Teaching English as a Second Language Quarterly (TESLQ)
(Formerly Journal of Teaching Language Skills)
41(2), Spring 2022, pp. 37-74
DOI: 10.22099/TESL.2021.39319.2928

Online ISSN: 2717-1604
Print ISSN: 2008-8191

Research Paper

Types and Tokens of Lexical Bundles in Civil Engineering Students' Genre-Specific English Articles

Hamid Reza Baba Ahmadi [*] **Hamed Babaie Shalmani** [**]

Department of English Language, College of Humanities, Rasht Branch, Islamic Azad University, Rasht, Iran

Abstract

The present study sought to find the most frequent lexical bundles, particularly the token of the bundles, in the introduction and discussion sections of Civil Engineering research articles. It also aimed to examine the forms or types of lexical bundles found in the introduction and discussion sections of the articles. To this end, a quantitative analysis was performed on the use of two-to-five-word lexical bundles, followed by a qualitative analysis of their functions and structures based on the structural taxonomy of lexical bundles proposed by Biber, Conrad, and Cortes (2004). AntConc, as a corpus analysis tool, was employed for the extraction of lexical bundles as well as the examination of their types and tokens. Likewise, concordancers and word and keyword frequency generators were used for the analysis of word clusters and lexical bundles. A total of 790 lexical bundles in the introduction section and 279 lexical bundles in the discussion section were analyzed based on a set of pre-established criteria (e.g., the frequency cut-off of 20 per million words). The findings revealed that the lexical bundles consisted largely of two-word strings, and the five-word strings were the fewest. Regarding the structural classification of bundles, phrasals were the most frequent ones, whereas clausal ones were the least frequent ones. The study carries important implications, especially for materials developers and course designers who may

Received: 15/03/2021 Accepted: 11/12/2021

[*] M.A. in TEFL, email: babaahmadi.h.r@gmail.com
[**] Assistant Professor of TEFL, email: babaie@iaurasht.ac.ir, Corresponding Author

hopefully benefit from the results in designing EAP materials finely tailored to the linguistic needs of Civil Engineering students.

***Keywords**:* AntConc, EAP, Genre Analysis, Lexical Bundles

Lexical bundles are "the most frequently occurring lexical sequences in a register" (Biber et al., 2004, p. 23) or sequences of words that are syntactically and semantically compositional (Jalali & Moini, 2014). Based on Biber and Conrad's (2001) classification, a frequency-driven approach is employed to analyze the most prevalent and frequently occurring lists of fixed structures or word associations. These compounds are termed lexical bundles to convey the concept of word combinations that appear consistently, with a moderately high recurrence, inside a specific genre or register. Unlike non-compositional idioms (whose meaning cannot be inferred from that of the constituent parts), lexical bundles are semantically vivid and straightforward, and their importance can be acknowledged from their segments. Another difference between these two types of lexical structures is that the former is undeniably less frequent in written and oral discourse (Biber & Conrad, 2001).

Hyland (2008b) defines lexical bundles as multiword expressions known as clusters or chunks. They are the most frequently occurring patterns, including three or more words that are generally neither colloquial nor complete forms of meaning even though they have solid linguistic associates (Biber & Barbieri, 2007; Biber, Conrad, & Reppen, 1999). Some typical instances of lexical bundles include expressions like *on the one hand*, *in conclusion, extensive research into, an*d *for the purpose of*.

Since lexical bundles are the core components of everyday speech and written texts, mastering these formulaic units of language, which are extensively used in academic genres, is vital to academic fluency; however,

these word combinations are diverse in different fields of studies and various parts of writing. Several studies have been conducted to examine this diversity (e.g., Hyland, 2012; Pérez-Llantada, 2014). Also, José and Marco (2000) argue that linguistic structures and elements are quite different in various genres, and we need corpus-based analysis to discover them.

Undoubtedly, different fields benefit from varieties of lexical bundles; therefore, university students should become familiar with these bundles to better understand research articles and write and produce research articles. To effectively use lexical bundles in academic discourse, understanding the constituent elements and the contexts of use plays a significant role (Biber et al., 1999).

It is now common among higher education students and professors that research articles play a significant role in their scientific growth. Furthermore, as Staples and Reppen (2016) noted, with a growing number of L2 writers studying at the university level, the question of how to address their linguistic needs has become increasingly important. Being familiar with the genre of research articles is one of the goals many students need to pursue in different fields of study. Knowing the lexical units associated with a particular genre and the features of their structures can help them achieve fluency in that genre.

Academic writing involves different tasks that include writing assignments, article writings, term projects, case studies, technical and laboratory reports, as well as thesis writing. The students of engineering must promote their academic writing; nevertheless, university professors are sometimes not satisfied with their students' performance in writing classes, especially in academic writing tasks. Teachers and professors always complain about the poor writing skills of engineering students, especially

those of their graduates. In general, these students have difficulties in three major areas: content, structure, and language.

Katnic-Bakarsic (2004) likewise argues that discourse of academic genres is a dismissed style in applied linguistics and that it is regularly viewed as a totally impartial class of talk. The phrasing of scholarly discourse could imply an expansive scope of oral and composed types in which a portion of the content might be nearly a discourse employed for administration, interaction, news agency correspondences, and compositions. It may comprise contention, strains, and inconsistency (e.g., scholarly conversations, doctoral papers, and colloquia).

Article and essay writing is regarded to be one of the most demanding academic requirements for a host of students doing their M.A. or Ph.D. studies; however, this onerousness worsens when it comes to writing in a second language. The growing number of second language writers notwithstanding, many students find meeting the requirements of article writing the most arduous academic task. Second language writers have remarked that they did not appear to have built up a similar way of data presentation in articles and other academic essays, which are mostly regulated and reviewed by English reviewers and scholars (Sadeghi & Khajehpasha, 2015). University students and second language writers are not, by any means, the only ones who are enduring in such matters. In English for Academic Purposes (EAP) classes, teachers have practiced similar opinions as they battled over second language writers' obscure and questionable structures.

According to Schmitt (2010), vocabulary is regarded as a cornerstone of language learning. It plays a significant role in students' ability to take part in academic discourse communities. Students must be familiar and somehow become experts in manipulating the language of the academia and the

terminologies of their disciplines to be able to understand scholarly works and be understood by others (Wright, 2019). Academic language should be learned through extensive reading and studying of the written and even spoken texts produced by the experts and skilled members of that specific genre.

Hyland (2012) contends that academic proficiency is fundamental to each movement done in a college and that expert types of scholastic knowledge are the only thing that is important in an academic organization. He argues that instructors should help understudies build their familiarity with the norms of scholarly composition to have the option to adequately build up their learning, set up their professions, and further their comprehension of their discourse community.

Along the same lines, Ren (2021) argued that analysis of formulaic patterns could also take academic disciplines and discourse functions into account. He also noted that lexical bundles and their counterpart formulaic frames could be used as valuable instruction materials for graduate-level academic writing classes. Novice writers should be capable of using the common lexical bundles of a particular discourse community (Wray, 2005), but it has been found that incidental acquisition of language formulas is likely to be ineffective (Boers & Lindstromberg, 2012). Therefore, overt teaching of lexical bundles is necessary for the apprentice writer to grasp these linguistic constructs. Shin and Kim (2017) likewise conducted an experiment and examined the potential for teaching articles using lexical bundles with adult English language learners of varying proficiency and found that lexical-bundle-based article instruction could be beneficial for promoting accurate article use by L2 learners.

Yin and Li (2021) investigated the intradisciplinary and interdisciplinary variations of research articles with a lexical bundle approach and concluded

that explicit teaching of lexical bundles plays a key role in the correct understanding of disciplinary features. It was found that lexical bundles could be used as an intradisciplinary mark between the two business-type disciplines: Distinctive bundles such as resultative expressions could distinguish finance and accounting articles even though shared bundles such as topic-related strings also reflect potentially characteristic features of business disciplines. In addition, the results suggest that the bundle usage as an interdisciplinary mark could indicate obvious variations between business, biology, and applied linguistics articles. For instance, the business research articles contain many lexical bundles, suggesting that the business disciplines may have a stronger phraseological behavior than the other two disciplines.

Review of Related Literature

The interdisciplinary investigation of academic discourse specifically deals with fulfilling the communicative purposes of academic and non-academic communities. It is also worthy of mention that this field seeks paths to success in interactions between discoursal participants in well-established cultures (Bennet, 1991). Academic discourse affords scholars the golden opportunity to have access to syntactic and lexical assets that are essentially required to fully interpret the discourse accommodating complete structures that further give writers the chance to construct well-formed written and oral discourse within a rich context that demands language use Katnic-Bakarsic, 2004). This developed discourse pays dividends owing to its rhetorical features enriching the writing genre, the degree of expressive norms manifest, and other features that help the interpretations or presentation of texts.

In certain studies (e.g., Hyland, 2007; Katnic-Bakarsic, 2004), the language employed in academic genres represents a sense of power and

supremacy. In other words, academic discourse is regarded as the language of authority and power, which differentiates the academic genre from other discourses in a wide range of cultures and brings about the formation of purposeful interaction, especially between authors and information receivers.

Corpus findings have indicated that both conversation and academic texts benefit a significant number of lexical bundles, although they are rarely used or even misused by university students (Bychkovska & Lee, 2017; Cortes, 2013). Biber and Barbieri (2007) argued that lexical bundles are even more common in this register than they are in the instructional registers. Other studies have shown that knowing holistically recurrent sequences of words as chunks could lower the processing demand for the user (Siyanova, Conklin, & van Heuven, 2011; Tremblay, Derwing, Libben, & Westbury, 2011). Cortes (2004) also asserted that these expressions are not acquired naturally.

Ellis, Simpson-Vlach, and Maynard (2008) assert that every genre has certain expressions, and mastering the genre involves mastering these expressions. Most studies over these decades have been conducted in a specific corpus and tried to investigate different lexical bundles and their similarities and differences in various disciplines, fields, genres, and different parts of writing (Cortes 2004; Cortes 2013; Herbel-Eisenmann & Wagner, 2010; Jalali & Moini, 2014). Biber et al. (2004) studied lexical bundles and compared bundles found in the registers. The study showed that the grammatical structure of lexical bundles is a unique feature of registers. Also, Hyland (2008a) identified lexical bundles in the genres of academic papers, master theses, as well as Ph.D. dissertations in four disciplines. He concluded that different types and disciplines genres drew on unique lexical bundles.

Jalali and Moini (2014) studied the introduction sections of medical research papers for the structure of lexical bundles and concluded that authors

of medical research papers depend on noun phrases for writing articles. Jalali and Zarei (2016) studied the use of lexical bundles qualitatively in higher education students' writing and found that they can use lexical bundles as well as professional writers. Rafiee and Keihaniyan (2013) investigated two broad corpora of journalistic writing and more than two million words and found out referential bundles were the most common type in the journalistic genre.

Recent corpus-based studies have discovered that there are syntactically and semantically compositional groupings for EAP-specific words. These combinations of words are constructed based on a specific EAP domain or some technical words that represent the rhetorical and discoursal functions highly projected in different sections of academic papers such as the introduction, hypotheses, conclusions, summaries, etc. Surely, one of the most significant features and characteristics of academic discourse is lexical bundles (Jalali & Moini, 2014).

Over the past 20 years, numerous teachers and specialists have started using jargon terms in second language schools (Folse, 2004). Immense electronic assortments of genuine and authentic language known as corpora offer progressed arrangement and distinguishing proof of subtle structures in the English language that may likewise be utilized to pervade other languages (Granger & Paquot, 2008; Stubbs, 2007). Yet, nobody would dispute the sparseness of research into this intricate issue, although scholars in the realm of corpus studies can present more data concerning this matter.

Broadly speaking, one of the most fundamental features of corpus studies is the analysis of the recurrence, which is commonly known as the frequency that gives an utterly quantitative design to the research conducted in the area of corpus linguistics (Hunston, 2006). Different comparative studies on frequency can be carried out from numerous points of view. For

example, the frequency of lexical items in oral or written language productions or those used by native and non-native discourse communities can be obtained. Interestingly, corpus specialists have extensively heeded the latter in recent years (Chen & Baker, 2010; Granger & Paquot, 2008; Siyanova & Schmitt, 2008; Wang & Shaw, 2008).

Phrasal verbs inter alia other word combinations (especially the lexical verbs phrased in written discourse) have great potential to improve a corpus field in terms of information assemblage and data collection. The investigation of corpus produces invaluable bits of knowledge on perceiving lexical bundles as accommodating syntactic features that are significantly useful in applied linguistics and for second language fulfillment (Gass & Selinker, 2001).

The argument that instruction and employment of chunks, bundles, and prefabricated patterns can transform a novice writer's performance into an authentic one is highly tenable (Kjellmer, as cited in Nesselhauf, 2005). The formulaic discourse is believed to harness the students' writing skills, providing it is taught and learned purposefully with minimal alertness. Students' use of bundles should be coupled with consciousness and appropriateness so that they find their language production similar to that of native speakers. Durrant and Schmitt (2010) found that the written and spoken productions of second language learners who use high-recurrence collocations share very few features with those of their counterparts that do not use them. This is because when they use high-recurrence word associations, they cannot employ less commonly used lexical compounds the native speakers find notable and effective in their communication. This argument is in line with the "use-based models of procurement" by Durrant and Schmitt (2010, p. 21), which contends that second language writing must accommodate the formulaic discourse.

Menon and Mukundan (2012) warned about the overrepresentation of these lexical patterns when they unanimously suggested that not only should students be instructed on the most frequently used lexical patterns, but they should also be taught how to use them appropriately concerning the context in which they occur. They noted that the adaptability of certain examples can be self-assertively impeded by utilization and that these subjective lexical compounds represent a huge deterrent for students. In any case, these investigations have, at most, uncovered the overall agreement that students who have little or no knowledge of these prefabricated patterns are not sufficiently proficient and fluent second-language speakers and writers.

Drawing on Swales' (2004) Create A Research Space (CARS) model, researchers analyzed the patterns of use for lexical units that appeared in academic papers for specific genres. The findings showed that native English writers employed more strategies than Iranian speakers of English, yielding richer texts.

Alipour, Jalilifar, and Zarea (2013) analyzed the rhetorical organization of the lexical bundles found in the introduction sections of articles published in Iranian domestic and foreign international journals and in relation to three different disciplines, namely English for Specific Purposes (ESP), Discourse Analysis (DA), and English for General Purposes (EGP). One hundred and twenty sections were examined in the study. Some consistency in the international corpus was observed concerning variations across sub-disciplines in both corpora. For example, the introduction sections in the international journals showed differences in the utilization of bundles. In the generic organization, the intra-sub-disciplinary variations were also identified within sub-disciplines.

In yet another study by Alipour et al. (2013), lexical bundles were identified and compared in the genre of research papers in three fields of study, namely Computer Science, Applied Linguistics, and Physics. The results showed that significant variations exist in the structures and functions of the bundles in these disciplines and that the writers of these disciplines rely on different norms to communicate appropriately with the members of their communities.

Parvizi (2011) performed a genre analysis on the introduction section of dentistry research papers written in English and Persian in a seminal study. The author examined 70 sections using Swales' (2004) classification of moves and steps used in the introduction section of articles. The findings showed a statistically significant difference in certain moves used in the introduction section of English and Persian research papers.

Lexical bundles that are recurrent multiword sequences and generally incomplete play a very significant role in academic literacy and corpus-based studies. Biber and Barbieri (2007) emphasize that "failure to understand their (lexical bundles) textual and interpersonal functions will influence student success in dealing with both spoken and written language situations" (p. 284). Therefore, understanding these recurring sets of words is vital, considering that they are the major factors in conveying meaning (Sinclair, 2008). Since lexical bundles are specific to a particular corpus, it would be very important to pay attention to these combinations of words to be successful in a discourse community.

Biber and Barbieri (2007) also noted that "each register employs a distinctive set of lexical bundles associated with the typical communicative purposes of that register" (p. 265). Whereas a growing body of research has corroborated the view that teaching genre-specific lexical bundles holds great

promise for university students to meet the standards required of them to be successful in their academic studies, very few studies, if any, have sought to identify the most frequent and hence the most useful bundles students need to master in relation to the themes and concepts associated with their fields of study. To become successful in their academic studies, students need to get familiar with genre-specific jargon terms when discussing spoken and written discourse issues: Sometimes, they are required to give oral presentations on discipline-specific topics in lecture-based classes. Other times, they are required to elaborate on genre-specific themes through composition and essay writing activities. Effective organization of ideas is indispensable to presenting a fascinating lecture or writing an efficient essay, which can be partly accomplished through the mastery of word combinations that the experts in the field commonly use. Command of lexical bundles is assumed to aid in the efficient structuring of ideas in spoken and written communication, such that when mingled with one's knowledge of genre-specific terminologies, it could serve as a unified knowledge base for university students, helping them to achieve great structural coherence in their oral and written assignments.

Inspired by these assumptions, the present study sought to identify, classify, and describe features of the lexical bundles that are commonly found in the written productions of Civil Engineering students in an attempt to provide a solid knowledge base for both materials developers to design materials tailored to the linguistic needs of these students and the students themselves to achieve greater fluency and also improved efficiency when doing their oral and written assignments. It was hypothesized that a thorough examination of these lexical bundles could allow for a deeper analysis of the types and tokens of genre-specific word combinations typically favored and

used by expert writers in the field. To this aim, English articles written by the students were used as the sample in genre analysis.

There are also three reasons why Civil Engineering students were selected as the population whose research papers were analyzed in this study: First, the written productions of these students could be readily accessed by the researchers of the study, given the fact that one of the researchers was a Civil Engineering graduate, had taught the students of this field, and could easily access their research articles, accordingly. Second, no study to date has closely examined the linguistic needs of Civil Engineering students to provide a knowledge base for course designers to produce needs-based materials for these students or for the students to achieve great fluency and efficiency in their academic performance by mastering the most useful lexical units in their field. Third, academic writing involves doing different tasks that typically include writing assignments, article writing, writing term projects, doing case studies, preparing technical and laboratory reports, as well as thesis writing. Engineering students are expected to promote their academic writing skills to become successful in their academic studies. Part of the requirements for their studies is how well they can prepare a coherent written report on their normal skill-based activities. Yet, university professors sometimes are not satisfied with their students' performance in writing classes, especially with their academic writing tasks. Teachers and professors always complain about the poor writing skills of engineering students, especially those of their graduates. In general, these students have difficulties in three major areas: content, structure, and language.

Specifically, then, the present study aimed to categorize the most frequently used lexical bundles used by the writers of English articles in the field of Civil Engineering by identifying the token of the most frequent

bundles in the corpora as well as exploring the forms or types of these bundles that the authors typically employ in the field. However, the sample chosen for corpus analysis was limited to those bundles that frequently appeared in the introduction and discussion sections of the articles. This is because these two sections provide the most information about the topic and hence are expected to feature a wide array of the bundles used in the corpora. The other sections provide a little amount of text and generally comprise numerical and mathematical data not typically used in genre analysis.

In line with the overriding aims of the present study, the following questions were formulated to guide this research:

1. What are the most frequent lexical bundles in Civil Engineering research articles' introduction and discussion sections?
2. What forms of lexical bundles are typically used in Civil Engineering articles' introduction and discussion sections?

Method

Design

The present study drew on a quantitative research design to explore the use of two-to-five-word lexical bundles and a qualitative investigation of their functions and structures. In particular, the researchers adopted Biber et al.'s (2004) structural taxonomy of lexical bundles (see Appendix for a copy of the taxonomy). Corpus studies can investigate two categories known as *corpus-driven* and *corpus-based* research. As the present study did not limit its scope to a predefined category of lexical bundles, it is regarded as an essentially corpus-driven study: The text is seen as an integral part of its verbal context, and, finally, no discontinuity is presumed between this and the broader context of situation and culture (Tognini-Bonelli, 2001).

Materials

The materials used in the present study consisted of data or corpora that were collected based on the content of journal articles published by experts in the Civil Engineering field. AntConc is a corpus analysis tool that was used in this study for the extraction of lexical bundles as well as for the calculation of their types and tokens. The tool also consisted of concordance or word and keyword frequency generators used to analyze word clusters and lexical bundles. These all are explained in further detail below.

Data. The corpora comprised 30 scientific, academic research articles published by *ScienceDirect* in the field of Civil Engineering within a given time interval over the past eight years (from 2012 to 2020). The scientific articles were gathered from distinguished ScienceDirect engineering journals because, as Reller (2014) argues, this database is one of the most prominent and prestigious databases. All the engineering research articles included in the corpora were downloaded from this database.

AntConc. AntConc consists of a concordancing tool, word, and keyword frequency generator for clustering words and analyzing lexical bundles. It also features a word distribution plot that diagrammatically shows the distribution of words in a given corpus. As a freeware multi-platform application that makes it ideal for individual authors, schools, or colleges with a limited budget, AntConc can be used in either Windows- or Linux/Unix-based operating systems. Furthermore, it is a single executable file that can be merely copied and launched on a computer (Anthony, 2004, 2013).

Procedures

A specialized corpus was selected for the current study. To this aim, research papers published in journals specializing in the field of Civil Engineering were chosen from peer-reviewed journals by considering three factors, namely topic, text type, and author profile, as recommended by Salazar (2014). Simply put, the corpus included the scholarly papers that were authored by scholars in English and in the field of Civil Engineering. The size of the corpus in this paper was determined as one million, and each sub-corpus consisted of roughly 500.000 words, which is the same as the size of the corpus used by Pan, Reppen, and Biber (2016).

This study aimed to make a list of lexical bundles; therefore, the list of lexical bundles needed to be manually inspected so that the researchers could exclude irrelevant and meaningless word combinations. To this end, once the articles were selected, the PDF files of the papers were converted into a plain text file format that would be recognized by the application and could help with a detailed analysis. Also, extra information and details in the research articles like tables, authors' names, interview quotes, figures, and page numbers were removed so that the ultimate list could only contain a meaningful set of bundles based on the criteria employed for their inclusion. The exclusion criteria proposed by Salazar (2014) served as a guide for the study in weeding out irrelevant word combinations. The modified criteria are presented in Table 1. After applying the exclusion criteria, the list of eligible lexical bundles was compiled to extract lexical frames from the relevant lexical bundles.

Table 1

Exclusion Criteria for Lexical Bundles

Bundles	Examples
Bundles consisting acronyms	*GDP, per capita, OECD*
Bundles with random numbers	*at least, one*
Random section titles	*Fig 1 b, Table 2 in*
Meaningless bundles	*it, that is, studies eg*
In-text citations	*Beck et al., Gatignon Anderson, 1988*

The current study aimed at finding two-to-five-word lexical bundles. To adhere to well-established conventions, the researchers chose the Biber et al.s' (2004) frequency approach. They proposed that lexical bundles ranging from three to six words can be researchable, not those with fewer than three or those comprising over seven words.

The software, AntConc (Anthony, 2013), was used to retrieve lexical bundles automatically by allowing the user to choose from among the available options. This corpus analysis tool identifies the lists of N-grams (lexical bundles) by referencing two statistics: frequency and Mutual Information (MI). As recommended in the literature, the minimum cut-off frequency value and MI score were set at 20 times per million words and 3.00 and above, respectively.

As can be seen in Figure 1, the N-Grams tool returned 83448 types and 98704 tokens of lexical bundles for the introduction. The frequency was set to five, and the same type of lexical bundles had to exist in three different articles at a minimum. Both criteria were set to allow for relatively frequent yet evenly distributed lexical bundles.

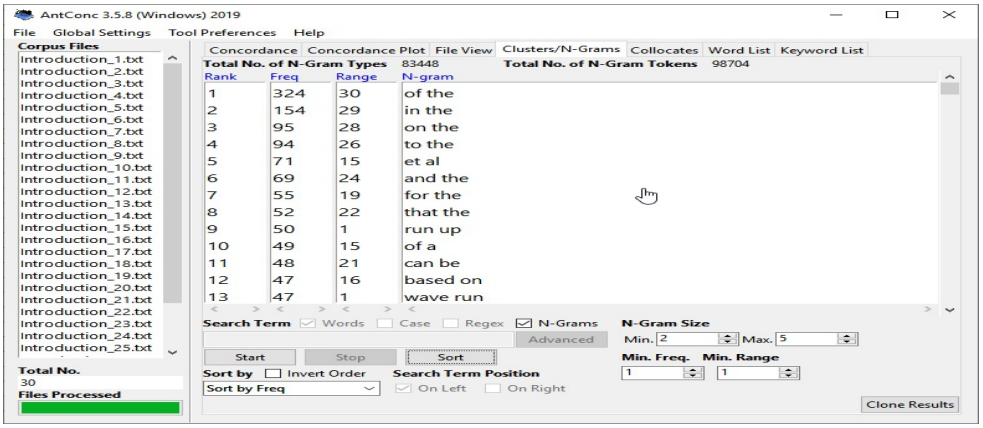

Figure 1

The N-Grams tool output, showing the types and tokens of lexical bundles identified in the introduction section of the articles

Following the extraction of the bundles, the next step was to check the dispersions of the word combinations in the corpus. Based on the literature, a phraseological sequence has to occur in three to five texts (Biber & Barbieri 2007) or 10% of texts to avoid the idiosyncrasies of particular writers (Hyland, 2008b). The present study determined that word combinations that occurred in at least 10% of the texts had to be kept in the output list accordingly.

The N-Grams tool returned 83448 types and 98704 tokens of lexical bundles in the introduction and 42385 types and 48012 tokens of bundles

identified in the discussion section of the articles. The frequency was set to five, and the same type of lexical bundles had to exist in three different articles at a minimum. Both criteria were set to allow for relatively frequent yet evenly distributed lexical bundles.

The type-token ratio (TTR) of the corpora in the present study was based on Biber's (2006) method; however, instead of using one million, as he did, 100000 was used for many of the analyses in this study, as the target size of the individual discipline corpora. Moreover, the smaller normalized number allows for variation in the TTR to be more obvious; the smaller the corpus typically, the greater the TTR and vice versa, and TTR is often used as a measure of lexical variation in which the higher the TTR, the greater the diversity of words (Covington & McFall, 2010). Once the normalized number of types was established, different sections of the corpora could be compared.

Results

As noted, the cut-off frequency of 20 per million words and distribution number of incidences (range) in five different articles were used as a selection criterion. The N-Grams tool returned a total number of 83448 types and 98704 tokens for identified lexical bundles in the introduction section and 42385 types and 48012 tokens for the discussion section of the articles. Yet, 790 types of lexical bundles in the introduction section and 279 types of lexical bundles in the discussion section remained on the list, following the exclusion criteria by which the frequency of lexical bundles was set to five. Having employed the exclusion criteria, the researchers obtained the list of eligible lexical bundles so they could extract lexical frames from the list of relevant lexical bundles just identified.

The Most Frequent Lexical Bundles

The list of lexical bundles mainly consisted of two-word strings, which accounted for 77% or 610 of the total 790 excluded target bundles in the introduction section and 79.5% or 222 of the lexical bundles identified in the discussion section (see tables 2 and 3 below).

Table 2

The First 20 Most Frequent Lexical Bundles in the Introduction Section of the Articles

Rank	Frequency	Lexical bundles
1	324	of the
2	154	in the
3	95	on the
4	94	to the
5	71	as the
6	69	and the
7	55	for the
8	52	that the
9	50	run up
10	49	of a
11	48	can be
12	47	based on
13	47	wave run
14	47	wave run up
15	40	by the
16	40	due to
17	34	to be
18	32	it is
19	32	such as
20	31	as a

Table 3

The First 20 Most Frequent Lexical Bundles in the Discussion Section of the Articles

Rank	Frequency	Lexical bundles
1	210	of the
2	97	in the
3	57	on the
4	47	to the
5	43	for the
6	38	can be
7	32	and the
8	32	that the
9	28	by the
10	27	as a
11	26	with the
12	21	from the
13	20	based on
14	20	in this
15	20	of a
16	20	the project
17	19	additive construction
18	19	to be
19	18	should be
20	17	the results

As can be seen, the tables display the most frequent two-word lexical bundles found in the corpus that were arranged in the descending order of normalized frequency (per million words = pmw).

Total Frequency of the Lexical Bundles

Out of 790 lexical bundles identified in the introduction section of the articles, 610 (77.2 %), 138 (17.4 %), 30 (3.7 %), and 12 (1.5 %) of the bundles

belonged to the two-word, three-word, four-word, and five-word categories, respectively (see Table 4 below).

Table 4

Total Frequency Reported for the Lexical Bundles Identified in the Introduction Section of the Articles

Type of Lexical Bundles	Frequency	Percent
Two-Word Bundles	610	%77.215
Three-Word Bundles	138	%17.468
Four-Word Bundles	30	%3.797
Five-Word Bundles	12	%1.518

Similarly, of 279 lexical bundles that were found in the discussion section of the examined articles, 222 (79.5 %), 46 (16.4 %), 8 (2.8 %), and 3 (1.0 %) lexical bundles belonged to the two-word, three-word, four-word, and five-word lexical bundles, respectively (see Table 5 below).

Table 5

Total Frequency Reported for the Lexical Bundles Identified in the Discussion Sections of the Articles

Type of lexical bundles	Frequency	Percent
Two-Word Bundles	222	%79.569
Three-Word Bundles	46	%16.487
Four-Word Bundles	8	%2.867
Five-Word Bundles	3	%1.075

Table 6 displays the most frequent lexical bundles in the introduction sections of the articles in terms of word strings classified as one-word, two-word, three-word, four-word, and four-word strings.

Table 6

Examples of Lexical Bundles Identified in the Introduction Section and in Terms of Word Strings

Two-Word Bundles	Three-Word Bundles	Four-Word Bundles	Five-Word Bundles
of the	based on the	wave run up on	run up on vertical piles
in the	plastic dilation rate	the plastic dilation rate	wave run up on vertical
on the	one of the	run up on vertical	of wave run up on
to the	the use of	up on vertical piles	the vertical part of ground
and the	in order to	of wave run up	ground to roof conversion factor
for the	on vertical piles	reconstruction of historical buildings	it should be noted that
as the	the plastic dilation	regular wave run up	regular wave run up on
run up	to improve the	for the prediction of	run up on single piles
can be	according to the	steel plate shear wall	vertical part of ground motions
based on	the effect of	the vertical ground motion	wave run up on single

As can be seen, they are far more limited, especially in the four-word and five-word strings categories. It can be observed that there are only 7 four-word strings and 3 five-word strings, which are far fewer than strings of the same type identified in the introduction section of the articles.

Table 7

Examples of Lexical Bundles Identified in the Discussion Section in Terms of Word Strings

Two-Word Bundles	Three-Word Bundles	Four-Word Bundles	Five-Word Bundles
of the	based on the	a result of the	as a result of the
in the	of the project	allowable individual sliding distance	run up on single piles
on the	the influence of	corrugated steel plate concrete	wave run up on single
for the	first passage probability	regular wave run up	
can be	in this paper	run up on single	
and the	the use of	up on single piles	
that the	as well as	wave run up on	
by the	heat saving potential		
as a	in this study		
for the	of coupling beams		

Structural Classification of Bundles in the Introduction Sections

In the latter phase of analysis, the lexical bundles were structurally categorized using the taxonomy Biber et al. (2004) proposed. To analyze the distribution of a variable, the data were organized according to the turn-out of different results in each category. As for categorical variables, the frequency distributions in the introduction section are presented in Table 8 below.

Table 8

Structural Classification Reported for the Lexical Bundles Identified in the Introduction Section

Rank	Category	Frequency of Bundles	Percentage
1	Noun phrase	191	24.17%
2	Adjective phrase	117	14.81%
3	Prepositional phrase + of	99	12.53%
4	Other prepositional phrase	74	9.36%
5	Preposition = articles (a. an, the)	38	4.81%
6	Noun phrase with other post modifier fragments	37	4.68%
7	Other expressions	36	4.55%
8	Verb phrase + (noun)	34	4.30%
9	Passive verb phrase	30	3.79%
10	Verb phrase+that-clause	30	3.79%
11	Connectors and conjunctions	26	3.29%
12	Noun phrase with prepositional phrase fragment	24	3.03%
13	Infinitives + noun phrase	22	2.78%
14	Modal verbs	18	2.27%
15	Reductions (passive / active)	14	1.77%

The table shows the frequency and percentage of bundles in each major category, and sub-categories are given concerning their rank and frequency. In the introduction section of the articles, about 51.51% of lexical bundles are phrasal, while clausal bundles form about 31.79% of lexical bundles. The remaining percent of lexical bundles belonged to other types, such as preposition + articles (a. an, the), modal verbs, reductions, connectors and conjunctions, and other expressions with 4.81% and 2.27 % 1.77%, 3.29%, and 4.55%, respectively.

Among those bundles recognized as phrasals, noun phrases were the most recurrent ones. They had the highest frequency with 24.17% of the whole lexical bundles in the introduction section, and adjective phrases and prepositional phrases ranked second and third with 14.81% and 12.53%. However, the lowest frequency in the introduction section could be counted for infinitives + noun phrases, modal verbs, and reductions (passive/active) that gaining 2.78%, 2.27%, and 1.77%, respectively. Table 9 presents examples of each category found in the introduction section of the articles.

Table 9

Examples of Individual Categories of Lexical Items Identified in the Introduction Section

Categories	Example(s)
Noun phrase	the project team, the design
Adjective phrase	passive confinement, historical buildings
Prepositional phrase + of	mechanical properties of, the concept of
Other prepositional phrases	according to the, for this
Preposition = articles (a. an, the)	of the art, of the plastic, performance of the
Noun phrase with other post modifier fragments	the project team, air quality in
Other expressions	part of ground, in terms of
Verb phrase + (noun)	to achieve, to develop
Passive verb phrase	are summarized, been carried out
Verb phrase+that-clause	concluded that, showed that the, it means that
Connectors and conjunctions	Since, therefore
Noun phrase with prepositional phrase fragment	the cross section of the effects, velocity stagnation head of the
Infinitives + noun phrase	To open transport, to roof conversion factor
Modal verbs	it should be, it would, it should be noted
Reductions (passive / active)	related to the

Similarly, Table 10 below displays the structural classification of the bundles identified in the discussion section of the articles. The frequency and percentage of bundles in 15 major categories and sub-categories are presented concerning their rank and frequency.

Table 10

Structural Classification of Lexical Bundles Identified in the Discussion Section

Rank	Category	Frequency of bundles	Percentage
1	Noun phrase	79	28..32%
2	Adjective phrase	37	13.26%
3	Prepositional phrase + of	35	12.54%
4	Other prepositional phrases	28	10.03%
5	Connectors and conjunctions	13	4.65%
6	Noun phrase with other post modifier fragments	12	4.30%
7	Other expressions	11	3.94%
8	Verb phrase + (noun)	10	3.58%
9	Passive verb phrase	10	3.58%
10	Preposition + articles (a. an, the)	9	3.22%
11	Reductions (passive / active)	8	2.86%
12	Noun phrase with prepositional phrase fragment	8	2.86%
13	Verb phrase+that-clause	7	2.50%
14	Infinitives + noun phrase	7	2.50%
15	Modal verbs	5	1.79%

As can be seen in the table, in the discussion section of the selected articles, about 84.2% of lexical bundles are phrasal, and only 2.50% formed the clausal bundles as the second-lowest bundles in the discussion section. The remaining number of lexical bundles comprised preposition + articles (a. an,

the), modal verbs, reductions, connectors and conjunctions, and other expressions with 3.22%, 1.79 %, 2.86%, 4.65%, and 3.94%, respectively.

As for the bundles that appeared in the introduction section, noun phrases were the most recurrent bundles among the phrasal bundles, as they had the highest frequency with 28.32% of the whole lexical bundles that appeared in this section of the articles, and the adjective phrase and prepositional phrase ranked the second and the third with 13.26% and 12.54% shares of the bundles, respectively. However, the lowest frequency in the discussion section was observed for the categories of verb phrase+that-clause, infinitives + noun phrase, and modal verbs with 2.50%, 2.50%, and 1.79% shares of the total frequency, respectively. Table 11 below presents examples of each category found in the discussion section of the articles.

Table 11

Examples of Individual Categories Identified in the Discussion Section

Category	Example(s)
Noun phrase	the influence, the reliability
Adjective phrase	the long term, the experimental results
Prepositional phrase + of	the influence of, effect of
Other prepositional phrases	in this paper, of the project
Connectors and conjunctions	Therefore, as well as, in addition
Noun phrase with other post modifier fragments	steel plate concrete, the time dependent
Other expressions	the first, according to
Verb phrase + (noun)	found that wave
Passive verb phrase	can be drawn,
Preposition + articles (a. an, the)	as a result of the,
Reductions (passive / active)	presented in, using a
Noun phrase with prepositional phrase fragment	wave run up on
Verb phrase+that-clause	showed that
Infinitives + noun phrase	To show the strength
Modal verbs	will be, could be

Figures 2 and 3 likewise graphically illustrate the lexical bundles identified in the introduction and discussion sections of the articles.

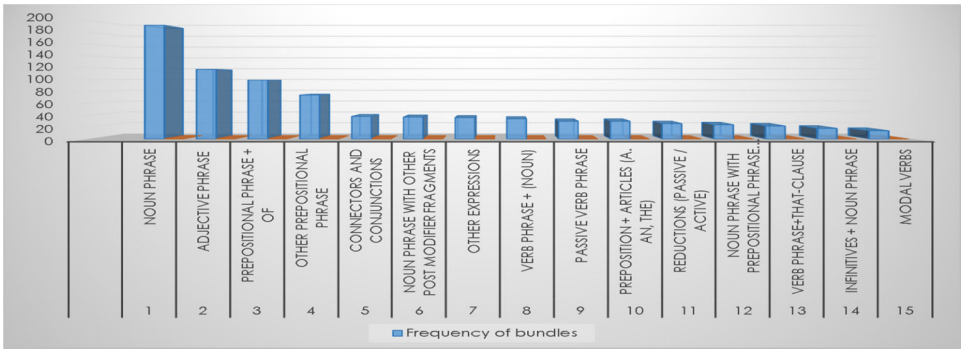

Figure 2

Lexical bundles identified in the introduction section of Civil Engineering articles

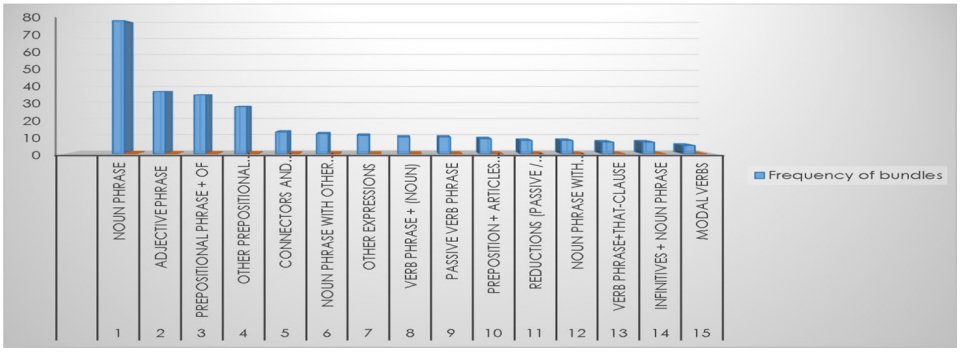

Figure 3

Lexical bundles found in the discussion section of Civil Engineering articles.

As can be seen, in both sections of the articles, the first three most frequent lexical bundles were noun phrases, adjective phrases, and

prepositional phrases with approximately the same frequency, the closest of which belonged to *prepositional phrase + of* with 12.53% and 12.54% of the total frequency in the two sections studied. Similarly, the lowest frequency in both the introduction and discussion sections of the articles belonged to the same categories of infinitives + noun phrase, modal verbs, and reductions (passive/active) with 2.78%, 2.27%, 1.77% of the bundles appearing in the introduction section, and 2.50%, 2.50%, 1.79% of the bundles appearing in the discussion section.

Discussion

The present study examined the types and frequencies of the lexical bundles that appeared in the introduction and discussion sections of Civil Engineering research papers. The analyses were performed focusing on frequency rates and forms of lexical bundles in the articles, and the findings were compared with each other in two sections of academic writing.

Based on the findings provided here, there appears to be a considerable convergence between the lexical bundles utilized in scholars' writings in the introduction and discussion sections of Civil Engineering articles. The results of the present study corroborated those of other similar studies (Alipour, et al., 2013; Farnia & Barati, 2017; Hyland, 2008b) in that they examined corpora of a wide range of lexical bundles and found that the English research papers written by students of medicine and other fields likewise contained a large and varied number of two-to four-word lexical prefabricated patterns.

The results of this study revealed that 51.51% and 84.2% of lexical bundles are phrasal in nature. The clausal bundles form about 31.79% of lexical bundles in the introduction sections, and 2.50% formed the clausal bundles in the conclusion section of the articles. Through the structural

examination of bundles, it was found that introductory and concluding parts accommodated the largest number of noun phrases or phrasal bundles. This finding is actually in line with those of Hyland (2008a). His study found that phrasal lexical groups were more frequent than clausal packs, lending backing to past investigations' thoughts or discoveries. For instance, Biber et al. (1999) found that the greater part of the lexical packs in scholarly compositions comprises chiefly phrasals, not clausals.

In the present study, prepositional phrases ranked third both in the introduction and conclusion sections of the Civil Engineering articles. This finding shows that phrasals are frequently used in academic writings. Similarly, in the corpora of applied linguistics, chemistry, and education, 75%, 55%, and 84% of bundles were phrasal in nature, respectively (Jalali, 2014; Parvizi, 2011; Valipoor, 2010). It was also found that the combination *prepositional phrase + of* was among the bundles used frequently in the category of phrasal bundles.

Another point worthy of mention is the idea that the present study results should be interpreted in light of the transfer from the native languages of the scholars, potentially affecting their use of the bundles in their research articles. The potential influence of cross-linguistic differences on the use of lexical bundles has been noted by many researchers (Gardner & Davies, 2007; Paquot, 2007). For example, Paquot (2007) found that learners of French used translational equivalents of frequent bundles in L2 writing; therefore, the current paper also highlights the necessity for research on the potential influence of lexical bundles cross-linguistically.

Because the linguistic makeup of specific registers highly contributes to a thorough analysis of lexical bundles, it can be considered an important pointer for defining the success of language users within these discourse

communities. Then, writers in English should integrate proper use of lexical bundles to create operational and fruitful discourse.

Conclusion

The scarcity of clausal bundles compared to the phrasal bundles used by scholars, as found in the current study, seems to suggest a failure in their adequate mastery of academic writing (Cortes, 2008; Durrant & Schmitt, 2010). High predictability scores of lexical frames are pedagogically valuable. English language learners can be exposed to phraseological variations through these lexical frames allied with specific lexical bundles.

Material developers and course designers can design materials consisting of multiword combinations to augment learners' knowledge of the uses and functions of lexical bundles. Effective implementation of bundles by the writers can be accomplished through explicit teaching of the bundles and emphasizing their appropriate use in a diverse range of texts.

Since the lexical bundles are register-specific, they are indicators of success in discourse communities. It would be of paramount significance for EAP learners to pay close heed to these features to be able to produce a convergent discourse. Like other texts practiced in universities, research papers contain lexical bundles prevalent in the university discourse. One area that learners need to know in EAP courses is vocabulary and word units that appear in research articles. Consequently, activities that raise awareness of lexical bundles and show their structures and functions should be used in classrooms.

In this research, the list of identified lexical bundles (e.g., *the purpose of this, purpose of this research, of this research is to*) overlap greatly in form and function with those identified in previous studies, largely because of the

methodology applied in extracting all n-grams of predefined length from the corpora. It is not always easy to categorize formulaic language into neat sets of two-, three-, or four-word chunks. As a matter of fact, linguists have endeavored to draw boundaries across the varying degrees of formulaicity in language. The beauty of lexical bundles lies in the fact that they are recognized based on frequency alone. Yet, a level of subjectivity is observed in deciding which lexical bundles are suitable for practicing in the classroom. This warrants the design of more effective ways of providing the most important and highly productive register-specific lexical bundles.

References

Alipour, M., Jalilifar, A., & Zarea, M. (2013). A corpus study of lexical bundles across different disciplines. *The Iranian EFL Journal, 9*(6), 11-35.

Bennet, A. (1991). *Discourses of power, the dialectics of understanding the power of literacy*. New York: Bergin and Garvey.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes, 26*, 263-286.

Biber, D., & Conrad, S. (2001). *Variation in English: Multi-dimensional studies*. London: Pearson.

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

Biber, D., Conrad, S., & Reppen, R. (1999). *Corpus linguistics: Investigating language structure and use*. Cambridge: University Press.

Bychkovska. T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes 30*, 38-52.

doi: 10.1016/j.jeap.2017.10.008

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

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43-57.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and move in research article introductions. *Journal of English for Academic Purposes 12*, 33-43.

Covington, M., & McFall, J. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics 17*(2), 94-100.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26*(2), 163-188.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 41*(3), 375-396.

Farnia, M., & Barati, S. (2017). Writing introduction sections of research articles in applied linguistics: Cross-linguistic study of native and non-native writers. *Indonesian Journal of Applied Linguistics, 7*(2), 486-494. doi: 10.17509/ijal.v7i2.8357

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching.* Ann Arbor, MI: University of Michigan Press.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*(2), 339-359.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

Hamzelou Moghadam, S., Zainal, Z., & Ghaderpour, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia-Sociol and Sciences, 66*, 555-563.

Herbel-Eisenmann, B., & Wagner, D. (2010). Appraising lexical bundles in mathematics classroom discourse: Obligation and choice. *Educ Stud Math, 75*, 43-63.

Hunston, S. (2000). *Evaluation and the planes of discourse: Status and value in persuasive texts.* Oxford: Oxford University Press.

Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, *28*(2), 266-285.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

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

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*,150-169.

Jalali, H. (2014). Examining novices' selection of lexical bundles: The case of EFL Postgraduate students in applied linguistics. *Journal of Applied Linguistics and Language Research, 1*(2), 1-11.

Jalali, Z. S., & Moini, M. R. (2014). Structure of lexical bundles in the introduction section of medical research articles. *Procedia: Social and Behavioral Sciences, 98*, 719-726. doi: 10.1016/j.sbspro.2014.03.473.

Jalali, H., & Zarei, G. R. (2016). Academic writing revisited: A phraseological applied linguistics high-stake genres from the perspective of lexical bundles. *Journal of Teaching Language Skills, 34*(4), 87-114.

José, M. & Marco, L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes, 19*, 63-86.

Katnic-Bakarsic, M. (2004). Discourse stylistic as contextualized stylistics. *Journal for Philogical Research, 15*(2), 37-48.

Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform, corpus analysis toolkit. IWLeL 2004: An *Interactive Workshop on Language e-Learning*, 7-13. Retrieved from http://www.laurenceanthony.net/research/iwlel_2004_anthony_AntConc.pdf

A critical look at software tools in corpus linguistics. *Linguistic Research*, *30*(2), 141-161.

Menon, S., & Mukundan, J. (2012). Collocations of high frequency noun keywords in prescribed science textbooks. *International Education Studies, 5*(6), 149-160.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* Amsterdam: Benjamins.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes, 21,* 60-71.

Paquot, M. (2007). *Exemplification in learner writing: A cross-linguistic perspective.* Amsterdam: Benjamins.

Parvizi, N. (2011). *Identification of discipline-specific lexical bundles in education (Unpublished thesis)*. University of Kashan.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes, 14*, 84-94.

Rafiee, M., & Keihaniyan, M. (2013). A comparative analysis of lexical bundles in journalistic writing in English and Persian: A contrastive linguistic perspective. *International Journal of Foreign Language Teaching and Research, 1*(2), 37-44.

Reller, T. *(2014). RELX annual reports and financial statements*. London: RELX Group.

Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes*, *50*, 1-16.

Sadeghi, K., & Khajepasha, A. S. (2015). Thesis writing challenges for non-native MA students. *Research in Post-Compulsory Education, 20*, 357-373.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins Publishing Company.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual.* New York, NY: Palgrave Macmillan.

Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, *69*, 79–91. doi: 10.1016/j.system.2017.08.002Sinclair, J. (2008). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Siyanova, C. A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review, 64*(3), 429-458.

Siyanova C. A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase "Time and Again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of experimental psychology. Learning, memory, and cognition, 37*,776-84.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17-35. doi: 10.1016/j.jslw.2016.02.00

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

Swales, J. (2004). Then and now: A reconsideration of the first corpus of scientific English. *Iberica*, *8*, 5-21.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569-613. doi: 10.1111/j.1467-9922.2010.00622.x

Valipoor, L. (2010). *A corpus-based study of words and bundles in chemistry research articles (Unpublished thesis*). University of Kashan.

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, *32*, 201-232.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes, 54,* 1-14.

Yin, X., & Li, S. (2021). Lexical bundles as an intradisciplinary and interdisciplinary mark: A corpus-based study of research articles from business, biology, and applied linguistics. *Applied Corpus Linguistics*, *1*(1), 1-11. doi: 10.1016/j.acorp.2021.100006

Appendix

Structural Taxonomy of Lexical Bundles (Biber et al., 2004)

1. Noun phrase with of- phrase fragment the beginning of the, the shape of the
2. Noun phrase with other post-modifier fragments the way in which, the extent to which
3. Prepositional phrase with embedded of-phrase as a result of, in the case of fragment
4. Other prepositional phrases (fragment) at the same time, on the other hand
5. Anticipatory it + verb / adjective phrase it is possible to, it should be noted that
6. Passive verb + prepositional phrase fragment is shown in figure, is based on the
7. Copula be + noun / adjective phrase is one of the, is part of the, is due to the
8. (Verb phrase+) that- clause fragment has been shown that, that there is no
9. (Verb/ adjective +) to-clause fragment are likely to be, has been shown to, to be able to
10. Adverbial clause fragment as we have seen, if there is a
11. Pronoun/ noun phrase+ be (+…) this is not the, there was no significant
12. Other expressions as well as the, than that of the