming discourse (e.g. *in this essay I*, *on the other hand*). Referential bundles are used to specify an entity or important attributes of an entity (e.g. *in the context of*, *the extent to which*). The subcategories of each function and examples are shown in table 3.

Table 3. Functional Taxonomy of Lexical Bundles (Biber et al., 2004)

Categories	Subcategories	Examples
Stance	Epistemic	*I don't think so, are more likely to*
	Attitudinal/modality	*you have to be, to be able to*
Discourse organizers	Topic introduction/focus	*take a look at, I would like to*
	Topic elaboration/clarification	*nothing to do with, on the other hand*
Referential	Identification/focus	*is one of the, one of the most*
	Imprecision	*or something like that, and things like that*
	Specification of attributes	*a lot of people, in terms of*
	Place/time/text reference	*at the same time, the end of the*

4. Results and Discussion

4.1 Comparison of Types and Frequency of Lexical Bundles in the Two Corpora

The numbers of types and frequency of the finalized bundles in two corpora are presented in table 4. As shown in table 4, L1-Chinese students used 3.7 times more types of lexical bundles than their L1-English peers, and the normalized frequency of L1-Chinese bundles was 4.4 times higher than L1-English bundles. This finding is consistent with many previous studies (e.g. Bychkovska & Lee, 2017; Pan et al., 2016; Wei & Lei, 2011), suggesting that L1-Chinese writers used both a greater variety and number of lexical bundles than L1-English writers, reflecting the unitary nature of the writing from L1-Chinese students.

Table 4. Number of Types and Frequency of Lexical Bundles in the Two Corpora

	No. of types	Raw frequency	Normalized frequency (pmw)
L1-Chinese corpus	177	2156	14321
L1-English corpus	48	489	2916

Note. pmw = per million words

A reason for L1-Chinese writers using more types and tokens of bundles may be that Chinese students are often trained in English class to be more aware of using language patterns, including prefabricated clusters, as a safeguard against being unidiomatic and inappropriate in their writing. Another explanation is likely to be the language transfer from Chinese. For example, *with the development of*, the fifth most frequently used bundle in L1-Chinese corpus (53 tokens), was non-existent in L1-English writing, mostly because *with the development of* is a very common and natural expression in Chinese to introduce the background of a topic, such as *with the development of economy, with the development of science and technology, with the development of China, with the development of social media*, all of which occurred many times in L1-Chinese corpus. In addition, the overuse of conversational bundles by L1-Chinese students tends to spur the inflation of the overall bundle types and frequency. For instance, the bundles with the quantifier *a lot of* or *lots of* (*a lot of people, a lot of time, have a lot of, there are a lot of, there are lots of*) occurred 39 times in L1-Chinese corpus, while L1-English writers used none of them. Lastly, it is worth noting that the misused bundles – *last but no least* (11 tokens), *as the development of* (6 tokens), *the last but not* (6 tokens) – also contribute to the greater types and tokens of bundles from L1-Chinese corpus.

Another finding from the comparison between two finalized bundle lists was that nine lexical bundles are shared by the two corpora (see table 5). A further examination into the full bundle lists showed that the four shared bundles – *on the other hand, at the same time, is one of the and one of the most* – are also among the top one third of the most frequent bundles in both corpora. *On the other hand* is the favorite bundle for both L1-Chinese and L1-English students, but L1-Chinese students used it far more frequently than their English counterparts. Likewise, the bundle *at the same time* was used 5.5 times more frequently by L1-Chinese students than by L1-English students. The considerably higher frequency of some bundles in L1-Chinese corpus can be explained in part by the cases of overuse. For example, *on the other hand* is used in the context of contrast, but some Chinese students used it in different logical relations other than contrast. Some examples are shown below.

Example 1: *Parent's bad behavior and lack of supervision are reason of youth crime. On the other hand, the rise of youth crime is also shaped by the society, which has not paid enough attention to the youth in all aspects.*

Example 2: *Some parents even think the law is too easy to learn for their children so that they usually ignore this aspect of education. On the other hand, in many places of schools, they just casually spread the legal knowledge and do not want to pay more attention on this.*

Example 3: *So the widespread use of social media has led to the lack of face-to-face communication between people. This is one of the disadvantages. On the other hand, students indulging in social media often forget to spend time in order to play Weibo and return to WeChat, resulting in poor academic performance.*

Table 5. Shared Bundles of the Two Corpora

	L1-Chinese corpus		L1-English corpus	
	Raw freq	Normalized freq (pmw)	Raw freq	Normalized freq (pmw)
on the other hand	60	399	28	167
at the same time	55	365	11	66
is one of the	23	153	17	101
the best way to	14	93	8	48
one of the most	12	80	20	119
the beginning of the	9	60	7	42
a great deal of	7	46	11	66
as a result of	7	46	15	89
it is obvious that	6	40	9	54

Note. freq = frequency; pmw = per million words

4.2 Comparison of Structural Types and Tokens of Lexical Bundles

Table 6 shows the proportional distribution of main structural categories in the two corpora, and table 7 gives detailed information about the types and tokens of each category of lexical bundles, as well as the log-likelihood (LL) value of raw frequencies of the bundles using the spreadsheet of Paul Rayson's log likelihood calculator at http://ucrel.lancs.ac.uk/llwizard.html. As shown in table 6, L1-Chinese students used a higher proportion of VP-based and clause-based bundles than L1-English students in both types and tokens, whereas L1-English relied more on NP-based and PP-based bundles. This result is consistent with the findings of some previous research (Pan et al., 2016; Bychkovska & Lee, 2017; Lu & Deng, 2019). It is also noticeable that L1-Chinese students used adjective-based bundles while their L1-English peers did not. The adjective-based bundles used by L1-Chinese students are *more and more important, difficult for them to, nowadays more and more, world more and more*. This result indicates that the pattern *more and more* was overused by L1-Chinese students, which may be due to the inadequate paraphrasing skills or grammatical transfer from the Chinese expression "越来越."

Table 6. Proportional Distribution of Main Structural Categories in the Two Corpora

Categories	Types (%)		Tokens (%)	
	L1-Chinese	L1-English	L1-Chinese	L1-English
NP-based	23.6	35.4	24.4	35.0
PP-based	21.3	22.9	27.7	27.6
VP-based	25.3	22.9	22.6	22.5
Clause-based	27.5	18.8	24.0	14.9
Adjective-based	2.2	0	1.3	0
Total	100	100	100	100

Table 7. Distribution of Structural Categories in the Two Corpora

Categories	Subcategories	Types		Tokens		LL
		L1-Ch	L1-En	L1-Ch (pmw)	L1-En (pmw)	
NP-based	NP with of-phrase fragment	18	13	187(1242.17)	128(763.33)	18.39***
	NP with other post-modifier fragment	18	4	256(1700.51)	43(256.43)	192.07***
	Other NP	5	-	79(524.77)	-	-
	Subtotal	41	17	522(3467.45)	171(1019.76)	226.16***
PP-based	PP with embedded of-phrase fragment	12	5	113(750.61)	59(351.85)	23.57***
	Other PP fragment	26	6	486(3228.31)	76(453.23)	379.61***
	Subtotal	38	11	599(3978.93)	135(805.08)	369.05***
VP-based	Copula be + NP/adjective phrase	14	6	201(1335.17)	67(399.56)	85.35***
	VP + dependent clause fragment	3	-	29(192.64)	-	-
	VP with active verb	13	4	135(896.75)	35(208.72)	74.08***
	VP with passive verb	-	1	-	8(47.71)	-
	VP with infinitive verb	15	-	124(823.68)	-	-
	Subtotal	45	11	489(3248.24)	110(655.99)	301.72***
Clause-based	Anticipatory it + copula be fragment	9	2	78(518.12)	16(95.42)	51.5***
	Dependent clause fragment	16	3	232(1541.09)	27(161.02)	208.74***
	NP + active verb fragment	15	1	129(856.90)	7(41.74)	146.92***
	NP + copula be fragment	9	3	79(524.77)	23(137.16)	38.85***
	Subtotal	49	9	518(3440.88)	73(435.34)	427.08***
Adjective-based	Adjective phrase	4	-	28(186.00)	-	-
Total		177	48	2156(14321.49)	489(2916.16)	1321.84***

Note. pmw = per million words; LL = log-likelihood value; *** = significant at p<0.001 level.

4.2.1 NP-Based Lexical Bundles

While L1-Chinese students used a smaller percentage of NP-based bundles, table 7 shows that they used more types and tokens of NP-based bundles in all the three subcategories than L1-English students. The most striking difference lies in the use of noun phrases with other post-modifier fragment. L1-Chinese students used 4.5 times as many bundles of this structural pattern as L1-English students. The bundle *the best way to* was shared by both groups, while the NP-based bundles with a relative clause, namely appositive clause in this case, *the fact that the* and *the fact that they*, were only found in L1-English group. This result converges with the finding of Chen and Baker (2010) that L1-Chinese students did not use this type of clause as frequently as L1-English students did. Another interesting finding in this subcategory was that in L1-Chinese corpus three bundles – *last but no least, last but not the, the last but not* – are misused versions of the phrase *last but not least* which does not occur in L1-English bundle list.

We also found substantial differences in the use of NP beginning with *a/an*, *the* and *one* in the two corpora. Lu and Deng (2019) found that L1-Chinese students used less NP-based bundles beginning with an indefinite article, but in our study L1-Chinese students employed significantly more types and tokens of noun phrases beginning with *a/an* than their L1-English peers (LL=109.49, p<0.001), as shown in table 8. It can be seen from table 9 that most nominal phrases beginning with a/an used by L1-Chinese students contain quantifiers – *a large number of, a lot of people, a lot of time, an increasing number of, a great deal of* – only one of which appears in L1-English bundles. Of those bundles, *a lot of people* and *a lot of time* are typically conversational phrases, neither of which is in the list of L1-English bundles. This type of nominal phrase with an informal quantifier *a lot of* characterizes learner writing (Chen & Baker, 2016; Bychkovska & Lee, 2017). The excessive use of *a lot of* by Chinese students may be caused by their inability of finding equivalent collocations or unawareness of academic register.

The two groups of students do not differ much in the types and tokens of NP beginning with *the*, but the bundles are truly different. The most heavily used NP beginning with *the* for L1-Chinese students is *the development of the* which does not exist in L1-English bundles. The concordance results show that *the development of the* was frequently used as part of the pattern *with the development of the + noun* which will be discussed in the analysis of PP-based bundles. *The most important thing*, a typically conversational bundle, was used by L1-Chinese students at a relatively high frequency (16 tokens), but not by L1-English students. This result is consistent with the finding of Shin (2019) that learners tended to use *the most important thing* much more than the natives. The heavy use of this bundle by leaners may be attributed to their inadequate English proficiency of locating a more precise word for a certain concept or unawareness of academic register.

Table 8 shows that L1-English students used NP beginning with *one* twice as frequently as L1-Chinese students did. They shared the bundle *one of the most*, which is known as a common NP bundle for both learners and natives in many previous studies (Chen & Baker, 2010; Bychkovska & Lee, 2017; Shin, 2019; Lu & Deng, 2019). Besides, L1-English students also used *one of the greatest*, but Chinese students did not.

Table 8. Types and Tokens of NP Beginning with *a/an, the* and *one* in the Two Corpora

	L1-Chinese		L1-English		LL
	Types	Tokens(pmw)	Types	Tokens(pmw)	
NP beginning with *a/an*	12	147 (976.47)	3	25(149.09)	109.49***
NP beginning with *the*	10	111(737.33)	9	94(560.57)	3.84
NP beginning with *one*	1	12(79.71)	2	27(161.02)	4.42*
Total	23	270(1793.51)	14	146(870.67)	52.11***

Note. pmw = per million words; LL = log-likelihood value; * = significant at p<0.05 level; *** = significant at p<0.001 level.

Table 9. NP Beginning with *a/an, the* or *one* in the Two Corpora

	L1-Chinese	L1-English
NP beginning with *a/an*	*a good way to (55)*	*a great deal of (11)*
	a large number of (17)	*a member of the (7)*
	an important role in (11)	*an example of this (7)*
	a lot of people (10)	
	a lot of time (8)	
	an increasing number of (8)	
	a good way for (7)	
	a great deal of (7)	
	a direct consequence of (6)	
	a good command of (6)	
	a great way to (6)	
	a negative impact on (6)	
NP beginning with *the*	*the development of the (20)*	*the end of the (17)*
	the most important thing (16)	*the fact that the (16)*
	the best way to (14)	*the only way to (11)*
	the rapid development of (13)	*the rest of the (11)*
	the joint efforts of (11)	*the best way to (8)*
	the beginning of the (9)	*the fact that they (8)*
	the spread of the (9)	*the idea of a (8)*
	the meaning of the (7)	*the majority of the (8)*
	the last but not (6)	*the beginning of the (7)*
	the power to overcome (6)	
NP beginning with *one*	*one of the most (12)*	*one of the most (20)*
		one of the greatest (7)

Note. The number in the parentheses refers to the raw frequency of the bundle in the corpus.

4.2.2 PP-Based Lexical Bundles

Table 7 shows that L1-Chinese students used more than 3 times as many PP-based bundles as L1-English students did. A close examination reveals that the bundles they used begin with different prepositions.

As shown in table 10, both L1-Chinese and L1-English students used PP bundles beginning with *in* more frequently than other PP with other prepositions. Besides, L1-Chinese students used more types and tokens of PP beginning with *in* than L1-English students. One reason might be the inflation of the prepositional phrase *in my opinion* in L1-Chinese corpus, which occurs 62 times in 6 bundles, suggesting a unitary nature of writing from L1-Chinese writers. Chinese students also used more nouns in the phrase frame *in the + noun + of*, such as *parts*, *face*, *process* and *course*, while L1-English students only used *case* as noun in this frame. This indicates that Chinese students tend to use the phrase frame *in the + noun + of* to describe the background or the progress, but L1-English students to refer to a certain topic. In addition, L1-English students used the structure of PP with a relative clause – *in a way that*, which is absent in L1-Chinese corpus.

With respect to PP beginning with *as*, the two groups shared the bundle *as a result of*, though L1-English students used it more frequently than L1-Chinese students. We also found that L1-Chinese students used more conversational bundles – *as soon as possible, as for me I*, suggesting that they may be struggling with academic register. A misused bundle *as the development of* was found in 6 cases of L1-Chinese student writing. The students confused preposition *as* and conjunction *as*, shown in the following examples.

Example 1: *As the development of technology, we can update our status on social media platforms.*

Example 2: *As the development of economic, some parents buy anything for their children if they ask and even agree their children to do all they want.*

Both L1-Chinese and L1-English students used three types of PP beginning with *at*, but Chinese students used them much more frequently than L1-English students. For instance, L1-Chinese students used *at the same time* 55 times, significantly more than L1-English students. The two groups of students converged on the use of PP beginning with *at* describing concepts about time.

The PP bundle beginning with *on* – *on the other hand* – ranked the top in both L1-Chinese and L1-English bundle lists, though Chinese students misused it in many cases. The bundle *on the road of* was only used by Chinese students. Concordance results show that Chinese students used this bundle mainly in the context "*go on the road of crime*" or "*go on the road of committing crimes*", an unidiomatic English expression, which is translated literally from the Chinese idiom "走上犯罪道路."

L1-Chinese students used three PP beginning with *with*, but L1-English did not. The three bundles are all related to development – *with the development of, with the rapid development*, and *with the progress of*. We found students tend to use these bundles at the beginning of a sentence to set a background for the topic (e.g. *with the development of the times, with the development of the era, with the development of the society, with the progress of science and technology, with the rapid development of China's economy*). The heavy use of these patterns is likely to be affected by language transfer from the Chinese expression "随着…的发展."

Table 10. Types and Tokens of PP Beginning with Different Prepositions in the Two Corpora

	L1-Chinese		L1-English		LL
	Types	Tokens(pmw)	Types	Tokens(pmw)	
PP beginning with *in*	14	165(1096.03)	3	33(196.80)	110.87***
PP beginning with as	8	76(504.84)	1	15(89.45)	51.53***
PP beginning with *at*	3	69(458.34)	3	28(166.98)	22.59***
PP beginning with *on*	3	105(697.48)	1	28(166.98)	56.17***
PP beginning with *with*	3	68(451.70)	-	-	-
PP beginning with *for*	3	35(232.49)	-	-	-
PP beginning with *to*	-	-	2	22(131.20)	-
PP beginning with *of*	1	6(39.86)	1	9(53.67)	0.569
Others	3	72(478.27)	-	-	-
Total	38	596(3959.00)	11	135(805.08)	365.78***

Note. pmw = per million words; LL = log-likelihood value; *** = significant at p<0.001 level.

Table 11. PP Beginning with Different Prepositions in the Two Corpora

	L1-Chinese	L1-English
PP beginning with *in*	*in many parts of (the world) (21)* *in my opinion it (20)* *in the face of (16)* *(only) in this way can (16)* *in the process of (12)* *in our daily life (11)* *in recent years the (11)* *in my opinion I (10)* *in my opinion the (10)* *in the first place (10)* *in my opinion there (8)* *in my opinion we (8)* *in the course of (6)* *it in my opinion (6)*	*in the case of (18)* *in a way that (8)* *in the long run (7)*
PP beginning with *as*	*as long as we (18)* *as soon as possible (13)* *as one of the (10)* *(as) a matter of fact (8)* *as a result the (8)* *as a result of (7)* *as for me I (6)* *as the development of (6)*	*as a result of (15)*
PP beginning with *at*	*at the same time (55)* *at the first time (8)* *at the age of (6)*	*at the same time (11)* *at the end of (9)* *at the beginning of (8)*
PP beginning with *on*	*on the other hand (60)* *on the one hand (39)* *on the road of (6)*	*on the other hand (28)*
PP beginning with *with*	*with the development of (53)* *with the rapid development (9)* *with the progress of (6)*	
PP beginning with *for*	*for a long time (21)* *for the reason that (8)* *for the sake of (6)*	
PP beginning with *to*		*to the fact that (14)* *to a certain extent (8)*
PP beginning with *of*	*of the world and (6)*	*of the most important (9)*
others	*all over the world (46)* *from all over the (17)* *all over the country (9)*	

Note. The number in the brackets refers to the raw frequency of the bundle in the corpus.

4.2.3 VP-Based Lexical Bundles

Table 7 shows that L1-Chinese students used significantly more types and tokens of VP-based bundles than L1-English students, as many previous studies (Bychkovska & Lee, 2017; Chen & Baker, 2010; Pan et al., 2016; Lu & Deng, 2019) have demonstrated. L1-Chinese students also used more subcategories of VP-based bundles. A close examination revealed that L1-Chinese students used the subcategory *copula be + NP/adjective phrase* three times as frequently as L1-English students. But both groups showed heavy use of certain phrase frames in this subcategory, such as *copula be + more and more* (e.g. *become more and more, is more and more, are more and more*) in L1-Chinese corpus, which is absent in L1-English corpus, and *modal verb + be able to* (*will be able to, would be able to, should be able to*) in L1-English corpus. The heavy use of *more and more* could be explained by Chinese students' lack of knowledge about academic register.

Both groups used more VP with active verb than VP with passive verb. But L1-Chinese students relied more heavily on VP with active verb than L1-English students, with no VP with passive verb found. The only VP with passive verb in L1-English corpus is *should be allowed to*. Another noticeable difference is that L1-Chinese students used 15 types of VP with infinitive verb (e.g. *to solve the problem, to fight against the, to address the issue, to cope with the*), constituting one third of all VP-based bundles, while the native corpus contained no instance of this subcategory.

4.2.4 Clause-Based Lexical Bundles

Clause-based bundles top the list for bundle type in L1-Chinese corpus, accounting for 27% of the total. They are also used significantly more frequently by L1-Chinese students than by L1-English students. Dependent clause fragment was the favorite subcategory for both groups. A close examination revealed that the top two dependent clause fragment bundles are *as we all know* and *as far as I* in L1-Chinese corpus, neither of which exists in L1-English corpus. Concordance results show *as far as I* is included in such spoken phrases as *as far as I am concerned* and *as far as I see* placed at the beginning of a sentence to introduce the writer's stance. The most popular dependent clause fragment for native English students is *when it comes to*, often used with a low level of formality. The heavy use of these spoken or informal phrases suggests that both the two groups of students may have little awareness of academic register or knowledge about how to introduce their views in academic writing.

With respect to NP + active verb fragment, both groups used this subcategory to introduce a topic or a view (e.g. *I think we should, this essay will discuss, we should not only, some people think that, I would like to*), but Chinese students used significantly more types with higher frequency. In the subcategory of anticipatory it + copula be fragment, the adjectives following *copula be* are more diverse in L1-Chinese corpus, including *easy, difficult, necessary, important, obvious*, while only two of them – *important and obvious* – are in L1-English corpus. In the subcategory of NP + copula be fragment, 6 out of 9 bundles in L1-Chinese corpus and 1 out of 3 in L1-English corpus begin with existential *there*, but Chinese students tend to make grammatical mistakes using existential *there*, as illustrated in the following examples:

Example 1: *There are lots of words cannot translate directly in English, such as "braised lion head", this is a famous dish in China.*

Example 2: *There are lots of reasons will cause you stressful.*

Example 3: *There are many people show their personal life on the media platform, ...*

Example 4: *There are many people die for it.*

Chinese students' frequent use of *there be* and sometimes with mistakes may result from the L1 transfer as the phrase frame *there be + sb/sth + main verb* is very common in Chinese.

4.3 Comparison of Functional Types and Frequency of Lexical Bundles

Table 12 compares the proportions of main functional categories of lexical bundles in the two corpora. The favorite functional category for L1-Chinese students is stance bundle, accounting for 36.7% of all bundle types, while half of the bundle types L1-English students used are referential. Since stance bundles are characteristic of conversations (Biber et al., 1999; 2004), Chinese students' writing shows features of spoken discourse. Discourse organizers are the least common bundles for both two groups. Table 13 compares the three functional categories of lexical bundles in the two corpora. The results show that L1-Chinese students used all the three functional categories of bundles significantly more frequently than L1-English students.

4.3.1 Stance Bundles

In terms of stance bundles, L1-Chinese students used 39 epistemic bundles, many of which are personal. For example, 6 epistemic bundles contain *in my opinion* and 5 include *I think/believe*, either of which is existent in L1-English epistemic bundles. It can be seen that Chinese students tend to make their stance more bluntly in argumentation. Moreover, Chinese students used bundles with a variety of evaluative adjectives (e.g. *is a good way, is not a good, the best way to, is the most important, a great way to*) to emphasize personal opinions without providing solid evidence, so they may sound overly subjective or biased for readers (Pan et al., 2016). On the other hand, L1-English students preferred impersonal epistemic bundles. The top two frequently used epistemic bundles for L1-English students are *the fact that the* and *to the fact the*.

A close examination into attitudinal/modality bundles shows that L1-Chinese students used more types and tokens of bundles in this subcategory than their English peers. Both groups used more obligatory/directory and ability bundles than desire and intention/prediction bundles. We found that Chinese students like to use bundles that contain *pay attention to* (*pay more attention to, pay attention to the*) and pronoun *we* (*so we have to, we need to do, we should not only*) to express obligations and directives, which constitute over one third of all obligatory/directory bundles. These bundles, more of a preaching, may hardly resonate with readers. However, L1-English students prefer impersonal bundles that contain passive verb (*should be allowed to*) or clause (*it is important to, it should not be*) to express obligations and directives, which are more objective and convincing to readers.

4.3.2 Discourse Organizers

Both groups used the smallest proportion of discourse organizers with a preference to topic elaboration bundles. The most frequently used discourse organizer is *on the other hand*, shared by the two corpora. L1-Chinese students employed significantly more types and tokens of discourse organizers than native English students. A close examination found that L1-Chinese students tend to use a number of sequential adverbials which constitute half of the topic introduction bundles (e.g. *last but not least, first of all we, in the first place*) at the beginning of a sentence for new topic being introduced, while native students only used *when it comes to*. Chinese students' heavy use of sequential adverbials may be explained by their intention of clarifying the logic of information, but the logic is often achieved only on the surface.

As regards the topic elaboration bundles, L1-Chinese students used a wider range of bundles to explain a topic than L1-English students. 11 out of 35 such bundles for Chinese students are related to the elaboration on solutions (e.g. *to solve the problem, how to cope with, to address this issue, to deal with it, ways to deal with*). L1-English students only used 3 topic elaboration bundles – *on the other hand, have the right to*, and *significantly changed people's lives*. We also found that L1-Chinese students are more identical in the use of some bundles. For instance, 15 students used *coin has two sides* (15 tokens) to elaborate on the other side or influence of a subject. This may result from the study of test-oriented writing templates for some students, since *every coin has two sides* is known as a skeleton-key phrase in many argumentative writing templates. It is also found that Chinese students have difficulties in distinguishing academic register with non-academic register. 15 students used the bundle *as long as we* (18 tokens) to explain the condition in a context. *As long as* is more frequently used in spoken discourse (74.17 pmw) than in academic discourse (28.37 pmw), shown in British National Corpus.

4.3.3 Referential Bundles

With respect to referential bundles, both groups used a large percentage of referential bundles. L1-Chinese students favored quantity specification bundles (e.g. *more and more people, a large number of, a lot of people, a lot of time*), accounting for about one third of all referential bundles, while L1-English students only used two such bundles – *a great deal of* and *the majority of*. The 19 quantity specification bundles used by Chinese students are marked by informality, including 5 bundles containing *a lot of* and *lots of* (e.g. *a lot of people, have a lot of, there are lots of*), as well as 8 bundles containing *more and more* (e.g. *more and more people, become more and more, are more and more*), none of which exist in L1-English bundles. The overuse of these informal quantifying bundles features low-proficiency learners (Chen & Baker, 2016).

Following quantifying bundles, L1-Chinese students used 18 types of framing attributes which are the most popular referential bundle type for L1-English students. The framing attributes used by Chinese students are more similar illustrating the progress of a topic, 7 out of 18 consisting of *development* or *progress* (e.g. *with the development of, with the progress of*) while those by native students are more diverse.

In terms of the time/place/text-deictic bundles, native students only used time-deictic bundles, whereas Chinese students used an almost equal number of time and place-deictic bundles. It is worth noting that the place-deictic bundles used by Chinese students are all too general, e.g. *all over the world, in many parts of (the world), all over the country, people around the world*. These expressions may reflect that learners have a tendency to overgeneralize a certain topic (Chen & Baker, 2009), which may decrease the writer's credibility in readers' mind.

Table 12. Proportional Distribution of Main Functional Categories in the Two Corpora

Categories	Types (%)		Tokens (%)	
	L1-Chinese	L1-English	L1-Chinese	L1-English
Stance bundles	36.7	41.7	36.3	38.4
Discourse organizers	29.4	8.3	29.5	11.5
Referential bundles	33.9	50	34.3	50.1
Total	100	100	100	100

Table 13. Distribution of Functional Categories in the Two Corpora

Categories	Subcategories	Types		Tokens		LL
		L1-Ch	L1-En	L1-Ch (pmw)	L1-En (pmw)	
Stance bundles	Epistemic	39	10	516(3427.59)	97(578.46)	361.33***
	Attitudinal/ modality — Obligatory/ directory	11	4	108(717.40)	30(178.91)	55.61***
	Ability	10	4	121(803.76)	42(250.47)	48.94***
	Desire	4	1	31(205.92)	11(65.60)	12.20***
	Intention/ prediction	1	1	6(39.86)	8(47.71)	0.11
	Subtotal	26	10	266(1766.94)	91(542.68)	109.51***
	Total	65	20	782(5194.53)	188(1121.14)	457.67***
Discourse organizers	Topic introduction	17	1	288(1913.07)	12(71.56)	345.76***
	Topic elaboration/clarification	35	3	347(2304.99)	44(262.40)	300.77***
	Total	52	4	635(4218.06)	56(333.96)	633.61***
Referential bundles	Identification/focus	9	5	100(664.26)	60(357.81)	14.89***
	Framing attributes	18	11	208(1381.67)	117(697.73)	36.58***
	Quantity specification	19	2	192(1275.38)	19(113.31)	184.06***
	Place/time/text-deixis	14	6	239(1587.59)	49(292.21)	157.86***
	Total	60	24	739(4908.90)	245(1461.06)	315.78***
Total		177	48	2156(14321.49)	489(2916.16)	1321.84***

Note. pmw = per million words; LL = log-likelihood value; *** = significant at $p<0.001$ level.

5. Conclusion

This study compared lexical bundles used by L1-Chinese and L1-English university students in argumentative writing. Significant differences were found in the structures and functions of lexical bundles used by the two groups of students. The structural analysis on the bundles showed that L1-Chinese students used significantly more types and tokens of all the major structures than L1-English students. As regards proportional distributions, L1-Chinese students exhibited a preference for clause-based bundles, which marks the academic writing of lower-proficiency writers (Biber, Gray, and Poonpon, 2011). The functional analysis on the lexical bundles demonstrated that L1-Chinese students used all the three functional categories of bundles significantly more

frequently than L1-English students. Both groups of students employed a relatively higher frequency of stance bundles, among which Chinese students showed a strong inclination to personal epistemic bundles, downplaying the objectivity of the text. In addition, Chinese students' writing is marked by a wide use of conversational referential bundles of quantity specification (e.g. *a lot of people*), which implies their lack of awareness of academic register.

The marked differences between L1-Chinese and L1-English writing found in this study could provide some pedagogical insights. First, some common errors revealed in the lexical bundle use by Chinese students may result from their inadequate lexico-grammatical knowledge. For example, the misuse of *last but not least* and *on the other hand*, the grammatical errors in the sentence beginning with *there be*, and the inability to distinguish *as* and *with* in describing the topic of progress were quite common in Chinese students' writing. Teachers should arrange more tasks targeting at these mistakes in class to help students improve language accuracy. Second, this study found that L1-Chinese students had little awareness of academic register. A large number of lexical bundles typical of spoken register such as *a lot of people, more and more people, as far as I (am concerned), as long as we* marked Chinese students' writing. It is suggested that teachers could integrate corpus-based vocabulary teaching into language class, comparing and contrasting texts of spoken and academic registers in order to help students use appropriate words and expressions in academic prose. Third, some lexical bundles used by Chinese students are not idiomatic due to the transfer of Chinese language conventions. For example, the phrase frame *with the development of + noun* frequently placed at the beginning of a background introduction in Chinese students' writing might be affected by conventional expressions about progress, advancement and development in Chinese. To help students achieve idiomacy, teachers could incorporate sessions about the comparison and contrast of Chinese and English language into the teaching syllabus. Last, many Chinese students were likely to copy the words in the prompts into their own writing, and thus their word choices tend to converge, which gave rise to the inflation of lexical bundles. This may be caused by their inflexibility in the use of vocabulary, so teachers should focus on honing the skills of paraphrasing in teaching.

This study has some potential limitations. First, while all the texts in L1-English and L1-Chinese corpora are argumentative writing, the range of topics, the level of difficulty of topics and the writing requirements are different. L1-English texts are mostly over 500-word essays with pre-research and references, and the topics are more specific and difficult ranging from government policies to legal and social issues. On the other hand, L1-Chinese texts are compositions with 200 to 300 words and without pre-research and references. The topics are relatively easy and more general. Since the context of writing (e.g. genres, disciplines, authors, audience) has a vital impact on the use of lexical bundles (Hyland, 2008), this imbalance may affect the distributions of lexical bundles to some extent. We suggest that the lexical bundles in L1 and L2 writing with same prompts and requirements could be explored in the future research. Second, the reference corpus in this study was adapted from LOCNESS corpus which was built around 30 years ago. To represent the features of students' writing in the present times, more updated reference corpus is recommended for future studies.

Acknowledgments

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: differences between high- and low-proficiency levels. *Language Assessment Quarterly, 13*(1), 55-71. https://doi.org/10.1080/15434303.2015.1126718

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116. https://doi.org/10.1016/j.jeap.2006.05.001

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-189). Amsterdam: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus linguistics by the Lune: a festschrift for Geoffrey Leech* (pp. 71-93). Frankfurt: Peter Lang.

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

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *TESOL Quarterly, 45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research, 10*(3), 245-261. https://doi.org/10.1191/1362168806lr195oa

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

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131-145). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.9.09cor

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2013). The purpose of this study is to: connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43. https://doi.org/10.1016/j.jeap.2012.11.002

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

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

Hyland, K. (2008). 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. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics, 32*, 150-169. https://doi.org/10.1017/S0267190512000037

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System, 53*, 13-23. https://doi.org/10.1016/j.system.2015.06.011

Hu, Y. J., Shi, H. M., & Ji, P. (2017). A corpus-based contrastive analysis on the structures and functions of lexical bundles in argumentative writing of English learners and L1-English speakers. *Foreign Languages Research, 34*(4), 58-62.

Lewis, M. (2009). *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency, and proficiency.* Saarbrücken, Germany: VDM Verlag.

Li, M. X., & Liu, Y. B. (2016). A comparative corpus-based study of Lexical bundles used in Chinese and native expert academic writings. *Modern Foreign Languages, 39*(4), 507-515+584.

Liu, X. F., & Deng, J. L. (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

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283-304. https://doi.org/10.1075/ijcl.11.3.04nes

Pan, F., & Liu, C. (2019). Comparing L1-L2 differences in lexical bundles in student and expert writing. *Southern African Linguistics and Applied Language Studies, 37*(2), 142-157. https://doi.org/10.2989/16073614.2019.1625276

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

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42*, 220-231. https://doi.org/10.1016/j.system.2013.12.003

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* Amsterdam, the Netherlands: John. Benjamins. https://doi.org/10.1075/scl.22

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

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214-225. https://doi.org/10.1016/j.jeap.2013.05.002

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal, 42*(2), 155-166. https://doi.org/10.1177/0033688211407295

Anthony, L. (2020). *AntConc (Version 3.5.9)* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software