



Exploring lexical bundles in recent published papers in the field of applied linguistics

Peyman Nasrabady , Majid Elahi Shirvan  and Seyed Ehsan Golparvar 

Department of Foreign Languages, University of Bojnord, Bojnord, Iran

ABSTRACT

Multiword units can contribute significantly to the fluency of linguistic production. Lexical bundles, as a type of multiword units, are sequences of three or more words used with high frequency in natural discourse. Primary studies regarding the introduction of structural and functional classifications of lexical bundles have been followed by other researchers to compare them across disciplines based on the primary classifications. However, no comprehensive study has been conducted to explore and investigate the structural as well as the functional characteristics of the lexical bundles in research articles in the field of applied linguistics. This study aimed at exploring lexical bundles used in a substantial number of academic papers published in this specific field through different journals since 2008 to identify all the lexical bundles and to capture all the possible structural and functional characteristics they express. To achieve this, a corpus of papers published in the field of applied linguistics was compiled and investigated. A considerable number of lexical bundles identified were similar to those previously found in the literature while a reasonable number of them expressed new structural and functional characteristics. The study concludes with theoretical as well as some pedagogical implications of the findings.

ARTICLE HISTORY

Received 8 February 2019

Accepted 11 June 2020

KEYWORDS

Lexical bundles; structural characteristics; functional characteristics; applied linguistics research articles; corpora

1. Introduction

Corpus linguistics provides empirical data, embodies statistical measures, and deals with large amounts of linguistic data (Biber, Conrad, and Reppen 1994; Granger 1998; Gries 2010). Among analytical processes that result from corpus linguistic studies, one can refer to wordlists and frequencies, lexical variation (type/token ratio), concordance lines, collocations, and lexical bundles (Biber et al. 1999; Coxhead 2000; Granger 2000; Sinclair 1991). Frequency is a basic feature of this kind of investigation; however, a corpus-based study does not simply count linguistic features, but also involves qualitative interpretations of numerical data as well. The goal of corpus-based research, according to Biber, Conrad, and Reppen (1998), is not only to report quantitative linguistic data but also to bring to light patterns of language use through the analysis of language data.

Moreover, Sinclair (1991) highlighted the need to consider fuzzy boundaries of rules in word associations in language which may oppose the informants' intuition. He further argued in favor of idiom principle to account for the unrandomness of word associations in language. According to Sinclair (1991), idiom principle refers to the use of prefabricated word sequences that are available to language users for use in speaking or writing without considering grammatical issues. This has been followed by researchers considering language beyond the word level through linguistic input and take collocations and multiword units into account (see Cortes 2013; Granger and Meunier 2008; Ward 2007; Wray 2000, 2008; Wray and Perkins 2000).

Multi-word expressions are believed to be a source of difficulty for non-native users (De Cock 2002; Granger 1998) and necessary to achieve native-like competence and fluency to the extent that they are important aspects in language teaching and learning (Howarth 1998; O'keeffe, McCarthy, and Carter 2007). Scientific writing in which authors need to produce cohesive and coherent language to communicate their ideas is not an exception. Gledhill (2000), for example, argues for the notion of "phraseological accent" which is put on conventionally by the widespread use of formulaic constructions in a specific discourse community and is a salient feature of technical writing, though unusual in general English. This trend in linguistic investigation involves examining formulaic language.

Formulaic language is defined as sequences of words "stored and retrieved whole from the memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray and Perkins 2000, 1). They are produced automatically, and unconsciously and have been investigated for their pragmatic functions in various contexts (Altenberg 1998; Nattinger and DeCarrico 1992). Formulaic language has also been studied for its variation across disciplines. Investigating the use of formulaic sequences, namely lexical bundles, Kashiha and Heng (2014) found that the "desire" lexical bundles under the category of stance expressions are used differently by academic lecturers in the fields of politics and chemistry. According to their findings, the lexical bundle *I'd like to* or *I would like to* was used for instructions purposes in the context of chemistry lectures; however, the same bundle functioned as a point of departure in most cases in the lectures in the field of politics. Such highly frequent word combinations have been primarily investigated as lexical bundles in a chapter of the *Longman Grammar of Spoken and Written English* (Biber et al. 1999).

Lexical bundles have attracted researchers' attentions in the field of corpus linguistics (e.g. Ådeland Erman 2012; Biber and Barbieri 2007; Biber, Conrad, and Cortes 2004; Chen and Baker 2016; Cortes 2013; Csomay 2012; Durrant 2017; Hyland 2008; Esfandiari and Barbary 2017; Staples et al. 2013). They are defined as "a recurring sequence of three or more words" that "commonly go together in natural discourse" (Biber et al. 1999, 990). They have been also regarded as multiword expressions in which "words follow each other more frequently than expected by chance" (Hyland 2008, 6). They are classified into structural and functional types (Biber, Conrad, and Cortes 2004) and further analyzed in research articles, doctoral dissertations, and master's theses and grouped into an academic functional categorization (Hyland 2008).

There have been studies trying to investigate lexical bundles in terms of their structural and functional categorizations in different settings and disciplines (Biber and Barbieri 2007; Cortes 2006, 2013; Grabowski 2015). In a study aiming at capturing the connection between lexical bundles and different moves existent in research article introductions, Cortes (2013) found a new characteristic of lexical bundles called triggers which "started

the move and were used at the beginning of the clause that triggered the move or step” (p. 41). This characteristic had not been identified before in the literature.

Therefore, this study attempts to explore lexical bundles in a large corpus compiled based on an impressive number of recent papers published in the field of applied linguistics. If, in addition to the categories previously found in the literature, some new structural and functional lexical bundle characteristics were observed; then, it could be asserted that the more specific the register under study, the higher the possibility of new lexical bundles as well as new structural and functional characteristics. Also, pedagogically, since lexical bundles help to “shape meanings in specific contexts” and contribute “... to our sense of coherence in a text” (Hyland 2008, 4), the generated lists of lexical bundles from the recent papers in the field of applied linguistics can be used more strategically in ESP and EAP text books.

2. Review of literature

Progression beyond the word in some studies dealing with formulaic language (Hyland 2008, for example) is believed to be first initiated by the work of Jespersen (1924), and Firth ([1951]1957). More recently, Pawley and Syder (1983) explained “Memorized sequences” and believed that conversation includes a small number of novel linguistic creations (p. 205). This can mean that memorized expressions contribute to a great extent to the conversation. This was followed by the notion of “lexical phrase” introduced by Nattinger and DeCarrico (1992) as “chunks of language of varying length” (p. 1). They also believed that these “routinized formulas” can “play a large part in both acquiring and performing language” (p. 1). These studies were followed by many studies taking formulaic language into account: Biber et al. (1999); Cortes (2002, 2004); Wray (2000, 2002), to name a few. It shows that formulaic language as an effective unit of meaning has been the concern of researchers for many years. Moreover, Sinclair (2004, 25–26) considers the recognition of words as co-dependent units of meaning and further scrutinizes the equality of word to unit of meaning to claim that words together “make a meaning that is different from the normal putting together of their individual meanings.” In another study, Sinclair (2004), introducing the “Lexical Item,” highlights the importance of surrounding elements of a word in building strong patterns. This shows that the concern of many researchers has been linguistic patterns which make up the surrounding of the words in texts. One of these patterns is referred to as lexical bundle in literature. Lexical bundles are defined as frequent use of the same strings of words in natural discourse (Biber et al. 1999), and frequent happenings of the same words as expressions more than expected by chance (Hyland 2008). Two primary studies of lexical bundles are those of Biber, Conrad, and Cortes (2004) and Hyland (2008). These studies were based on different corpora and proposed structural and functional classifications for lexical bundles.

Biber, Conrad, and Cortes (2004, 381, 384–388) grouped them into structural and functional types. Regarding their structural categorization, lexical bundles can incorporate “verb phrase fragments” (e.g. *can be used to*), “dependent clause fragments” (e.g. *that there is a*), and “noun phrase and prepositional fragments” (e.g. *one of the things*). Functionally, they are sub-categorized into stance expressions (e.g., *I don’t know that, the fact that the, if you want to, it is important to, going to be a, it is possible to*). Next functional subcategory of lexical bundles is called discourse organizers which “reflect

relationships between prior and coming discourse" (e.g. *in this chapter we, as well as the*). Last functional subcategory of lexical bundles is called referential expressions which "make direct reference to physical or abstract entities, or to the textual context itself (e.g. *is one of the, something like that, the rest of the, the size of the, in terms of the, in the United states, shown in figure N*).

Exploring academic disciplinary variation of lexical bundles, Hyland (2008) proposed a functional classification for them and subcategorized them into research-oriented, text oriented, and participant-oriented types (pp. 13–14). Research-oriented lexical bundles "help writers to structure their activities and experiences of the real world" (e.g. *at the beginning of, the use of the, a wide range of, the structure of the, in the Hong Kong*). Text-oriented lexical bundles are "concerned with the organization of the text and its meaning as a message or argument" (e.g. *in addition to the, as a result of, in the present study, in the case of*). Participant-oriented lexical bundles focus on the writer or reader of the text (Hyland 2008) and include stance features which carry writer's stance towards a following proposition (e.g. *may be due to, it should be noted that*).

The above two primary classifications of lexical bundles have been followed by many studies investigating lexical bundles in different contexts based on those classifications. Biber and Barbieri (2007) investigated the use of lexical bundles in university spoken and written registers. They found that nonacademic spoken registers (management sessions, for example) make use of many lexical bundles which are the building blocks of that register. They further argued that bundles under the category of stance expressions (e.g. *you might want to*) are the most common in the spoken registers. They also found that although written course management "makes frequent use of bundles from all three functional classes," stance lexical bundles are very common in that register. In another study, Cortes (2006) collected research articles of American History, and identified the most frequent four-word lexical bundles used in them. After that, she classified those lexical bundles functionally and divided them into discourse organizers (e.g. *on the other hand*) and referential expressions (e.g. *some of the most*) in order to teach them to a group of students from a writing-intensive history class. She concluded that although this instruction increased the awareness and the interest of the students in those lexical bundles, their assignments showed no difference in the use of those lexical bundles between pre- and postinstruction. Another study which takes lexical bundles into account is Csomay (2012). Csomay (2012) investigated more than 1000 discourse units in university class sessions and identified 84 lexical bundles. Class sessions were divided into two phases in this study: "opening and instruction phase." Csomay (2012) found that stance bundles (e.g. *you don't how to*) were the most frequent in the first phase and referential bundles (e.g. *at the end of the*) were the most frequent in the instructional phase. Exploring specific registers, some studies; however, have found bundles with structural or functional characteristics not previously identified in the literature. For example, exploring lexical bundles in history and biology writing, Cortes (2004) found that lexical bundles like *the power of the* "reflect an inherent connection to the social connection to the social disciplines, especially those related to social events or issues" (p. 406). In addition, no comprehensive study has yet tried to identify the lexical bundles used in published papers in the field of applied linguistics in an attempt to explore all the bundles used in this specific register as well as all their possible new structural and functional characteristics. Therefore, investigating a large corpus of papers published in

the field of applied linguistics, this study aimed at identifying the most frequent 3-, 4-, 5-, and 6- words lexical bundles as well as their structural and functional patterns in this register.

3. Method

To explore all the lexical bundles dominant in the field of applied linguistics, both a corpus-based and a corpus-driven approach (Biber 2010) were used to first investigate the corpus for its most frequent lexical bundles used and; second, to find what new subcategories, in addition to primary ones, might emerge.

3.1. The applied linguistics corpus

The corpus on which this study is based is a 16-million-word corpus compiled based on the research articles collected from the field of applied linguistics. To set up the applied linguistics Corpus (ALC),¹ the articles published in some well-established journals in the field of applied linguistics were chosen because many articles are being published every year in those top journals and they consist of highly cited papers. And also, other meta-analytic studies usually refer to the same journals (see Boo, Dörnyei, and Ryan 2015, for example). The ALC consists of 1738 published research articles from five different journals: Applied Linguistics, Modern Language Journal, Studies of Second Language Acquisition, System, and TESOL Quarterly (all articles published in them from 2009 to 2018). Table 1 gives more information about the numbers of articles published at the time of this study which were included in the corpus. It should be stated that, as the purpose of this study was to investigate the papers published in the field of applied linguistics to explore the structural and functional characteristic of the lexical bundles, all the papers published in the five different journals, at the time of conducting this study, were included in the corpus to ensure a comprehensive exploration of all the possible lexical bundles in this specific register.

3.2. Identification of the lexical bundles

The tool used to explore the ALC was Sketch Engine. It is an online language exploration tool (Kilgariff et al. 2014). Sketch engine can take large corpora for investigation and offers many different functions such as making concordance lines, finding collocations, and making wordlists. For the purpose of this study the ALC was uploaded into the sketch engine and the n-gram function was used to generate lists of 3-, 4-, 5-, and 6-word lexical bundles out of the ALC. After choosing the n-gram, setting the n from 6 to 6 (to generate

Table 1. Journals included in the ALC.

Journal	Year	Number of articles
TESOL Quarterly	2009–2018	219
The Modern Language Journal	2009–2018	378
System	2009–2018	668
Studies in Second Language Acquisition	2009–2018	199
Applied Linguistics	2009–2018	274
Total 5	-	1738

only 6-word lexical bundles), and clicking on *Make word list*, all the 6-word lexical bundles appeared on the screen. The same process was done to obtain lists of 3-, 4-, and 5-word lexical bundles.

There are two criteria in the literature when studies deal with lexical bundles: frequency and range (Biber, Conrad, and Cortes 2004; Hyland 2008; Cortes 2013). Frequency refers to how many times a recurrent expression should occur in a corpus to be regarded as a lexical bundle. In a chapter of the book *Longman Grammar of Spoken and Written English*, Biber et al. (1999) took a frequency of 10 (10 times per million words) as the cut-off point for a recurrent expression to be regarded as a lexical bundle. Further studies took different frequency cut-off points for lexical bundles: 40 times per millions words (Biber, Conrad, and Cortes 2004); 20 times per million words (Hyland 2008). In this study cut-off points of 50, 30, 15, and 10 times per million words are taken for 3-, 4-, 5-, and 6-word lexical bundles respectively. The reason for different frequency cut-off points is that the frequency of lexical bundles decreases as they contain more words (Cortes 2013). Therefore, the criteria used in this study to identify bundles of different length which meet the conservative frequency cut-off points is as follows: First, using the n-gram feature of the Sketch Engine, all the 6-word recurrent expressions in the ALC were identified. Then, as the frequency cut-off point for 6-word lexical bundles was 10 times per million words and the corpus under investigation in this study (the ALC) contained 16 million words, only the 6-word recurrent expressions which were repeated at least 160 times (10 multiplied by 16) in the ALC were considered as 6-word lexical bundles, and those with lower numbers were not included in the list. Similarly, the same criterion was applied for identifying 5-, 4-, and 3-word lexical bundles with the exception that 5-word recurrent expressions should occur at least 240 times (15 multiplied by 16) in the ALC, 4-word recurrent expressions should occur at least 480 times (30 multiplied by 16) in the ALC, and 3-word recurrent expressions should occur at least 800 times (50 multiplied by 16) in the ALC to be considered as lexical bundles. Also, all those with lower numbers were not included in the corresponding lists. Taking advantage of this set of criteria resulted in an adequate level of accuracy of the findings regarding the frequency cut-off points of the lexical bundles.

On the other hand, range refers to how often a lexical bundle occurs in a corpus, in that they should occur in a predetermined number or a specified percentage of files in the corpus. The common range cut-off point for a recurrent expression, while meeting the frequency cut-off point, is five to be considered as lexical bundle. It means that a recurrent combination of words should occur in at least five different texts in a corpus to avoid author idiosyncrasy (Biber, Conrad, and Cortes 2004). In another study, Hyland (2008) took a 10% range cut-off point for lexical bundles for the same purpose to ensure that a recurrent expression is typical of the entire corpus. The same approach is taken in this study. Any combination of words had to occur at least in 10% of the text in the ALC to be regarded as a lexical bundle. To achieve this, using the MS Word Split tool, the ALC was divided into ten equal parts, with each part containing 10% of the pages as well as 10% of the words of the whole corpus. Then, each 3-, 4-, 5-, and 6- word lexical bundle which had met their corresponding frequency cut-off points were individually searched in each of the ten parts. Subsequently, each recurrent expression had to occur at least once in all the ten parts, making up one tenth of the whole corpus each, to be regarded as a lexical bundle. This resulted in removing recurrent expressions which rendered writer

idiosyncrasies and; thus, were not identified as lexical bundles. For example, the recurrent expression *ideal L2 self* was frequent enough (851 times in the whole corpus) to meet the frequency cut-off point for 3-word lexical bundles (50 times per million words) while no more than seven different parts of the ALC included the instances of *ideal L2 self*. Therefore, the recurrent expression *ideal L2 self* was not eligible to be considered as a lexical bundle and was removed from the list. This process allowed the researchers to ensure that the lexical bundles found in their study meet the range cut-off point as well.

Before giving more information about the way bundles were considered for investigation in this study, a point about the studies which deal with lexical bundles in the literature needs to be explained. Usually, studies which take lexical bundles into consideration target only 4-word bundles because smaller lexical bundles (3-word) are usually combined into bigger ones (Biber, Conrad, and Cortes 2004; Hyland 2008).

But this study takes a wider perspective. It attempts to investigate all of the lexical bundles with different length to see if 3-, 5-, and 6-word lexical bundles reveal the same structural and functional identities as 4-word lexical bundles or not. And if not, what new structural and functional patterns might emerge in the specific register under study (published papers in the field of applied linguistics). To achieve this, first the ALC was uploaded into the Sketch Engine and all the 6-word lexical bundles were extracted out of it. After doing so, all of the 6-word bundles were removed from the ALC. In this way the ALC contained no frequent 6-word lexical bundles; therefore, investigating it for 5-word lexical bundles resulted in only 5-word lexical bundles which were not embedded in 6-word lexical bundles because 6-word lexical bundles were removed before investigating 5-word lexical bundles. So, after removing 6-word lexical bundles, which met the frequency and the range criteria, from the ALC, the ALC-minus-6-word-lexical-bundles was uploaded into the Sketch Engine for 5-word lexical bundles investigation. The following is an example of the process: First, it was found that *It is important to note that* is a 6-word lexical bundle in the ALC. Second, all of the instances of these lexical bundles were removed from the ALC. Third, the same 6-word lexical bundle subtraction was done for all the other frequent 6-word bundles. This resulted in the ALC-minus-six-word-lexical-bundles. Forth, the ALC-minus-six-word-lexical-bundles were uploaded into the Sketch Engine as a new corpus for investigating smaller lexical bundles. This process made it possible to claim that if a 5-word lexical bundle incorporated the word combination *It is important to note* was a pure 5-word lexical bundle which do not expand into any 6-word lexical bundles (e.g. *It is important to note that*), because all of the 6-word lexical bundles were removed before the 5-word lexical bundles investigation. Similarly, after the 5-word lexical bundles were extracted out the ALC-minus-six-word-lexical-bundles, the 5-word lexical bundles list was saved and the 5-word bundles were removed from the ALC-minus-six-word-lexical-bundles. Then, the new corpus which was saved as ALC-minus-five-word-lexical-bundles was uploaded into the Sketch Engine for 4-word lexical bundles investigation. The same process was done for 4-word and 3-word lexical bundles. So, the results were lists of pure 3-, 4-, 5-, and 6-word lexical bundles in which smaller bundles were not expanded into the bigger ones.

In addition, for a more accurate identification of lexical bundles structures, another feature of Sketch Engine named Corpus Query Language (CQL) was used. As the corpora uploaded into the Sketch Engine could be tagged automatically for parts of speech of words (all past tense *be* verbs are tagged *BD*, for example), CQL was used to investigate specific structural lexical bundle patterns used in the ALC. Using this feature, one can, for

example, search a corpus for infinitives: all the words *to* followed by the base form of the verbs (*to* + base form of the verbs). Results and Discussion section gives more information regarding how this feature was used in this study.

In order to conclude with more reliable results, all the procedures done in this study were performed twice by the two researchers individually and the results showed 99% agreement. The one percent disagreement was because some bundles could express two different identities and each researcher took a different perspective toward them.

4. Results and discussion

4.1. ALC lexical bundles

As the purpose of this study was to explore the lexical bundles used in the context of applied linguistics academic writing, a general percentage of all the lexical bundles present in the ALC as well as structural and functional analyses are presented in the following sections. After presenting lexical bundles in each category, lists of lexical bundles which were not previously found in the literature but are present in research articles in the field of applied linguistics as a specific register are discussed. These discussions are followed by new structural and functional subcategorizations of the lexical bundles exclusively found in the ALC.

4.1.1. Proportion of lexical bundles in the ALC

A total of 1,045 different lexical bundles, regardless of their frequency, were identified in the ALC. The most frequent 3-, 4-, 5-, and 6-word lexical bundles were *as well as*, *on the other hand*, *at the end of the*, and *It is important to note that* respectively. Some of these bundles are also identified as frequent bundles in the literature (See Hyland 2008). The 1,045 recurrent expressions amount to 1,652,275 words (words in bundles with different length multiplied by their frequency) which make up around 10% of the 16 million words in the ALC, highlighting the point that lexical bundles with different length contribute considerably to discourse of applied linguistics academic writing. Figure 1 illustrates this point.

4.1.2. Structural analysis of ALC lexical bundles

Structural analysis of lexical bundles generated from the ALC is discussed according to the structural classification proposed by Biber, Conrad, and Cortes (2004). Table 2 illustrates lexical bundles that incorporate verb phrases.

As shown in Table 2, there are many 3-word lexical bundles which do not expand to larger ones in the ALC. For example, the bundle *were asked to* has occurred 1,399 times in ALC and is not expanded to any larger lexical bundle because further manual analysis of the concordance lines including this lexical sequence showed that it has not combined with any other words to make a frequent 4-word, 5-word, or 6-word lexical bundle. This bundle which incorporates a passive verb mostly pointed to situations where participants of a study needed to perform an action. The following is an example of this lexical bundle:

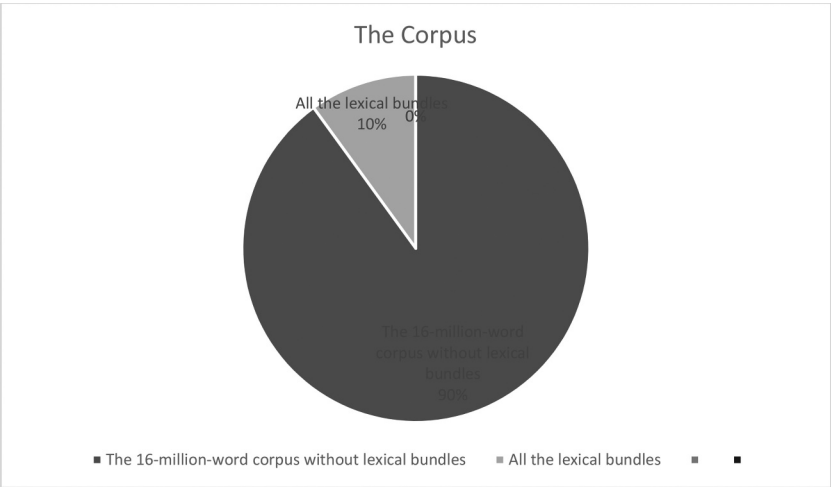


Figure 1. Percentage of all lexical bundles in the 16-million-word corpus.

Table 2. Most frequent lexical bundles that incorporate verb phrases in the ALC.

3-word	4-word	5-word	6-word
were asked to can not be may not be needs to be found to be was used to	can be used to was found to be is based on the can be explained by can be found in	can be seen in the has been found to be were not included in the	is based on the fact that be explained by the fact that

(1) Once again, teachers **were asked to** make recommendations about fictional students based on their academic achievement as revealed by the grades on their record cards. Teachers were then invited to reflect on the reasons behind their decisions.

The bundle *needs to be* (see Table 2) is another example which has occurred 839 times (more than fifty times per million words) and, as the approach to identify lexical bundles in the methods section suggests, is not expanded to any larger lexical bundle in the papers published in the field of applied linguistics. This lexical bundle is always followed by a past participle in the ALC which makes a passive voice in texts. It should also be noted that although in the study by Hyland (2008) this bundle was marked as “a locative or logical relation” (p. 11), the bundle *needs to be* generally refers to an urgent situation in the ALC.

(2) However, perhaps the larger issue that **needs to be** addressed at this juncture are the differences between beliefs and ideologies.

Some bundles; however, may seem controversial. As shown in Table 2, recurrent expressions *found to be*, *was found to be*, *is based on the*, *is based on the fact that*, *can be explained by*, and *be explained by the fact that* may seem to be only three different recurrent expressions combined with some words to emerge as larger bundles and; thus, should not be regarded as six different lexical bundles. However, this is not the case. As explained in the method section, all larger lexical bundles were removed from the ALC before the investigation of smaller ones. This means that a 4-word lexical bundle like *can be explained*

by has been frequent enough to meet the frequency cut off point (forty times per million words) and range cut off point (occurring at least once in ten different parts of the corpus) after removing all of the instances of its 6-word *be explained by the fact that* counterpart. The following are examples of concordance lines of these kinds of bundles:

- (3) Gestures have been **found to be** used for an array of purposes, most of which have been hitherto unreported in the second language studies literature, though some are well-documented in research involving native speakers of English.
- (4) The reliability index **was found to be** acceptable (34 participants, 51 items, Cronbach's alpha =.79).
- (5) As discussed in Section 1.2 above, register research **is based on the** claim that all linguistic variation is functionally associated with the situational context.
- (6) This **is based on the fact that** grammar learning is a complex process and relies on both memorization of individual items and processing of the relationships among them.

As it is apparent from the first two extracts above, the smaller bundles are not preceded or followed by the same words as the larger bundles do. These results indicate that 3-word lexical bundles reasonably contribute to the discourse of applied linguistics academic writing. As a result, their omission, as it is common in some previous studies (see Cortes 2004; Hyland 2008, for instance), in the belief that they usually expand into 4- or 5-word lexical bundles, may adversely affect the results.

The next structural subcategory proposed by Biber, Conrad, and Cortes (2004) refers to lexical bundles which consist of dependent clauses. Examples of the ALC bundles which fit into this subcategory are presented in Table 3.

Regardless of their length, these bundles were structurally in line with those previously found by Biber, Conrad, and Cortes (2004).

Most frequent lexical bundles which fit into the third structural subcategory proposed by Biber, Conrad, and Cortes (2004), bundles with noun phrases and prepositional phrases, are given in Tables 4 and 5. The reason for dividing this subcategory into two tables in this study is that prepositional phrases could act as adjectives or adverbs in sentences in the ALC; therefore, it was necessary to present them in a new table in order to provide further explanation about the subcategory under discussion.

Table 4 also illustrates bundles with varying lengths and the only one which may seem to be expanded to another lexical bundle is *English as a foreign language*. This bundle seems to be combined into *learning English as a foreign language*. The following extracts show that *English as a foreign language* is mostly used to modify a following noun in the ALC; however, the bundle *learning English as a foreign language* does not act as an adjective phrase. Accordingly, smaller-length lexical bundles may seem to simply expand into larger ones;

Table 3. Most frequent lexical bundles that incorporate dependent clauses in the ALC.

3-word	4-word	5-word	6-word
there is a this is a that there are it has been	that there is a the following is the	participants in this study were the participants were asked to The results showed that the there is a need for the results of this study the participants in this study	there was no significant difference between purpose of this study was to purpose of this study is to

Table 4. Most frequent lexical bundles that incorporate noun phrases in the ALC.

3-word	4-word	5-word	6-word
the use of the number of the role of use of the the amount of	the use of the a wide range of the extent to which the ways is which The results of the	English as a foreign language English as a second language English as a lingua franca the extent to which the the results of this study the participants in this study	the Common European Framework of Reference learning English as a foreign language the findings of the present study

however, their structural and functional characteristics may change as they expand, as seen in the results of this study. This point makes the presence of bundles smaller or larger than 4-word ones significant to ensure more comprehensive results.

- (7) The effects of focused and unfocused written corrective feedback in an *English as a foreign language* context.
- (8) This study investigated children’s motivation for *learning English as a foreign language* (EFL) and intrinsic motivation for learning in general.

Table 5 presents lexical bundles with prepositional phrases. As stated above, they can act as adjectives or adverbs in sentences. Lexical bundles which function as adjectives are presented before those which act as adverbs.

The following are examples of bundles which consist of prepositional phrases functioning as adjectives in sentences:

- (9) After piloting the test with 30 students in another class *from the same* department, 14 items were eliminated in cases where more than one answer could be correct or if the item was too easy.
- (10) Although the learners *in our study* were highly familiar with each other and used to working together in class, some pairs and groups were more collaborative and produced more LREs than others, therefore creating more L2 learning opportunities.

As shown in extract (9), the bundle *from the same* modifies the noun *class* and acts as an adjective. The recurrent expression *in our study* functions as an adjective by giving more information about its preceding noun *learners* in extract (10).

Some lexical bundles in **Table 5** function as adverbs in the ALC:

Table 5. Most frequent lexical bundles that incorporate prepositional phrases in the ALC.

3-word	4-word	5-word	6-word
<i>Prepositional phrases acting as adjectives</i>			
from the same	from the perspective of	in the foreign language classroom	of English as a foreign language
in our study	in second language acquisition		of English as a second language
in the learning	in the present study in a second language		
<i>Prepositional phrases acting as adverbs</i>			
in the sense	in the United States in the present study	at the end of the at the beginning of the	at the time of the study

(11) [...] findings were also supported **in the sense that** verb movement and finiteness are not entirely related, as has been demonstrated in L1 acquisition.

The lexical bundle *in the sense that* in extract (11) functions as an adverb, in that it expresses the relationship between object of a proposition and the verb phrase *were supported* to answer the question of how? Or under what conditions? The findings in that study were supported.

Although a considerable number of lexical bundles found in the ALC fit into the structural categorization proposed by Biber, Conrad, and Cortes (2004), some lexical bundles found in the applied linguistics academic writing expressed new structural characteristics. The explanation regarding the new lexical bundles found in the ALC is as follows.

There are some lexical bundles which do not include a pure verb phrase, dependent clause, noun phrase, or prepositional phrase but functioned as adjective phrases in the ALC. These bundles are presented in Table 6. It should also be noted that the bundles *has been shown to be* and *can be seen in the* seem to fit under the subcategory of lexical bundles with verb phrases, but further concordance line analysis showed that these two bundles demonstrate a specific structural identity, acting as adjective phrases, they can be classified under the structural category of bundles as a new subcategory which act as adjective phrases and are called, structurally, adjectival phrases as well.

To illustrate the identity of the bundles presented in Table 6, an example of these bundles in the ALC is presented:

(12) In addition to a short questionnaire aimed at eliciting biographical information, the instruments **used in the present study** were (a) difficulty judgement questionnaires for all participants and (b) tests of explicit and implicit L2 knowledge for the learners.

As it is obvious in extracts (12) the lexical bundle *used in the present study* modifies *instruments*. This means that this bundle has modified its preceding noun, acting as adjectives. Adjectival phrases as lexical bundles found in the ALC are presented in Table 6.

There were also bundles which acted as adverbs in text and are called, structurally, adverbials: *so that the*, *depending on the*. The following is an extract from the ALC to discuss this kind of bundles in more detail:

(13) Circumlocutions are classified **depending on the** features selected for description. Thus, we distinguish among function, superordinate, description of physical characteristics of object, and locations

The lexical bundle *depending on the* in extract (13) modifies its preceding verb *classified*. This shows two points: first, lexical bundles smaller than 4-word ones can perform a specific action in text. Second, lexical bundles can act as adverbs in texts as well.

Some lexical bundles in the ALC incorporate dependent clauses which are not previously found in the literature. They can act as adjective clauses or adverb clauses in texts.

Table 6. Most frequent lexical bundles that incorporate adjectival phrases in the ALC.

3-word	4-word	5-word	6-word
		used in the present study	is beyond the scope of this
		has been shown to be	
		used in the current study	
		can be seen in the	
		plays an important role in	

Table 7 presents examples of this kind of bundles. It should be noted that some of the bundles presented in this table may also concord with the primary structural tables, but as they act as adjectives or adverbs specifically, they are particularly presented in this table. For a more detailed discussion of them extract (14) is generated from the ALC:

(14) They would like their teachers to provide some more specific follow-up questions, **so that they can** realize what their errors and mistakes are after clarification request feedback.

As extract (14) indicates, the bundle *so that they can* modifies the verb *provide*, in that the action of “providing” is done for a specific reason and the bundle *so that they can* answers the question of “why?” by expressing the relationship between its preceding and following propositions. In addition, these results show that lexical bundles smaller or larger than 4-word ones can contribute reasonably to the discourse of academic writing in the field of applied linguistics.

The next newly identified type of lexical bundles in the ALC incorporated a specific structural pattern. As shown in Table 8 the specific structure is Be + Adjective/Adverb + to. The Corpus Query Language (CQL) function of Sketch Engine was used to identify the frequency of this specific structural pattern ([lemma = “[a-z]*” & tag = “VB.*”] ([lemma = “[a-z]*” & tag = “JJ.*”][lemma = “[a-z]*” & tag = “RB.*”]) [word = “to.*”]). It was found that this structure occurred 20,289 times in the ALC (see Figure 2). And this number was reasonable enough to take this structure as frequent. Extracts (15) and (16) are provided to give examples and discussions for this type of bundles:

(15) That is, for a learner to understand a word’s meaning in listening, s/he must **be able to** recall (rather than recognize) that meaning once the form is recognized.

(16) Slightly better control in grammar, vocabulary use, and fluency by the former group may **be due to the fact that** students in the F2F class had an opportunity to freely and easily clarify grammar facts in each class whenever they wanted to.

These extracts show that the lexical bundle structure be + adjective/adverb + to can be used to express the ability (see extract 15), or the possibility of a specific reason for a result in a study. The extracts (15), and (16) further show that these 4-word lexical bundles found in the ALC are not expansions of 3-word lexical bundles since the lexical bundle *be able to* is not followed by the word *use* to argue for that kind of expansion (*be able to use* is a 4-word lexical bundle in Table 8).

Table 7. Most frequent lexical bundles that incorporate adverb/adjective clauses in the ALC.

3-word	4-word	5-word	6-word
Because it is	although it was suggested	although many studies have examined	
Because they are	while previous studies investigated		
which can be	so that they can		
who did not	when it comes to		

Table 8. Most frequent lexical bundles that incorporate be + adjective/adverb + to fragments in the ALC.

3-word	4-word	5-word	6-word
be able to	be able to use	be able to understand the	be due to the fact that
are likely to			is due to the fact that

Query [a-z]*, VB.*, JJ.*, RB.*, to.* 20,289 (937.09 per million) ⓘ

First | Previous Page 17 of 1,015 Go Next | Last

doc#0 inflected form, hij is liegt "he lies." It **is important to** know that Turkish learners produce more basic

doc#0 element between the subject and the verb, it **is impossible to** decide on the exact position of the verb. The

doc#0 As illustrated in (27c), the examples in Cycle 3 **are similar to** those in Cycle 1 (27a). The regressive pattern

doc#0 V to I and then from I to C (the latter movement **is unfamiliar to** learners with a Turkish L1 background). The

doc#0 in (32), in which the complementizer **was difficult to** perceive because it followed the stressed

doc#0 of inflected verbs appear in Cycle 3. Emine **is able to** use copulas, -s-patterns, and inflected forms

doc#0 be adjoined to AgrP, but because a CP projection **is necessary to** generate interrogative 00 00 Stage Syntactic

doc#0 . As verb movement and inflectional morphology **are hard to** apply, the learners resort to easier (more

doc#0 viewed as a pronoun (Klein & Perdue, 1992). 5. I **am indebted to** the anonymous SSLA reviewer who drew my

doc#0 , Bley-Vroman & Chaudron, 1994). 8. In Dutch, it **is possible to** express ongoingness with a prepositional

doc#0 , to say that foreign languages are languages **is not to** say that foreign languages are governed (

doc#0 that the component elements of the module **are unique to** it" (2005, p. 5), and "we need no longer assume

doc#0 entertain the possibility that they might **be reducible to** language-independent principles" (p. 9).

doc#0 property of substitutability, which **is essential to** show the learnability of structure dependence

doc#0 faculty are not unique to the language faculty **is not to** say that a language module does not exist. The

doc#0 to the FDH: In the context of the original FDH, it **was possible to** say that child language development used a rich

doc#0 to be affected by conscious awareness and may **be amenable to** explicit instruction, can manipulate objects

doc#0 sense. To take a famous example, it would **be wrong to** characterize Wes (Schmidt, 1983) as a failure.

doc#0 correct, see Valliant (1984). 6. This **is not to** say that the system is able to learn reliably

doc#0 (1984). 6. This is not to say that the system **is able to** learn reliably even in the face of misleadingly

First | Previous Page 17 of 1,015 Go Next | Last

Figure 2. Frequency of the structure be + adjective/adverb + to in the ALC generated by sketch engine.

Table 9. Most frequent lexical bundles that begin with a conjunction in the ALC.

Lexical bundles
but also to
and second language
and that the
and so on
because of the

Table 10. Most frequent lexical bundles that incorporate infinitive phrases in the ALC.

3-word	4-word	5-word	6-word
to be an	to be able to	to participate in the study	to determine the extent to which
to address the		to learn a foreign language	to communicate in a second language
to assess the		to examine the effects of	to answer the following research questions

ALC also included lexical bundles which began with a conjunction. These bundles are presented in Table 9.

ALC also included lexical bundles which incorporate infinitive phrases. Table 10 presents most frequent examples of these bundles in the ALC. Extracts below are given to show how these bundles are used in sentences:

(17) Rather than providing a description of the various components that contribute to the subjects' motivation and identifying the correlation between motivation and learning outcomes as in social-psychological or cognitive approaches to L2 motivation, this study aims **to address the** issue of motivation through the following question:

(18) Perhaps thinking in the L1 or resorting to the L1 may be "the only way a learner can ever begin **to communicate in a second language**" [...]

The lexical bundle *to address the* in extract (17) is used to signal the purpose of that study. In extract (18), the bundle *to communicate in a second language* is used to show the result of an assumption. In just two examples, two different functions of this kind of bundles (lexical bundles which incorporate infinitive phrases) were shown: to signal the purpose of studies, and to show the result of an assumption.

Up to now, the structural classification of lexical bundles in the ALC has been presented. The next section deals with the functional classification of the lexical bundles in the ALC.

4.1.3. Functional Analysis of ALC lexical bundles

The classification used for the functional analysis of the ALC lexical bundles was the one developed by Hyland (2008). It should be noted that, both Biber, Conrad, and Cortes (2004) and Hyland (2008) have developed a functional categorization for lexical bundles. The one in Biber, Conrad, and Cortes (2004) covers the discourse in classroom teaching and textbooks. On the other hand, Hyland (2008) has investigated research articles, doctoral dissertations, and master's theses for lexical bundles. In this study, the classification proposed in the latter study is taken for consideration because of the similarity in the target discourse of that study and the present study. It was found that overall, a considerable number of bundles used in published papers in the field of applied linguistics perform functions similar to those performed by lexical bundles previously found in the literature. In addition to the functional sub-categories proposed by Hyland (2008), some other sub-categories of lexical bundles had to be added to define some new bundles found in the ALC.

Most frequent examples of the bundles found in ALC which share the similar functional identity to those identified by Hyland (2008), which are called research-oriented lexical bundles, are presented in Table 11. It should be noted that, in line with the findings of Hyland (2008), noun phrase + of structures found in ALC were mostly associated with research-oriented function.

The next sub-category of the functional classification proposed by Hyland (2008) is called text-oriented lexical bundles which are "concerned with the organization of the text and its meaning as a message or argument" (p. 13). Most frequent examples of the

Table 11. Research-oriented lexical bundles in the ALC.

Lexical bundle	3-word	4-word	5-word	6-word
Location Procedure	in the study the use of	in the present study used in this study	at the beginning of the to participate in the study	at the time of the study purpose of this study was to
Quantification Description Topic	the number of the quality of learners of English	a wide range of the nature of the in second language acquisition	by the total number of in the form of a English as a foreign language	is beyond the scope of this learning English as a foreign language

Table 12. Text-oriented lexical bundles in the ALC.

Lexical bundle	3-word	4-word	5-word	6-word
Transition Signals	in line with	in addition to the	while at the same time	
Resultive Signals	found to be	the results of the	as a result of the	the findings of the present study
Structuring Signals	in the next	As shown in Table	as can be seen in	As can be seen in Table
Framing Signals	with respect to	in terms of the	on the basis of the	

Table 13. Participant-oriented lexical bundles in the ALC.

Lexical bundle	3-word	4-word	5-word	6-word
Stance Features	the fact that	are more likely to	is one of the most	is based on the fact that
Engagement Features	showed that the	it is important to	It should be noted that	It is important to note that

similar bundles found in the ALC are presented in [Table 12](#). The pattern which was mostly associated with this function in the ALC was prepositional phrase which is in line with the findings of Hyland (2008).

The final functional sub-category of lexical bundles is called participant oriented lexical bundles (Hyland 2008) which are “focused on the writer or reader of the text” (p.14). Most frequent examples of the similar bundles found in the ALC are presented in [Table 13](#). It should be noted that, the pattern which was mostly associated with this function was anticipatory it pattern and this was, also, in line with the findings of Hyland (2008).

Overall, although the functional classification proposed by Hyland (2008), efficiently defined a large number of lexical bundles in ALC, some bundles found in the applied linguistics academic writing expressed rather new functional identities. They are presented in the following.

Some bundles and functions newly found in the ALC tend to refer to two things in the texts: Doubling Lexical Bundles. Most frequent examples of these bundles are given in [Table 14](#).

The following examples are given to discuss the doubling lexical bundles in more detail.

(19) We observed that L2 proficiency interrelates with **L1 and L2** influence in different ways depending on the specific mechanism used, so that some mechanisms which are frequent at 4th grade decrease their presence, basically borrowing, and some other tend to increase, basically L2-based mechanisms as linguistic proficiency, awareness and cognitive development increase.

(20) Some researchers [...] further argue that culturally responsive pedagogy designed to address these imbalances is relegated to the margins in favor of standardized **teaching and learning**, but the theoretical literacy grounding provided for educators of language-minority students tends to be founded on frameworks that centralize the norms and values of the White, native-English-speaking population [...]

In extract (19) the bundle *L1 and L2* is mainly related to the field of applied linguistics and has a twofold identity: first language and second language.

Another type of bundles with a new function found in the ALC referred to a point of divergence in textual context. Contrasting Lexical Bundles are presented in [Table 15](#). The following example is given to show how contrasting lexical bundles are used in the ALC.

Table 14. Some most frequent bundles that refer to two things in the ALC: Doubling Lexical Bundles.

3-word	4-word	5-word	6-word
L1 and L2	language teaching and learning	the L1 and the L2	at the beginning and end of
teaching and learning	language learning and teaching	in second and foreign language	
teachers and students	between L1 and L2	the teaching and learning of	
reading and writing	in L1 and L2		

Table 15. Some most frequent bundles that refer to a contrast in the ALC: Contrasting Lexical Bundles.

3-word	4-word	5-word
Rather than the Rather than a It is not as opposed to	did not seem to was not statistically significant on the other hand	but at the same time were not included in the while at the same time

(21) In his study of English language and global hip-hop cultures, Pennycook attempts to elucidate the ways in which global languages and cultures offer alternative identities and forms of expression, **while at the same time** being reshaped to meet local needs, and then being sent back out again with new forms and meanings in a circular or flowing process.

In extract (21), the recurrent expression *while at the same time* is used to refer to a point of divergence, in that although different languages and cultures can provide different identities, they themselves are not immune to changes.

Another type of lexical bundles found in ALC which expressed a specific function referred to examples in texts of the register under study. The most frequent Exemplifiers in ALC are presented in Table 16.

To discuss how exemplifiers are used the following example is given:

(22) It refers to a teacher’s ability to predict language learner difficulties **such as the** negative transfer of rules and patterns from the L1 or speech sounds that exist in the target language but not in the L1 system.

In extract (22), the lexical bundle *such as the* is used to point to instances of learner difficulties.

Some bundles in the ALC were used mostly for research questions. Most frequent Questioning Lexical Bundles found in ALC are presented in Table 17. Examples below are given to show questioning lexical bundles in the text:

(23) This study was designed to address the following questions: 1. **To what extent** do L2 knowledge skills and L2 processing skills predict communicative success in L2 speaking, in a mixed-proficiency group of adult L2 learners?

Table 16. Some most frequent lexical bundles that refer to examples in the ALC: Exemplifiers.

3-word	4-word	5-word	6-word
Such as the Factors such as an example of	is an example of	such as the use of	the following is an example of

Table 17. Some most frequent bundles that incorporate a question in the ALC: questioning lexical bundles.

3-word	4-word	5-word
To what extent what are the	is there a relationship	what is the relationship between

Table 18. Some most frequent lexical bundles that refer to cause of something in the ALC.

3-word	4-word	5-word
Because it is	be due to the	due to the fact that
Because they are		due to the fact that the
Because of the		

(24) *Is there a relationship* between explicit and implicit attitudes towards NESTs and non-NESTs?

As it is apparent in extracts (23), and (24) these recurrent expressions are mostly used to point to or initiate research questions. Also, this function seems not to be identified in the literature before.

The last new type of lexical bundles which performed a specific function in the ALC referred to cause of something. Most frequent examples of Cause Initiator Lexical Bundles found in the ALC are presented in Table 18. Following examples are given to show cause initiator lexical bundles in the text:

(25) This could very well **be due to the** nature of the task itself, which focuses on exchanging cultural jokes in a task-oriented environment rather than a social environment.

(26) Those who said they like English very much were more likely to prefer NS accents and NS teachers after the project. This may be **due to the fact that** their preference toward English originated from their interests in Anglo-American cultures: these English lovers were also more likely to agree to the statement, 'to learn English well, one must know a lot about American or British cultures' ($r = 0.341$ $p = 0.012$).

The lexical bundle *be due to the* in extract (25) is used to refer to nature of a specific task in that study as the result of something. The recurrent expression *due to the fact that* in extract (26) is also used in that study to refer to the cause of why some people were more likely to prefer Native Speaker accents and Native Speaker teachers after a specific project. Also, this function of lexical bundles has not been identified in previous theoretical classifications.

5. Conclusion

5.1. Structural and functional characteristics of lexical bundles in the ALC

This study attempted to explore a fairly large corpus of papers published in the field of applied linguistics to identify lexical bundles with different length as well as all the possible structural and functional characteristics of them existing in this register. First of all, it was found that a considerable number of lexical bundles used in applied linguistics articles published from 2008 to 2018 were in accordance with those identified previously in the literature. On other hand, there were some bundles which were not identified before, or performed different functions in comparison to what they were supposed to do in the text based on what the literature suggests. It should also be noted that, besides the structural and functional identities of lexical bundles, their length was also considered in this study. This study showed that lexical bundles of different length can express different structural and functional identities and contribute reasonably to the discourse of applied linguistics academic writing.

Regarding the structural characteristics, lexical bundles which incorporated adjectival phrases (*is beyond the scope of this*, for example), adverbials (*depending on the*, for example), adverb/adjective clauses (*who did not*, for example), be + adjective/adverb + to (*be able to*, for example), infinitive phrases (*to address the*, for example), and those which begin with a conjunction (*but also to*, for example) were newly identified in the ALC which were not identified before in the literature. On the other hand, some lexical bundles expressed new functional characteristics in the ALC: doubling lexical bundles (*at the beginning and end of*, for example), contrasting lexical bundles (*did not seem to*, for example), exemplifiers (*factors such as*, for example), questioning lexical bundles (*what is the relationship between*, for example), and those which referred to cause of something in a text (*be due to the*, for example).

5.2. Length of lexical bundles and their different functions

There were some bundles with varying length in the ALC which do not expand to longer bundles, and even for many of those which do so, their structural identities changed while being expanded to a larger lexical bundle. It means that lexical bundles with varying length contribute reasonably to the field of applied linguistics. To clarify this point, the following extracts indicate the difference between the functions of a lexical bundle and its expansion. *is based on the* as a 4-word lexical bundle is used to point at a logical relation between the propositions in the following extracts, but *is based on the fact that* as its expansion to a 6-word lexical bundle is used to express a high degree of certainty since its following proposition is regarded by its author as a “fact.” This shows that although smaller bundles may seem to expand to larger ones, they are used differently in different texts in the ALC, making bundles with different length important linguistic patterns in applied linguistics academic writing.

(27) As discussed in Section 1.2 above, register research ***is based on the*** claim that all linguistic variation is functionally associated with the situational context [...]

(28) This ***is based on the fact that*** grammar learning is a complex process and relies on both memorization of individual items and processing of the relationships among them.

5.3. Implications

Theoretically, it can be asserted that the new structural and functional subtypes of lexical bundles found in this study are suggested to be taken into account by the researchers who aim to investigate them in published academic writing in the field of applied linguistics in order to achieve more comprehensive findings. Moreover, the researchers interested in the discourse of academic writing can consider doubling lexical bundles, as a newly found subcategory, for further research. They can investigate the frequency and range of lexical bundles together with the concordance lines including these bundles in disciplines other than applied linguistics. Although the lexical bundles along with their structural and functional characteristics found in this study may be applicable to other disciplines as well, the results of this study need to be treated with some caution. As this study has investigated only the specific genre of applied linguistics, the lexical bundles found in the present study may not be generalizable to other disciplines. As explained by Biber, Conrad,

and Cortes (2004), since the structural correlates of lexical bundles are register-bound, lots of variation can be seen across registers in terms of the typical discourse functions.

On the other hand, the vision of the language produced by L2 learners can be created through careful use of lexical bundles as a type of formulaic language. As noted by Boers et al. (2006), L2 speakers can achieve higher levels of speaking proficiency when they employ formulaic sequences in their speech. Moreover, an examination of the EFL learners whose written composition was highly scored showed that they preferred to use formulaic sequences more frequently than those with lower scores (Ohlrogge 2009).

Given the significance of the formulaic language (i.e. lexical bundles), it can be asserted that that syllabuses of language courses can be developed based on lexical-bundles. Although lexical bundles could naturally play a notable role in language tests and teaching textbooks, we should keep in mind that ESL tests or textbooks seem to usually target formulaic sequences only for explicit noticing purposes (giving random lists, for instance) through questionable selection bases (Gouverneur 2008; Hsu 2008).

Therefore, pedagogically, the findings of this study can be used in designing more effective materials. Since the bundles found in this study were of high frequency, shaping the specific discourse of academic writing in the field of applied linguistics, strategic use of the frequent lexical bundles extracted out of the ALC can help writers increase their chances of publishing their manuscripts; therefore, it seems important for writers to know the particular structural and functional characteristics of lexical bundles which contribute to shape and meaning of different writing registers. Accordingly, materials developers can be suggested to consider introducing strategic use of lexical bundles when developing materials. Additionally, around 10% of the whole corpus was captured by the different bundles which highlight the pivotal importance of the role of lexical bundles with different length in contributing to the cohesion and coherence of the texts in applied linguistics academic prose. By more frequent use of lexical bundle-sensitive materials, it can be implied that, writing teachers can devote more time dealing with field-specific bundles to expose their learners to a bundle-driven version of language which can be the beginning of a bundle-driven approach to language teaching. Rather than advanced level students, lower level learners can also benefit from this approach. Since the more exposure to a linguistic pattern will entrench that specific pattern in learners' memories (see Gries and Ellis 2015), this strong fixedness caused by the bundle-driven approach, started at lower levels, can result in a more practical use of lexical bundles in future.

Note

1. Applied Linguistics Corpus.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Peyman Nasrabady holds an M.A. in Teaching English as a Foreign Language from University of Bojnord, Iran. His main research interests are Corpus Linguistics and English for Specific/Academic Purposes.

Majid Elahi Shirvan is an assistant professor of Teaching English as a Foreign Language at University of Bojnord, Iran. His main research interests include psychology of language learning and English for Specific purposes.

Seyyed Ehsan Golparvar is an assistant professor of Teaching English as a Foreign Language at University of Bojnord, Iran. His main research interests include second language writing and English for Specific purposes.

ORCID

Peyman Nasrabady  <http://orcid.org/0000-0002-0596-8550>

Majid Elahi Shirvan  <http://orcid.org/0000-0003-3363-8273>

Seyyed Ehsan Golparvar  <http://orcid.org/0000-0002-2721-5760>

References

- Ädel, A., and B. Erman. 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. doi:[10.1016/j.esp.2011.08.004](https://doi.org/10.1016/j.esp.2011.08.004).
- Altenberg, B. 1998. "On the Phraseology of Spoken English: The Evidence of Recurrent Word-combinations." In *Phraseology: Theory, Analysis, and Applications*, edited by A. P. Cowie, 101–122. Oxford: Oxford University Press.
- Biber, D. 2010. "Corpus-based and Corpus-driven Analyses of Language Variation and Use." In *The Oxford Handbook of Linguistic Analysis*, edited by B. Heine and H. Narrog, 159–191. Oxford: Oxford University Press.
- Biber, D., and F. Barbieri. 2007. "Lexical Bundles in University Spoken and Written Registers." *English for Specific Purposes* 26 (3): 263–286. doi:[10.1016/j.esp.2006.08.003](https://doi.org/10.1016/j.esp.2006.08.003).
- Biber, D., S. Conrad, and R. Reppen. 1994. "Corpus-based Approaches to Issues in Applied Linguistics." *Applied Linguistics* 15 (2): 169–189. doi:[10.1093/applin/15.2.169](https://doi.org/10.1093/applin/15.2.169).
- Biber, D., S. Conrad, and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. New York: Cambridge University Press.
- Biber, D., S. Conrad, and V. Cortes. 2004. "If You Look at ... : Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25 (3): 371–405. doi:[10.1093/applin/25.3.371](https://doi.org/10.1093/applin/25.3.371).
- Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan, and R. Quirk. 1999. *Longman Grammar of Spoken and Written English*. Vol. 2. Cambridge, MA: MIT Press.
- Boers, F., J. Eyckmans, J. Kappel, H. Stengers, and M. Demecheleer. 2006. "Formulaic Sequences and Perceived Oral Proficiency: Putting a Lexical Approach to the Test." *Language Teaching Research* 10 (3): 245–261. doi:[10.1191/1362168806lr195oa](https://doi.org/10.1191/1362168806lr195oa).
- Boo, Z., Z. Dörnyei, and S. Ryan. 2015. "L2 Motivation Research 2005–2014: Understanding a Publication Surge and a Changing Landscape." *System* 55: 145–157. doi:[10.1016/j.system.2015.10.006](https://doi.org/10.1016/j.system.2015.10.006).
- Chen, Y.-H., and P. Baker. 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.
- Cortes, V. 2002. "Lexical Bundles in Freshman Composition." In *Using Corpora to Explore Linguistic Variation*, edited by R. Reppen, S. M. Fitzmaurice, and D. Biber, 131–146. Amsterdam: Benjamins.
- Cortes, V. 2004. "Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology." *English for Specific Purposes* 23 (4): 397–423. doi:[10.1016/j.esp.2003.12.001](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 (4): 391–406. doi: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 (1): 33–43. doi:10.1016/j.jeap.2012.11.002.
- Coxhead, A. 2000. "A New Academic Word List." *TESOL Quarterly* 34 (2): 213–238. doi:10.2307/3587951.
- Csomay, E. 2012. "Lexical Bundles in Discourse Structure: A Corpus-based Study of Classroom Discourse." *Applied Linguistics* 34 (3): 369–388. doi:10.1093/applin/ams045.
- De Cock, S. 2002. "Pragmatic Prefabs in Learners Dictionaries." In *Proceedings of the Tenth EURALEX International Congress, Copenhagen, August 2002, Denmark*, 13–17.
- Durrant, P. 2017. "Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories." *Applied Linguistics* 38 (2): 165–193. doi:10.1093/applin/amv011.
- Esfandiari, R., and F. Barbary. 2017. "A Contrastive Corpus-driven Study of Lexical Bundles between English Writers and Persian Writers in Psychology Research Articles." *Journal of English for Academic Purposes* 29: 21–42. doi:10.1016/j.jeap.2017.09.002.
- Firth, J. R. [1951]1957. "Modes of Meaning, Essays and Studies." In *Papers in Linguistics 1934–1951*, edited by John R. Firth, 190–215. London: Oxford University Press.
- Gledhill, C. 2000. *Collocations in Science Writing*. Tuebingen: Gunter Narr Verlag.
- Gouverneur, C. 2008. "The Phraseological Patterns of High-frequency Verbs in Advanced English for General Purposes: A Corpus-driven Approach to EFL Textbook Analysis." In *Phraseology in Foreign Language Learning and Teaching*, edited by F. Meunier and S. Granger, 223–243. Amsterdam: John Benjamins.
- Grabowski, Ł. 2015. "Keywords and Lexical Bundles within English Pharmaceutical Discourse: A Corpus-driven Description." *English for Specific Purposes* 38: 23–33. doi:10.1016/j.esp.2014.10.004.
- Granger, S. 1998. "The Computer Learner Corpus: A Versatile New Source of Data for SLA Research." In *Learner English on Computer*, edited by S. Granger, 3–18. New York: Addison Wesley Longman.
- Granger, S. 2000. "Optimizing Measures of Lexical Variation in EFL Learner." In *Corpora Galore: Analyses and Techniques in Describing English: Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora*, edited by J. Kirk, 249–257. Amsterdam and Atlanta: Rodopi.
- Granger, S., and F. Meunier, eds.. 2008. *Phraseology: An Interdisciplinary Perspective*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Gries, S. T. 2010. "Useful Statistics for Corpus Linguistics." *A Mosaic of Corpus Linguistics: Selected Approaches* 66: 269–291.
- Gries, S. T., and N. C. Ellis. 2015. "Statistical Measures for Usage-based Linguistics." *Language Learning* 65 (S1): 228–255. doi:10.1111/lang.12119.
- Howarth, P. 1998. "Phraseology and Second Language Proficiency." *Applied Linguistics* 19 (1): 24–44. doi:10.1093/applin/19.1.24.
- Hsu, J. Y. T. 2008. "Role of the Multiword Lexical Units in Current EFL ESL Textbooks." *US-China Foreign Language* 6 (7): 27–39.
- Hyland, K. 2008. "As Can Be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27 (1): 4–21. doi:10.1016/j.esp.2007.06.001.
- Jespersen, O. 1924. *The Philosophy of Language*. London: G. Allen.
- Kashiha, H., and C. S. Heng. 2014. "Discourse Functions of Formulaic Sequences in Academic Speech across Two Disciplines." *GEMA Online® Journal of Language Studies* 14 (2): 15–27.
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1 (1): 7–36. doi:10.1007/s40607-014-0009-9.
- Nattinger, J., and J. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- O'keeffe, A., M. McCarthy, and R. Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Ohlrogge, A. 2009. "Formulaic Expressions in Intermediate EFL Writing Assessment." *Formulaic Language* 2: 375–386.

- Pawley, A., and F. H. Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." *Language & Communication* 191: 225.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. 2004. "The Lexical Item." In *Contrastive Lexical Semantics*, edited by E. Weigand, 131–148. Amsterdam: Benjamins.
- Staples, S., J. Egbert, D. Biber, and A. McClair. 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. doi:[10.1016/j.jeap.2013.05.002](https://doi.org/10.1016/j.jeap.2013.05.002).
- Ward, J. 2007. "Collocation and Technicality in EAP Engineering." *Journal of English for Academic Purposes* 6 (1): 18–35. doi:[10.1016/j.jeap.2006.10.001](https://doi.org/10.1016/j.jeap.2006.10.001).
- Wray, A. 2000. "Formulaic Sequences in Second Language Teaching: Principle and Practice." *Applied Linguistics* 21 (4): 463–489. doi:[10.1093/applin/21.4.463](https://doi.org/10.1093/applin/21.4.463).
- Wray, A. 2002. "Formulaic Language in Computer-supported Communication: Theory Meets Reality." *Language Awareness* 11 (2): 114–131. doi:[10.1080/09658410208667050](https://doi.org/10.1080/09658410208667050).
- Wray, A. 2008. *Formulaic Language: Pushing the Boundaries*. Oxford, UK: Oxford University Press.
- Wray, A., and M. R. Perkins. 2000. "The Functions of Formulaic Language: An Integrated Model." *Language & Communication* 20 (1): 1–28. doi:[10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4).