A Learner Corpus of High-Stakes Placement Test Essays of Chinese Second Language Students Through a Corpus Analysis

Eunjeong Park

Department of Teaching and Learning, The Ohio State University

1945 N. High St., Columbus 43210, USA

Tel: 614-292-2461   E-mail: park.1752@osu.edu

Abstract

Learner corpora—repositories of authentic texts produced by foreign/second language learners (Granger, 2009)—have widely been used in second language (L2) research due to authenticity and practical insight. This study involves a learner corpus of placement test essays to examine L2 students' use of lexical bundles and explore their linguistic needs so that the learner corpus can be used as potential linguistic resources for L2 students with their perceived needs. 367 placement test essays from Chinese L2 students in a midwestern university were analyzed in the study. The results show the most frequent use of prepositional phrasal lexical bundles, less frequent use of discourse-organizing bundles, and frequent use of 'I' in stance bundles. This study revealed L2 students need to be aware of the use of discourse-organizing bundles for cohesion/logic in academic writing and implied the value of L2 students' lexico-grammatical use as a resource and investment with pedagogical significance.

Keywords: Learner corpus, Lexical bundles, Lexico-grammar, Corpus analysis

1. Introduction

Second language (L2) writing is a crucial skill for international L2 students who study in a foreign country in order to be successful in their academic context. It is obvious for them to learn the word and phrase combinations of disciplinary writing conventions as a fundamental step in academic settings (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Cortes, 2004; Hyland 2008a, 2008b). However, international L2 students face challenges in learning a foreign or a second language and completing academic writing tasks in college (Biber, 2006).

It is necessary for them to acquire the ability to understand complex academic discourse in English. Lexicogrammar, which is a continuum of lexis and grammar, is a system of wording and the representation of meaning making through words and structures (Halliday & Matthiessen, 2004). Lexicogrammar is considered as one of the diverse components of a language, leading applied linguists and researchers to conduct research on it. One of the features within lexicogrammar is a "lexical bundle" (Biber et al., 1999, 2004). Lexical bundles are the most frequent sequences of multiple words in a written/spoken register indicating formulaicity of lexicogrammar in a language (Biber et al., 1999). Corpus researchers (Cortes, 2004; Hyland, 2008b) maintain that the academic adaptation to the rhetorical styles and writing expectations of different disciplines is important in light of securing membership in the academic community. Therefore, it is important to understand the students' authentic use of lexical bundles so that higher educational institutions can enhance the curriculum and instruction of second language writing particularly in the programs of English as a Second Language (ESL).

The big data, such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), or the International Corpus of Learner English (ICLE), have been regarded essential in applying them to corpus-based instruction (e.g., Bychkovska & Lee, 2017; Lancaster, 2016; Neely & Cortes, 2009; Pickering & Byrd, 2008), which is valid particularly if there is no specialized corpus created in certain programs. Furthermore, a lot of corpus studies have focused on L2 students' written products in their composition classes due to high accessibility and efficiency. These corpus data are all valuable. However, few studies (e.g., Staples, Egbert, Biber, & McClair, 2013) have explored lexical bundles in high-stakes test essays as the potential resources of L2 writing.

To fill a gap in corpus research, this study aims at identifying the most frequently used four-word lexical bundles in placement test essays of international L2 students, particularly Chinese L2 students. Second, with a reference of Biber et al.'s (1999) study, structural and functional features of the lexical bundles are uncovered. Analyzing the features of the lexical bundles is crucial because the analysis provides not only the frequent use of lexical bundles in placement test essays but also their strengths and weaknesses of the lexical bundle use.

The following research questions guided this study:

1) What are the most frequently used four-word lexical bundles found in placement test essays written by Chinese L2 students?

2) What are the structural characteristics of the lexical bundles found in this study?

3) What are the functional characteristics of the lexical bundles found in this study?

4) What are the students' potential linguistic needs to improve their academic writing based on the structural and functional lexical bundle analyses?

2. Literature Review

2.1 Definition and Criteria of Lexical Bundles

Many researchers (e.g., Biber et al., 1999; Cortes, 2004; Sinclair, 1991; Stubbs & Barth, 2003) investigated the frequent use of lexical phrases in texts or corpora, and multi-word lexical

chunks exist across various registers—conversation, academic prose, and fiction (Biber et al., 1999; Stubbs & Barth, 2003). Since the work of lexical bundles in the Longman Grammar (Biber et al., 1999), Biber et al.'s (1999) lexical bundle framework has been used in a wide range of research, such as comparing different registers of textbooks and classroom discourse (e.g., Biber et al., 2004; Biber & Barbieri, 2007) and university registers (classroom management, office hours, study groups, service encounters, course management writing, and institutional writing) (Biber, 2006), the use of lexical bundles in different populations, such as native versus nonnative speakers (e.g., Ädel & Erman, 2012; Chen & Baker, 2010) and novices versus expert writers (e.g., Cortes, 2004; Hyland, 2008a), lexical bundles of academic disciplines (e.g., Cortes, 2004; Hyland, 2008b), and the comparison of lexical bundles in different languages, such as history articles written in English versus Spanish (e.g., Cortes, 2004, 2008) and a corpus of Korean texts consisting of academic prose and conversation (e.g., Kim, 2009).

Lexical bundles are defined as the most frequent sequences of words in a register (Biber, Conrad, & Cortes, 2004) and combinations of more than three words that frequently occur in a language (Biber, et al., 1999). Biber et al.'s (1999) operational definition of lexical bundles is "the combinations of words that recur most commonly in a given register" (p. 992). A lot of lexical bundle researchers have coined technical terms, such as *sentence stems* (Pawley & Syder, 1983), *chunks* (Sinclair, 1991), *prefabs* or *lexical phrases* (Nattinger & DeCarrico, 1992), *recurrent word combinations* (Altenberg, 1998; De Cock, 1998, 2000), *prefabricated patterns* (Granger, 1998), *phrasal lexemes* (Moon, 1998), *lexical bundles* (Biber & Conrad, 1999; Biber et al., 1999), *chains* (Stubbs, 2002; Stubbs & Barth, 2003), *recurrent sequences* (De Cock, 2003), *formulaic sequences* (Schmitt & Carter, 2004), *n-grams* (Stubbs, 2007a, 2007b; Henderson & Barr, 2010), *formulas* (Granger & Meunier, 2008; Wray, 2002),and *clusters* (Hyland, 2008b; Schmitt, Grandage, & Adolphs, 2004).

Lexical bundles are generally incomplete structural units (Biber & Conrad, 1999). Lexical bundles are distinguishable from idioms and collocations. Idioms usually include constant meanings in expressions and structurally complete units (e.g., *kick the bucket*, *bear in mind*); collocations are statistical associations between the two words that retain their own literal meaning (e.g., *little while*, *small amount*) (Biber & Conrad, 1999). In a sense, lexical bundles are regarded as extended collocations—a sequence of more than three words that present statistical co-occurrence in a register (Biber & Conrad, 1999). This co-occurrence can also have strong associations with grammatical patterns; so-called lexico-grammatical patterns (Biber et al., 1999). More linguists and language researchers may notice that lexico-grammatical features are essential in identifying the patterns of language use because both grammatical structures and lexical categories are recognized to "serve the same underlying communicative tasks or functions" (Biber et al., 1999, p. 13).

A set of defining criteria can build up an entity of lexical bundles. The first criterion is related to frequency. In terms of language research, frequency seems to be objective and clear-cut (Wray, 2008). Formulaic expressions must occur frequently in a register to be considered as lexical bundles. They are the most frequently occurring sequences of words in multiple texts in a register; however, frequency thresholds are somewhat arbitrary with threshold ranges between 10 and 40 instances per million words (Biber et al., 1999). Frequency cut-off points

for lexical bundles depend on the researchers' goals. For example, Biber et al. (1999) and Simpson-Vlach and Ellis (2010) determine a cut-off point of ten in one million words in their register-based research, while Biber, Conrad, & Cortes (2004) and Biber & Barbieri (2007) are more conservative in using the criteria of 40 in one million words in their search for lexical bundles and Cortes (2004) sets the cut-off point of 20 in one million words in her data. However, the actual cut-off frequency to identify lexical bundles may be arbitrary, depending on the research contexts and objectives (Sánchez Hernández 2013). The second criterion of lexical bundles is dispersion, indicating that they must recur in at least five or multiple texts in order to guard against idiosyncratic uses by individual speakers or writers (Biber, 2009; Cortes, 2006). In the ancestral study of Biber et al. (1999), lexical bundles are distributed across at least five different texts in each register. However, the minimum dispersion can vary across studies, depending on the aim and bulk of corpus data (Alipour & Zarea, 2013). The third criterion is noted that lexical bundles are not idiomatic in meaning. A lexical bundle operates as an entire unit like an idiom, but its meaning can be clearly understood from the constituents of the bundle (Biber, 1996), which is not in the nature of an idiom. Biber et al.'s (1999) study highlights the difference between idioms and lexical bundles. Idioms are mostly fixed expressions with a meaning, while lexical bundles are "the sequences of words that most commonly co-occur in a register" (Biber et al., 1999, p. 990). Idioms are possibly substituted with a single word for the sequence. However, a certain word in lexical bundles cannot be substituted. The next criterion refers to fixedness in texts. According to Cortes (2004), the fixedness is a result of the frequency determination in the bundle extraction process; thus, the fixedness of lexical bundles differs from that of other idiomatic word combinations. The fixed form within designated cut-off frequency is qualified as a lexical bundle, regardless of its idiomaticity.

Besides fixedness, lexical bundles do not embody a complete structural unit, particularly in academic writing, comprising parts of noun phrases or prepositional phrases. In other words, lexical bundles are recurrent word combinations, irrespective of idiomaticity and structural status (Biber et al., 1999, p. 990). For example, Biber et al.'s (1999) work reveals that 15 percent of the lexical bundles in conversation consist of entire grammatical clauses or phrases, while less than five percent of the lexical bundles in academic prose contain complete structural units; that is, more than 95% of the lexical bundles were not complete units. Cortes (2004) also supports the structural incompleteness of lexical bundles by stating that "[l]exical bundles are identified empirically, rather than intuitively, as word combinations that recur most commonly in a register, and therefore, lexical bundles are usually not complete structural units, but rather fragmented phrases or clauses with new fragments embedded" (p. 400). Moreover, most lexical bundles in conversation connect two clauses (e.g., *I want to know*, *well that's what I*), while lexical bundles in academic prose link two phrases (e.g., *in the case of*, *the base of the*) (Biber et al., 1999).

2.2 Characteristics of Lexical Bundles

2.2.1 Structural Characteristics of Lexical Bundles

Most lexical bundles do not characterize complete structural units, but they can be categorized according to the structural correlates (Biber et al., 1999). Structural

characteristics mainly hold noun phrase (NP), verb phrase (VP), prepositional phrase (PP), and dependent clause fragments. According to Biber et al. (1999), NP-based patterns include noun phrases with or without post modifier fragments and quantifier expressions; PP-based patterns are prepositional phrase fragments along with embedded of-phrase fragments; VP-based patterns include verb phrase fragments and [auxiliary + copular be]; and clausal patterns involve adverbial clause fragments, *wh-* and *to-*clauses.

2.2.2 Functional Characteristics of Lexical Bundles

Not only structural but also functional characteristics of lexical bundles are significant in better understanding the distribution of lexical bundles. Biber (2006) and Biber et al. (1999) emphasize three primary functions of lexical bundles in the registers: (1) stance bundles, (2) discourse organizers, and (3) referential bundles. According to them, stance bundles represent attitudes or assessments of certainty of the propositions (e.g., *it is necessary to, I don't think so*); discourse organizers manifest relationships between preceding and forthcoming discourse (e.g., *on the other hand, when it comes to*); referential bundles make a direct reference to the textual context to identify a particular attribute of the entity (e.g., *one of the things, in terms of the*).

3. Method

3.1 Corpus Data Collection

The Corpus of Learner English (CoLE), generated by a corpus team in the ESL program at Research One university, is the central corpus data of Chinese L2 students' placement test essays. The written placement test consists of two parts: a summary and a response of an assigned source. Students read and respond to a source— scientific research articles for graduate students, and research articles of a broader nature, topic-wise, for undergraduate students. The placement test essays measured how effectively the students summarize and critically think about the source as well as their use of proper documentation styles and citation techniques.

The students who took placement compositions at the testing center were given consent to having their essays collected into the corpus repository. Demographic information (i.e., age, gender, country, college, grad/undergrad, and a placement test level) was included in each composition. Placement test compositions were only included with students' consent, and they did not contain any identifying information about the author. Risks to students were minimized by de-identifying the data in each essay before it was uploaded into the corpus. Exclusion criteria involved something personally identifiable (e.g., names and family background), non-consenters' products, and the ones with below-thresholds.

The corpus was further subdivided into three proficiency levels (low, medium, and high), based on a range of placement test essay scores. The 367 compositions in the CoLE included all Chinese students (215 males and 152 females) with their consent. 192 undergraduate (52.3%) and 175 graduate (47.7%) students were from three Colleges of Arts and Sciences (135 undergrad/graduate students), Engineering (135 undergrad/graduate students), and Business (97 undergrad/graduate students). 314 students (85.6%) were at the age of 18-21, and 53 students (14.4%) were over 22 years old. 121 undergraduates received the low level;

71 undergraduates received the intermediate level; 81 graduates received the low level; 94 graduates received the intermediate level of the placement essay test.

3.2 Corpus Data Analysis

A list of the four-word lexical bundles was generated through AntConc software (Version 3.5.7.) (Anthony, 2018)—a corpus-based analytic tool. The use of frequency counts is the most straightforward approach for quantitative data by classifying linguistic items (McEnery & Wilson, 2001). The students' needs for improving lexico-grammatical academic writing were explored based on the structural and functional lexical bundle concordance analyses.

3.2.1 The Validity of the CoLE Lexical Bundle List

For content validity, the list of 39 lexical bundles was reviewed by a panel of experts (i.e, the CoLE team) whose academic background is applied linguistics and L2 composition if the lexical bundle items are valid in the study. Then, they were revised based on the experts' comment. The decision on the inclusion of these lexical bundles were also made by literature (Biber et al., 1999; Byrd & Coxhead, 2010; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).

3.2.2 The Reliability of the CoLE Lexical Bundle List

The inter-rater reliability of the CoLE lexical bundle list was estimated on the total items ($N =$ 39) with a CoLE team who can inspect each lexical bundle with structural and functional characteristics. Cohen's Kappa statistics in SPSS (Version 24, 2017) yielded .869 for the 39 items in Table 1. The reliability statistics range indicates a high level of inter-rater reliability with the interpretation of Kappa .81-1.00: almost perfect agreement (McHugh, 2012). In other words, regarding reliability, the results of the current study can be replicated consistently if other researchers examine the structural and functional characteristics of the same lexical bundles.

Table 1. Symmetric measures of inter-rater reliability statistics

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	-.026	.019	-.164	.869

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

4. Results

This section discusses the frequency-based lexical bundle list, structural and functional characteristics of the CoLE lexical bundles, and the students' needs to improve lexico-grammatical writing skills based on the results of lexical bundles.

4.1 The Frequency-Based Four-word CoLE Lexical Bundles

Research Question 1: *What are the most frequently used four-word lexical bundles found in placement test essays written by Chinese L2 students?*

Grounded on Biber et al.'s (1999) lexical bundle project, 4-word lexical bundles are identified for the corpus analysis because 4-word lexical bundles are used more commonly than 5-word ones and provide a more distinct range of structures and functions than 3-word lexical bundles (Hyland, 2008a), holding 3-word bundles in the structure of 4-word strings (Cortes, 2004). In line with the previous studies (Byrd & Coxhead, 2010; Salazar, 2014), a list of 39 lexical bundles is generated from the CoLE composed of 367 placement test essays with 158,390 tokens/words (*M*=431.58, *SD*=111.91) in Table 2. According to Biber et al. (1999), frequency thresholds are arbitrary, ranging between 10 and 40 instances per million words. Since the CoLE contains a small number of words (i.e., 158,390 tokens), the frequency of ten was chosen as the cut-off point for the lexical bundle list. The raw frequency was normalized per 100,000 words in order to compare one corpus with another.

Table 2. The summary of lexical bundles in the CoLE

Rank	Lexical bundles	Function	Structure	Raw Freq	Normalized Freq.(/100,000)
1	I agree with the	Stance-epistemic	C	51	32.20
2	as far as I	Stance-epistemic	C	47	29.67
3	for a long time	Referential	PP	38	24.00
4	on the other hand	Discourse organizer	PP	35	22.10
5	are more likely to	Stance-epistemic	VP	32	20.20
6	in the United States	Referential	PP	29	18.31
7	I totally agree with	Stance-epistemic	C	27	17.05
8	my point of view	Stance-epistemic	NP	27	17.05
9	at the same time	Referential	PP	22	13.89
10	as an academic major	Topic-specific	PP	16	10.10
11	eating organic food is	Topic-specific	NP	15	9.47
12	in my opinion I	Stance-epistemic	PP	15	9.47
13	last but not least	Discourse organizer	O	15	9.47
14	when I was in	Referential	C	15	9.47
15	with the development of	Referential	PP	15	9.47
16	a lot of time	Referential	NP	14	8.84
17	I think it is	Stance-epistemic	C	14	8.84

18	it is true that	Stance-attitudinal	VP	14	8.84
19	cooperate with the intelligence	Topic-specific	VP	13	8.21
20	from my perspective I	Stance-epistemic	PP	13	8.21
21	in this article the	Discourse organizer	PP	13	8.21
22	a large number of	Referential	NP	12	7.58
23	as we all know	Stance-epistemic	C	12	7.58
24	first of all the	Discourse organizer	O	12	7.58
25	the author believes that	Stance-epistemic	C	12	7.58
26	when it comes to	Discourse organizer	C	12	7.58
27	a lot of things	Referential	NP	11	6.94
28	in our daily life	Referential	PP	11	6.94
29	is no doubt that	Stance-attitudinal	VP	11	6.94
30	is one of the	Referential	VP	11	6.94
31	is the most important	Stance-attitudinal	VP	11	6.94
32	when they are young	Referential	C	11	6.94
33	a strong relationship with	Referential	NP	10	6.31
34	according to the passage	Stance-epistemic	PP	10	6.31
35	as for me I	Stance-epistemic	PP	10	6.31
36	IU immigration bridge program	Topic-specific	NP	10	6.31
37	more than three hours	Referential	O	10	6.31
38	sitting for a long	Topic-specific	NP	10	6.31
39	when I was a	Referential	C	10	6.31
	Total word tokens: 158,390			676	

Note: C (clause), NP (noun phrase), PP (prepositional phrase), VP (verb phrase), & O (other expressions).

4.2 The Structural Characteristics of the CoLE Lexical Bundles

Research Question 2: *What are the structural characteristics of the lexical bundles found in this study?*

Lexical bundles usually comprise syntactic fragments instead of complete structural units. This has been substantiated through several studies (e.g., Biber, Conrad, & Cortes, 2004; Byrd & Coxhead, 2010; Hyland, 2008b; Salazar, 2011, 2014; Simpson-Vlach & Ellis, 2010). Based on Biber et al.'s (1999) structural taxonomy of lexical bundles, the CoLE lexical bundles are structurally placed in the following categories: NP-, PP-, VP-based phrases and

clausal fragments.

4.2.1 Noun Phrases/Structures (NP)

The term "noun structures" means noun-headed phrases including determiner + (premodification) + head noun + (postmodification and complementation) (Biber et al., 1999). Noun structures found in CoLE include eight types (20.5%) with 109 tokens (16.2%), and this result does not coincide with previous lexical bundle studies, such as 32% in Biber et al.'s (1999) study and 34% in Salazar's (2014) study of noun structures. These studies concluded that the noun phrase is the most common structure in the lexical bundle list, which was not the case for the CoLE list. The eight noun structures identified are used to denote qualities (e.g. *my point of view*) and relations (e.g. *a strong relationship with*); to indicate quantities (e.g., *a large number of, a lot of time, a lot of things*), actions (e.g., *eating organic food is, sitting for a long*), and content (e.g. *IU immigration bridge program*).

4.2.2 Prepositional Phrases (PP) Fragments

The Prepositional phrase fragments in this context consist of a preposition and a complement in the form of a noun phrase (Biber et al., 1999). A variety of prepositional phrase fragments are included in the CoLE list. Twelve types (30.8 %) with 227 tokens (33.5 %) fall within several categories: impersonal epistemic stance (e.g., *according to the passage, as an academic major, in this article the*), personal epistemic stance (e.g., *as for me I, from my perspective I, in my opinion I*), topic elaboration (e.g., *on the other hand*), temporal reference (e.g., *at the same time, for a long time, in our daily life*), spatial reference (e.g., *in the United States*), and the intangible framing attribute (e.g., *with the development of*). The results confirm Biber et al.'s (1999) analysis that the actual structure of academic lexical bundles (70 % of all bundles) is preposition + noun phrase, or noun phrase + *of*, or anticipatory *it* fragments. The results also revealed that their structural use of noun phrases in the CoLE is similar to the one in the argumentative writing of native speakers of English (Bychkovska & Lee, 2017).

4.2.3 Verb Phrases/Structures (VP)

Verb structures in this context means verb-headed phrases accompanied by one or more auxiliaries specifying the action, state, or process (Biber et al., 1999). Verb structures involve six different types (15.4 %) with 92 tokens (13.6 %) in the CoLE list. There is one type of anticipatory *it* + adjective phrase (e.g. *it is true that*). This type is considered as an extraposed structure with a predicative adjective that indicates the stance of the writer, such as possibility and importance (Biber et al., 1999). The major verb structure is copula be + noun/adjective phrase (e.g., *are more likely to, is one of the*). According to Biber et al. (1999), these bundles play a role as subject predicative to the copula *be*. Interestingly, the examples of '*are more likely to*' and '*is no doubt that*' signify the stance of the writer.

Salazar's (2014) study revealed that the main verb structure was the passive verb followed by a prepositional phrase. Unlike her study, the current study's data does not show any expression of the passive verb structure among the 4-word lexical bundles.

4.2.4 Clausal Fragments

Clausal fragments in this context are defined as a structure consisting of an independent clause together with dependent clauses (Biber et al., 1999). Ten types (25.6%) with 211 tokens (31.2%) are shown in the CoLE list. As Biber et al.'s (1999) discussion of the adverbial clause fragment points out, lexical bundles with the subordinator 'as' are yielded for deictic reference to other main clauses (e.g., *as far as I, as we all know*). The other adverbial clause fragments involve temporal reference (e.g., *when I was a, when I was in, when it comes to, when they are young*). One type of (verb phrase +) *that*-clause fragment is the lexical bundle including a main clause verb followed by a *that*-clause (e.g., *the author believes that*).

4.2.5 Other Expressions

There are also a few lexical bundles in the list that do not fit into any of the previously reported categories. These include three types (7.7%) with 37 tokens (5.5%) (e.g., *first of all the, last but not least, more than three hours*) and have the features of adverbial phrases.

Table 3 and Figures 1 and 2 provide a detailed breakdown of results relative to Research Question 2.

Table 3. The Structural Classification of Lexical Bundles in the CoLE

Structural patterns in the CoLE	Type	%	Token	%	Examples
A. Phrasal					
1. NP-based					
Noun phrase with *of*-phrase fragments	4	10.2	64	9.5	*my point of view, a large number of, a lot of things, a lot of time*
Noun phrase with post modifier	1	2.6	10	1.5	*a strong relationship with*
(connector +) NP with of- phrase	-	-	-	-	
Other noun phrases	3	7.7	35	5.2	*eating organic food is, sitting for a long, IU immigration bridge program*
2. PP-based					
Prepositional phrase fragments	11	28.2	212	31.3	*according to the passage, as an academic major, as for me I, at the same time, for a long time, from my perspective I, in my opinion I, in our daily life, in the United States, in this article the,*

|					
					on the other hand
PP with embedded *of*-phrase fragment	1	2.6	15	2.2	*with the development of*
3. VP-based					
Pronoun/NP (+ auxiliary) + copula be	-	-	-	-	
Anticipatory it + VP/-adjective phrase	1	2.6	14	2.1	*it is true that*
Copula be + NP/adjective phrase	4	10.2	65	9.6	*are more likely, to, is no doubt that, is the most important,* *is one of the,*
Passive verb + PP fragments	-	-	-	-	
Other verb phrases	1	2.6	13	1.9	*cooperate with the intelligence*
B. Clausal					
Adverbial clause fragments	6	15.3	107	15.8	*as far as I, as we all know,* *when I was a, when I was in, when it comes to, when they are young*
(VP +) *that*-clause fragment	1	2.6	12	1.8	*the author believes that*
(verb/adjective+) *to*-clause fragments	-	-	-	-	
Personal pronoun + lexical verb phrase	3	7.7	92	13.6	*I agree with the, I totally agree with, I think it is*
C. Other expressions	3	7.7	37	5.5	*first of all the, last but not least, more than three hours*
Total	39	100	676	100	
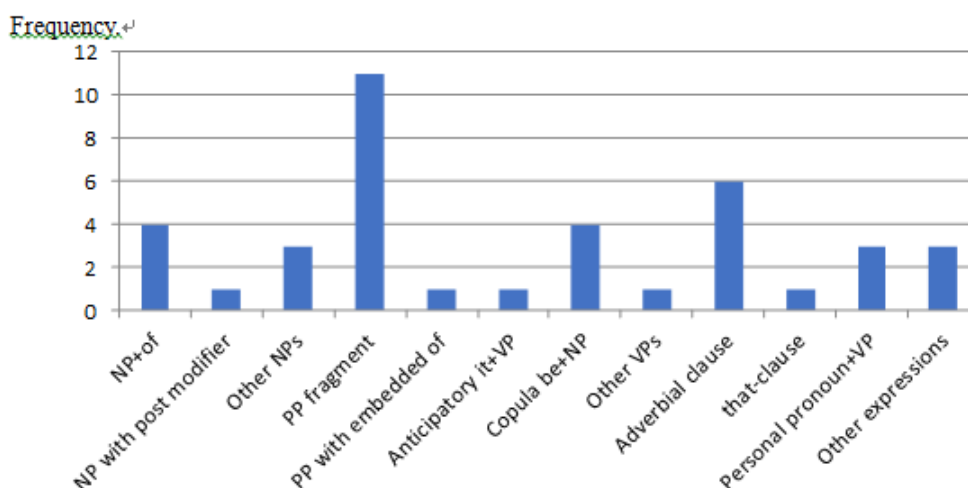
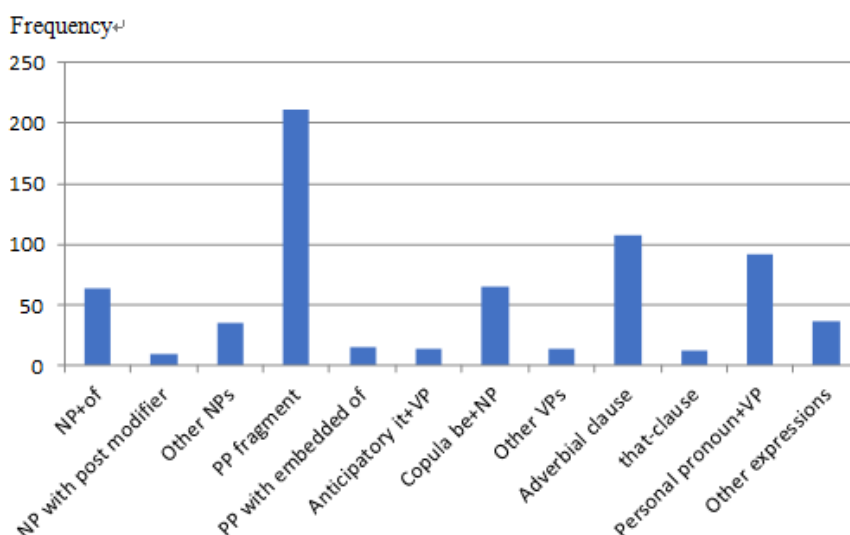Figure 1. The distribution of structural types

Figure 2. The distribution of structural tokens

4.3 The Functional Characteristics of the CoLE Lexical Bundles

Research Question 3: *What are the functional characteristics of the lexical bundles found in this study?*

As noted earlier, the CoLE list yields 39 four-word lexical bundles. In terms of functions performed, these include:

- 15 stance bundles (38.5%) with 306 tokens (45.2%)

- 14 referential bundles (35.9%) with 219 tokens (32.4%)

- 5 discourse organizing bundles (12.8%) with 87 tokens (12.9%)

- 5 topic-specific (12.8%) lexical bundles with 64 tokens (9.5%)

Among the 15 stance bundles, there are 12 epistemic and three attitudinal stance bundles in

the CoLE list. Most epistemic stance bundles (nine types) are personal with the proposition of certainty or probability. For example, '*I think it is*' and '*the author believes that*' express some degree of certainty. Impersonal epistemic stance bundles (e.g., *according to the passage, are more likely to*) also show the extent of certainty, as depicted below in the following CoLE examples:

(1) As for me, *I think it is* a good way for universities to help graduate students to remain in the USA for the following reasons.

(2) However, *the author believes that* people should work in their own ways.

(3) *According to the passage*, the author points out that the lifestyle in young age can impact the brain development when they get older.

(4) *According to this*, it seems that narcissistic women *are more likely to* use Facebook as a reflecting pool than narcissistic males do.

Attitudinal stance bundles express speaker attitudes towards the actions or events in the text. Biber et al. (2004) categorized them relative to four aspects: desire, obligation/directive, intention/prediction, and ability. Among them, the three lexical bundles below (e.g., *is no doubt that, is the most important, it is true that*) from the CoLE list include directive functions, such as these in CoLE:

(5) There *is no doubt that* everyone has the experience of feeling regret.

(6) The language *is the most important* factor when they choose which country to study.

(7) *It is true that* parents have a great impact during our growth period.

Discourse organizing bundles involve two functions: topic introduction and topic elaboration/clarification. Among the six discourse organizing bundles in CoLE, only one type serves topic introduction; the others play a role in topic elaboration. Here are some examples from CoLE:

(8) *In this article, the* author mainly talks about egret. (topic introduction)

(9) *On the other hand*, those organizations can ask mathematicians to do some research on the crime rate and the natural disaster. (topic elaboration)

(10) However, *when it comes to* abortion, every single step should be considered with great caution. (topic elaboration)

Fourteen referential bundles in CoLE also produce different categories, including: quantity, framing attributes, and time/place references. Some examples from CoLE include:

(11) The second factor is that *a large number of* citizens may have trouble in some fields. (quantity)

(12) Cost *is one of the* nonacademic barriers. (framing attribute)

(13) They have to protect their children and *at the same time*, they should also give their children personal space and time. (time reference)

Topic-specific lexical bundles found in CoLE focus on the literal content per se (e.g., *cooperate with the intelligence, eating organic food is*). Therefore, they do not serve any function in the text, but they may strengthen the text in a repetitive way.

Table 4 and Figures 3 and 4 provide a detailed breakdown of the results just reported:

Table 4. The functional categories of lexical bundles in the CoLE

| Functional categories	Type	%	Token	%	Examples
A. Stance					
Epistemic	12	30.8	270	39.9	*according to the passage, are more likely to, as far as I, as for me I, as we all know, from my perspective I, I agree with the, I totally agree with, I think it is, in my opinion I, my point of view, the author believes that,*
Attitudinal	3	7.7	36	5.3	*is no doubt that, is the most important, it is true that*
B. Discourse-organizing					
Topic introduction	1	2.6	13	1.9	*in this article the*
Topic elaboration	4	10.2	74	11.0	*first of all the, last but not least, on the other hand, when it comes to*
C. Referential					
Quantity	3	7.7	37	5.5	*a large number of, a lot of things, a lot of time*
Framing attributes	3	7.7	36	5.3	*A strong relationship with, is one of the, with the development of*
Temporal	6	15.3	102	15.1	*at the same time, for a long time, more than three hours, in our daily life, when I was a, when they are young*
Spatial	1	2.6	29	4.3	*in the United States*
Multi-functional	1	2.6	15	2.2	*when I was in*
D. Topic-specific	5	12.8	64	9.5	*as an academic major, cooperate with the intelligence, eating organic food is, IU immigration bridge program, sitting for a long*
Total	39	100	676	100	

Frequency
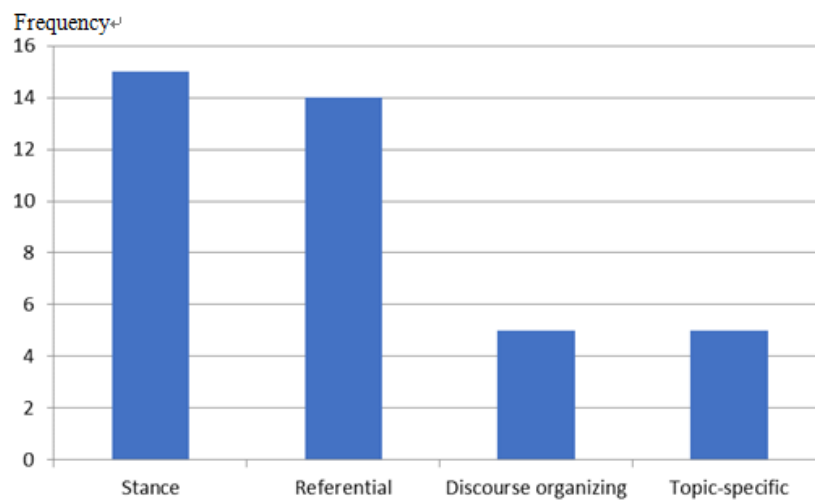
Figure 3. The distribution of stance, referential, discourse organizing, and topic-specific types
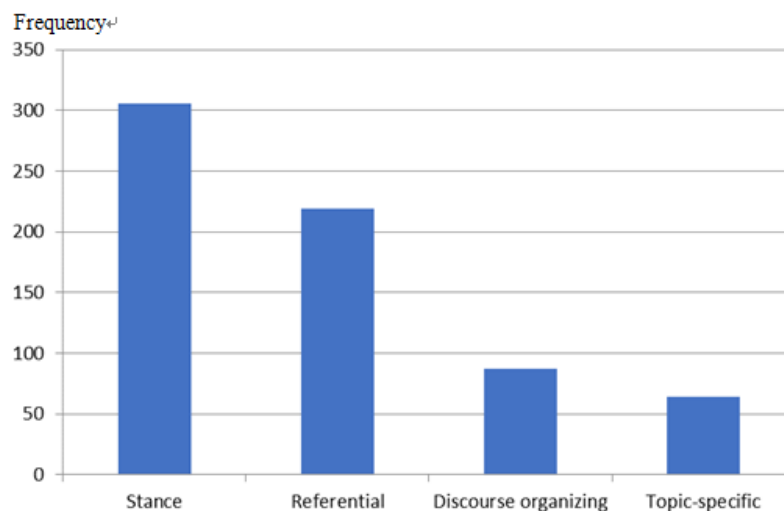
Frequency

Figure 4. The distribution of stance, referential, discourse organizing, and topic-specific tokens

4.4 The Students' Potential Linguistic Needs for Improving Lexico-Grammatical Academic Writing Based on the Lexical Bundle Concordance Analysis

Research Question 4: *What are the students' potential linguistic needs to improve their academic writing based on the structural and functional lexical bundle analyses?*

A needs-based approach (e.g., Braun, 2007; Jones & Schmitt, 2010; Yoon & Jo, 2014) is operationalized as the instruction modified based on the students' potential linguistic needs. The concordance analysis revealed the students' limitations of the lexical bundle use (i.e., relatively under-used lexical bundles). The under-used lexical bundles indicate a less functional use of stance, discourse-organizing, and referential bundles. They were examined by comparing the structural and functional features of the CoLE learner corpus with the reference corpora from Simpson-Vlach and Ellis (2010) and Hyland's (2012) studies.

Quantitative descriptive statistics (i.e., frequency & percentage) and qualitative analytic judgments play a major role in responding to this research question.

As for structural characteristics of lexical bundles, due to the small size of the CoLE lexical bundle list, no typical pattern was discovered. However, as indicated earlier, the results are not commensurate with previous research which indicates that the view of academic writing is "noun-centric" (Pickering & Byrd, 2008, p. 122). For example, Salazar (2014) pointed out that variations of the most frequently used frame are *the ___ of* and *a ___ of* in her study about native and nonnative scientific writing. It would be useful for international L2 students to know various expressions of the noun fragment with *of*-phrase (e.g., *the ___ of & a ___ of*). The current study with the CoLE also revealed that prepositional lexical bundles with 12 types (30.8 %) and 227 tokens (33.5 %) appeared the most in academic writing, as also seen in previous research (e.g., Biber et al., 1999; Pan et al., 2016), which shows international students should also be exposed to these prepositional lexical bundles.

Functional characteristics of lexical bundles were revealed through the judgment of the "relative frequency" (Bonk, 2001, p. 118) of the lexical bundles. The results regarding functional characteristics in the current study showed less functional use of discourse organizing lexical bundles (87 tokens, 12.8 %), compared to the functional use of stance (306 tokens, 45.2 %) and referential (219 tokens, 32.4 %) lexical bundles, as under-used lexical bundles (i.e., comparatively less frequently used lexical bundles in terms of functions). Several studies (i.e., Chen & baker, 2010; Hyland, 2012; Simpson, 2004; Simpson-Vlach & Ellis, 2010) discussed the importance of using discourse organizing lexical bundles in academic writing. Simpson (2004) highlighted the important function of discourse organizing bundles to summarize, sequence, and focus information by analyzing the Michigan Corpus of Academic Spoken English (MICASE). Chen and Baker's (2010) study noted that the student writing contained more discourse organizing bundles than the published academic work. Simpson-Vlach & Ellis' (2010) study showed the top ten and bottom ten lexical bundles of the academic writing items by frequency and mutual information (MI) and found that 'on the other hand' by frequency and 'due to the fact that' by MI, which are categorized as discourse organizing lexical bundles, are the top ten phrases of the original academic writing items. Hyland's (2012) study also presented the most frequent three-, four-, and five-word bundles in academic articles and theses, and the top-ranked four-word bundle was 'on the other hand', categorized as a discourse organizing lexical bundle. Based on the reference corpora from Simpson-Vlach & Ellis (2010) and Hyland's (2012) studies, the results of functional characteristics in the current study revealed relatively less use of discourse organizing lexical bundles (87 tokens, 12.8 %) according to the judgment of the relative frequency (Bonk 2001), which provide insights into students' potential linguistic needs regarding the use of lexical bundles.

The corpus results of CoLE in the current study also showed frequent use of self-mentions with the pronoun, 'I', in the outcome of stance bundles. In CoLE, stance bundles were used in the placement essays with 38.5 % of type and 45.2 % of tokens, indicating that self-mentions with the pronoun "I" were used extensively. Among 12 types of stance bundles, seven lexical bundles (e.g., *as far as I, as for me I, from my perspective I, I agree with the, I totally agree with, I think it is, in my opinion I*) contained 'I' to address the view of the authors in the CoLE.

Hyland (2005a) categorized lexical bundles with 'I' as self-mentions in interactional metadiscourse. Gilquin & Paquot (2008) found that personal structures (using pronouns 'I') are used more in L2 learners' papers than in ones of professional writers or native speakers of English. The results from the analysis of lexical bundles (especially seven bundles with "I" out of 12 types of stance bundles) in the current study revealed that L2 students may need to express their ideas or attitudes with a variety of lexical bundle expressions and sophistication. Therefore, it would be beneficial for L2 students to know different stance bundles including impersonal structures and use them in their academic papers.

5. Discussion

Structural and functional characteristics of the lexical bundle use were examined through the lexical bundle approach. In terms of structural characteristics, the prepositional phrasal lexical bundles were most used in Chinese L2 students' placement test essays. The results of this study are not aligned with some of the prior literature (e.g., Biber et al., 1999 (Note 1); Salazar, 2014) that noun phrasal lexical bundles are the most common structures in their academic lexical bundle lists. Yet, the results of types of prepositional phrases, such as preposition + noun phrase, noun phrase + *of*, and anticipatory *it* fragments are major items, confirm Biber et al.'s (1999) lexical bundle analysis.

For functional characteristic, two main issues emerged from the analysis and interpretation of the functional characteristics in the placement test essays: (1) relatively less frequently used discourse organizing lexical bundles, compared to stance and referential bundles in the learner corpus, and (2) frequent use of 'I' in the stance bundles. First, the results of the Corpus of Learner English (CoLE) list in this study contained relatively less frequency of discourse organizing lexical bundles, compared to stance and referential bundles, as opposed to the prior research (e.g., Chen & baker, 2010; Hyland, 2012; Simpson-Vlach & Ellis, 2010). Discourse organizing lexical bundles are particularly important to make relationships between the preceding and forthcoming discourse in texts. Therefore, in terms of logic and cohesion, it is recommended to use discourse organizing bundles appropriately in academic writing. This may lead to potential research to investigate the functional use of lexical bundles.

Second, Chinese L2 students primarily used stance bundles with first person singular pronouns. Stance bundles were used well in the placement essays with 38.5 % of type and 45.2 % of tokens. Among 12 types of stance bundles, however, it is noticeable that seven lexical bundles (e.g., *as far as I, as for me I, from my perspective I, I agree with the, I totally agree with, I think it is, in my opinion I*) contain 'I' to address the writer's perspective. Gilquin and Paquot (2008) state that personal structures (using pronouns 'I') are used more in L2 learners' papers than in ones of professional writers or native speakers of English. Professional native writers express their attitude towards the proposition in academic writing by using impersonal structures, such as *this article examines the* or *it is reasonable to*; whereas, L2 learners tend to "make their presence strongly felt in their writing" (Gilquin & Paquot, 2008, p. 48). L2 learners often tend to be too visible with the expressions *I think that* or *I would like to*, which are unnecessary in academic writing, while professional native writers prefer more impersonal phrases, less dependent on 'I' (Gilquin & Paquot, 2008). The

results from the analysis of lexical bundles reveal that L2 students need to express their ideas or attitudes with a variety of lexical bundle expressions and sophistication.

Some concerns occurred in the process of constructing the corpus data of the L2 students' placement test essays. One was why L2 learners' writing should be utilized for the instructional materials in ESL writing classes. Some people seemingly consider on-going learners' language as deficiency (i.e., interlanguage) and are hesitant to utilize their data. However, we argue that we should value and look at their product as a resource and investment. Furthermore, the list of the lexical bundles generated from the learner corpus includes all correct usage of them by checking with multiple professional writers and experts (at least four panels). Hence, it is of significance to explore L2 learners' language and apply it to pedagogical and instructional materials with authenticity.

The other concern was about the encouragement of lexical bundles in academic writing. Some studies (e.g., Staples et al., 2013) indicated that less proficient L2 learners use familiar lexical bundles repeatedly, while more proficient and native-like students can control the language use in that they tend not to rely on or use formulaic expressions. As Byrd & Coxhead (2010) regard lexical bundles as "just one of a variety different types of prefabricated and often repeated language" (p. 47), therefore, the use of lexical bundles may grant a small portion of writing proficiency. However, some other studies (e.g., Chen & Baker, 2010) showed that published academic writing exhibited the widest range of lexical bundles; whereas, L2 students' writing showed the smallest range. In other words, L2 students still need more linguistic components, and lexical bundles will benefit L2 writers by considering L2 writing based on linguistic components (i.e., lexical bundles) and helping them to be a good writer. Therefore, it is essential to consider the potential linguistic needs from the students' perspectives.

On the whole, the essence of L2 writing instruction is to provide linguistic resources that are beneficial to develop their academic writing skills with lexico-grammatical language knowledge. It is still crucial to instruct lexical bundles with high frequency in academic papers especially because the target participants in this study were novice language learners, producing errors of the use of the lexical bundles in their writing. In this sense, we should teach effective and appropriate use of academic lexical bundles with learner corpus-aided resources to L2 students with low-intermediate language levels in order to boost the prospect of being eligible members in an academic community.

6. Conclusion

Learner corpora play an important role as a beneficial resource for L2 learners to gain lexico-grammatical knowledge and potentially improve their academic writing. In particular, novice L2 writers must learn the discourse conventions (i.e., lexical bundles) of academic writing with adequate use of lexical bundles (Biber et al., 2013; Cortes, 2004). Lexical bundles are advantageous in L2 writing instruction with frequency, range, learnability, and usefulness in academic writing (Byrd & Coxhead, 2010; Nation, 2001). Furthermore, the utility of lexical bundles is promising as a basis for material design and curriculum development, creating a great impact on revealing the representativeness of the English language used by second language students (Byrd & Coxhead, 2010; Gilquin & Paquot, 2008;

Simpson-Vlach & Ellis, 2010). Learner corpora play an important role in selecting and structuring teaching contents (Granger, 2002). Lexical bundles have the potential for exploring phraseological differences between registers and disciplines (Römer, 2009). Hence, applied linguists and L2 writing researchers should strive to study L2 students' linguistic strengths and challenges in their learner corpora.

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*, 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Alipour, M., & Zarea, M. (2013). A disciplinary study of lexical bundles: The case of native versus non-native corpora. *Taiwan International ESP Journal, 5*(2), 1-20.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 99-122). Oxford: Oxford University Press.

Anthony, L. 2018. AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics, 1*(2), 171-197. https://doi.org/10.1075/ijcl.1.2.02bib

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam: J. Benjamins. https://doi.org/10.1075/scl.23

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 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. (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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Pearson.

Boers, F., Eykmans, 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

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*, 31-64.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language*

Learning & Technology, 14(2), 30-49.

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*, 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*, 391-406. https://doi.org/10.1016/j.linged.2007.02.001

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*, 33-43. https://doi.org/10.1016/j.jeap.2012.11.002

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59-80. https://doi.org/10.1075/ijcl.3.1.04dec

De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair, & M. Hundt (Eds.), *Corpus linguistics and linguistic theory* (pp. 51-68). Amsterdam: Rodopi.

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41-61. https://doi.org/10.1075/etc.1.1.05gil

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Ed.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Philadelphia, PA: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.6.04gra

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective.* Amsterdam: John Benjamins. https://doi.org/10.1075/z.139

Halliday, M. A. K., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd Ed.). London: Arnold.

Henderson, A., & Barr, R. (2010). Comparing indicators of authorial stance in psychology students' writing and published research articles. *Journal of Writing Research, 2*(2), 245-264. https://doi.org/10.17239/jowr-2010.02.02.8

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

Hyland, K. (2008b). 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

Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora, 4*(2), 135-165. https://doi.org/10.3366/E1749503209000288

Lancaster, Z. (2016). Do academics really write this way? A corpus investigation of moves and templates in "They Say/I Say." *The National Council of Teachers of English, 67*(3), 437-464.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282. https://doi.org/10.11613/BM.2012.031

Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Oxford University Press.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nattinger, J. R., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Neely, E., & Cortes, V. (2009). A little bit about: Analyzing and teaching lexical bundles in academic lectures. *Language Value, 1*(1), 17-38.

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

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.

Pickering, L., & Byrd, P. (2008). An investigation of relationships between spoken and written academic English: Lexical bundles in the AWL and in MICASE. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 110-132)*. Ann Arbor: University of Michigan Press.

Römer, U. (2009). English in academia: Does nativeness matter? *International Journal of English Studies, 20*(2), 89-100.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing*. Amsterdam, Philadelphia: John Benjamins.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.9

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127-151). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.9.08sch

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In T. Upton, & U. Connor (Eds.), *Discourse in the Professions: Perspectives from Corpus* Linguistics (pp. 37-64)*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.16.03sim

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English.

International Journal of Corpus Linguistics, 7(2), 215-244. https://doi.org/10.1075/ijcl.7.2.04stu

Stubbs, M. (2007a). Quantitative data on multi-word sequences in English: The case of the word "world". In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert (Eds.), *Text, discourse and corpora: Theory and analysis* (pp. 163−189). London: Continuum.

Stubbs, M. (2007b). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years on* (pp. 89-105). Amsterdam: Rodopi.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language, 10*(1), 61-104. https://doi.org/10.1075/fol.10.1.04stu

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*, 214-225. https://doi.org/10.1016/j.jeap.2013.05.002

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Note 1. Biber et al.'s (1999) study pointed out over 60% of all lexical bundles are noun phrasal (30%) or prepositional phrasal (33%) lexical bundles in academic prose.