Comparing Lexical Bundles in Medical Research Article Abstracts of Iranian and Foreign Journals

Elshan Varghaei [a]
e.varghaei@iauardabil.ac.ir
Islamic Azad University,
Ardabil Branch, Ardabil, Iran

Golsa Khodadadi [b]
Golsakhodadadi25@gmail.com
University of Tabriz, Iran

ABSTRACT

The art of persuasive speaking or writing in English academic settings depends on using multi-word expressions, also known as clusters, collocations, or lexical bundles, to a large extent. They are discipline-specific prefabricated word combinations that statistically tend to appear together. However, many novice foreign writers of English have difficulty using appropriate discipline-bound lexical bundles, which has made numerous applied linguists in English for Academic Purposes (EAP) study the genre of research articles from this perspective. Thus, the present study sought to compare the similarities and differences of four-word lexical bundles in a 409210-word corpus of Medical Research Article (MRA) abstracts from Iranian and foreign journals. To accomplish this, in a frequency-based approach, bundles were extracted utilizing the AntConc3.5.7 concordance program, and their structures and functions were analyzed by Biber et al.'s (1999) structural and Hyland's (2008a) functional taxonomies. The results showed similarities in the distribution of the bundles' main structural and functional patterns. However, besides the differences in the distribution of sub-structures and sub-functions, it was revealed that highly frequently shared lexical types in MRA abstracts of Iranian journals were less frequent in MRA abstracts of foreign journals and vice versa. This study helps novice medical researchers write unified abstracts which have a crucial role in getting research articles accepted or rejected. Furthermore, producing well-organized abstracts in internal Iranian journals can significantly enhance the rank of Iranian medical journals worldwide.

Keywords: Lexical bundles; Frequency; Structure; Function; Medical research article abstracts

INTRODUCTION

Recently, due to growth in scientific publication in English around the world and existing differences in stylistic conventions of every academic community, studying Research Articles (RA) from different perspectives has received considerable attention (Al-Khasawneh, 2017). In this case, the overall organizational patterns, as well as the linguistic features of RA abstracts, as the essential sections of RAs, have been explicitly examined (e.g., Anderson & Maclean, 1997; Pho, 2008; Kim, 2014; Darabad, 2016; Al-Khasawneh, 2017). A genre analysis of a research article abstract is worthwhile because: it is the only part being published in conference

---

[a] *Main author*
[b] *Corresponding author*

proceedings, being seen by potential referees while reviewing or editing, and it is the first and the only part that readers encounter when they search through electronic databases (Andrade, 2011).

However, contrary to the extensive studies of lexical bundles in academic registers of written and academic discourse (e.g., Charles, 2003; Nesi & Basturkmen, 2006; Hyland 2008a, 2008b; Adel & Erman, 2012; Heng, Kashiha, & Tan, 2014; Yang, 2017; Muslu, 2018; Lee, 2020), it seems the subject has been overlooked in RA abstracts of medicine to some extent. For example, while medical education and research are largely conducted in English (Maher, 1986; Swales, 1990; Fryer, 2012), it was noticed that there had been no examination into comparing the features of lexical types in MRA abstracts of foreign and domestic medical publications. Thus, to bridge the perceived gap, this study investigates the structure and function of the most frequently utilized lexical types in abstract sections of MRAs published in Iranian and foreign Journals. Lexical bundles are "statistically the most frequent recurring sequences of words in any collection of texts" (Hyland, 2012, p. 150), which are familiar to writers and readers who frequently use a unique genre (Hyland, 2008b). Since abstracts are essential in getting RAs accepted or rejected (Kanoksilapatham, 2009), novice researchers can produce well-structured abstracts using lexical bundles. They are extended collocations known to be highly influential in creating coherency and organization of different texts (Cortes, 2004, 2008; Biber, Conrad, & Cortes, 2003, 2004; Hyland, 2008b; Jeblonaki, 2009; Jalali et al., 2009).

CORPUS-BASED MULTI-WORD EXPRESSIONS

Word combinations have assumed importance as building blocks in academic discourse since the development of corpus techniques (Chen & Baker, 2010). Since then, under different terminologies including "recurrent word combinations" (Altenberg, 1998; De Cock, 1998); "phrasicon" (De Cock, Granger, Leech, & McEnery, 1998); "lexical bundles" (Biber, Johansson, Leech, Conrad & Finegan, 1999; Stubbs, 2007); and "n-grams," (Banerjee & Pederson, 2003), to name a few, multi-word units have been investigated from different aspects. According to John Sinclair (2005), a corpus is "a collection of pieces of language texts, in electronic form, selected according to external criteria to present, as far as possible, a language or language variety as a source of data for linguistic research" (p.16). In a corpus-based analysis of multi-word expressions, concordance is used as the primary technique to extract naturally occurring word combinations from ever-larger text collections (Salazar, 2011). To do so, a frequency cut-off, the number of times a lexical bundle should occur in a corpus, must be specified arbitrarily at the beginning (Hyland, 2006). Then, once the concordance process is finished using quantitative methods, the identified word combinations can be interpreted qualitatively in the context of their use (Salazar, 2011).

LEXICAL BUNDLES

As mentioned, multi-word units have been named differently. In 1999, Biber et al. introduced the term lexical bundles and distinguished it from other word combinations, like formulaic sequences and idioms. Afterward, the term has been used by various scholars (e.g., Cortes, 2004; Hyland, 2008a, 2008b; Chen & Baker, 2010; Adel & Erman, 2012) in a corpus-driven study of word combinations. In contrast with other formulaic expressions, lexical bundles are structurally incomplete extended collocations made up of three to six words that frequently tend to co-occur together (Biber et al., 1999). They are not idiomatic, and their meanings are easily understood from

the meaning of the individual words that make up the bundle (Cortes, 2004). On the other hand, lexical bundles are similar to other word combinations in that they are fixed expressions extracted by computer programs and electronic corpora. In the case of approaches to identifying lexical bundles, again, frequency has been used as the main criteria. However, "the actual frequency cut-off used to identify lexical bundles is somewhat arbitrary" (Biber, Conrad, & Cortes, 2004, p. 376). For larger corpora, the frequency cut-off is set at 10 to 40 times per million words (Biber et al. 1999; Biber, Conrad, and Cortes 2004), and for small corpora, the normalized frequency cut-off ranges from 2 to 10 occurrences per million words (Fattani, 2018). Furthermore, dispersion, the range of texts in which a lexical bundle is distributed, has been used as the second criterion in examining lexical bundles to avoid "idiosyncratic uses of individual writers or speakers" (Biber & Barbieri, 2007, p. 278).

PREVIOUS STUDIES OF LEXICAL BUNDLES

It is argued that recurrent word combinations, also known as lexical bundles, comprise a large portion of discourse (Schmitt & Carter, 2004). For instance, Erman & Warren (2000) indicated that 58.6% of spoken discourse and 52.3% of the written discourse in their corpus of London-Lund Corpus of Spoken English (LLC) and Lancaster-Oslo-Bergen (LOB) are part of multi-word combinations. In addition, lexical bundles are said to be essential building blocks in academic discourse (Hyland, 2008b) because they are registered specific, and their structures and functions vary among different disciplines (Allen, 2009). As such, the results have provoked researchers in applied linguistics to look at lexical bundles in several ways, using Biber et al. (1999) structural and functional taxonomies or Hyland's (2008a) functional patterns as their primary methods. For example, to identify the actual grammatical patterns of four-word lexical bundles across registers, Biber et al. (1999) concluded that contrary to spoken registers, which use clausal bundles, phrasal bundles are predominant in academic prose. In his study of lexical bundles in academic texts of various disciplines, Hyland (2008a, 2008b) also reported that lexical bundles' frequency, structure, and function vary in different disciplines. In addition to this, many other studies have shown that, in comparison to native English authors, non-native speakers of English either utilize the smallest range of lexical bundles or have difficulty using different types of bundles properly (e.g., Chen & Baker, 2010; Adel & Erman, 2012; Gungor & Uysal, 2016; Bychkovska & Lee, 2017; Shin, Cortes, & Yoo, 2018). Besides this, attempts have been made to create pedagogically useful lists of essential lexical bundles in English for Academic Purposes (EAP) or English for Occupational Purposes (EOP) texts by using corpora (e.g., Coxhead, 2000; Verdaguer et al., 2009; Simpson-Vlach & Ellis, 2010; Kim & Lee, 2018; 2019; Lee, 2020). For instance, Kim and Lees' (2018, 2019) Linguistic Academic Vocabulary List (AVL) is an academic word list for students in linguistics constructed upon Corpus of Linguistic Textbooks (CLT) with 1.14 million running words. Recently, in a corpus-driven study, Subramaniam & Kaur (2021) have also compared the use of passive verb bundles in the British Academic Written English (BAWE) sub-corpus and the Malaysian Polytechnics Electronic Engineering Learner Corpus (MyPolyEELC). Their findings showed that both learner corpora contained a variety of passive verb structural categories. Moreover, according to their analysis, contrary to L1 English learners who extensively used the "engagement" and "stance" categories, limited passive verb bundles with participant-oriented functions were disclosed in L2 Malaysian English learners. The results of such approaches can be pedagogically significant in those EAP courses that try to make novice researchers familiar with

the most frequently utilized discipline-bound lexical items in the target discourse community (Cortes, Jones, & Stoller, 2002).

Throughout history, the subject of lexical bundles has been studied in different genres of research articles (Hyland 2008a, 2008b; Salazar, 2011; Kwary, Ratri & Artha, 2017), textbooks (Hsu, 2015; Lee, 2020), and spoken discourse (Nesi & Basturkmen, 2006; Liu & Chen, 2020) to a large extent. However, some distinct features of lexical bundles in the genre of MRA have been taken for granted. Although some studies have concentrated on two-word discontinuous collocations (Marco, 2000), some have studied the structures and functions of lexical bundles in the introduction and discussion sections of MRA (Jalali & Moini, 2014; 2018), and the others have examined their functions in doctor talk (Panthong & Poonpon, 2020), no studies have been conducted over MRA abstracts. As stated, the abstract is one of the most important genres in academic writing. It is the summary of a research article, and it aims to help other researchers and readers decide to select appropriate readings (Darabad, 2016). Because of the word limit for the abstracts set by the editor (Wissberg & Buker, 1990), writing a coherent abstract is one of the main problems for novice writers of English. Hence, making novice researchers in the field of medicine familiar with the critical features of four-word lexical bundles in the abstract part is the central purpose of this study. Many previous types of research have indicated that lexical bundles are used differently in various sections of research articles (e.g., Ahmadi, Ghonsooly, and Fatemi, 2013; Shahmoradi, Jalali, & Ghadiri, 2021). In this regard, although English is used extensively in the field of medicine (Swales, 1990; Freyer, 2012), studying lexical bundles in MRA abstracts has been neglected. As a crucial component, the abstract provides a precise and understandable description of the article, enabling readers to assess whether a piece of writing is valuable and pertinent (Shahmoradi, Jalali, & Ghadiri, 2021). Therefore, Studying lexical bundles in MRA abstracts of foreign and Iranian journals can help novice researchers become familiar with the most frequently used lexical bundles in MRA abstracts and help them comprehend the structures and functions of lexical bundles in this section. As a result, medical students can retain textual coherence and organization of their abstract.

RESEARCH QUESTIONS

1. Are there any significant differences in the distribution of the most frequently used structural patterns of four-word lexical bundles between MRA abstracts of foreign and Iranian journals?
2. Are there any significant differences in the distribution of the most frequently used functional patterns of four-word lexical bundles in MRA abstracts of foreign and Iranian journals?
3. What are the most commonly shared 4-word lexical bundles between MRA abstracts of foreign and Iranian medical journals?

METHODOLOGY AND DATA GATHERING

Two types of corpora, MRA abstracts published in foreign and Iranian medical journals, were developed to answer the research questions. To do so, the initial step was to find reliable indexed databases in the field of medicine. However, Medical journal indexing is a contentious matter. That is, while Index Medicus has been the most comprehensive and thorough index of medical scientific journal publications since 1879, several other popular publishing databases have been developed during this time, including MedLine, PubMed, EMBASE, SCOPUS, EBSCO

Publishing's Electronic Databases, and SCIRUS, to name a few (Balhara, 2012). For this investigation, Iranian medical journals were chosen from PubMed, PubMed Central, SCOPUS, and Directory of Open Access Journal (DOAJ) databases, and foreign journals from Web of Science, Medline, PubMed, and PubMed Central. Although it is believed that "the corpus must be large enough to adequately represent the occurrence of the features being studied" (Biber, 2006, p. 251), this study utilized small corpora, which is helpful for pedagogical purposes (Cortes, 2004; Ghadessy et al., 2001; O'keeffe & McCarthy, 2007). As stated by Salazar (2011), topic and text types were also used as the main criteria in collecting the data. In terms of authorship, contrary to the other researchers (e.g., Wood, 2001, Salazar, 2011), being native or non-native was not considered in the data collection process. This is because scientific writing is published after a rigorous peer-review process, and the published article is a representative sample of MRA regardless of the authors' first language (Mbodj-Diop, 2016).

Considering the criteria mentioned earlier, the target size of the corpus for this study was 1228 MRA with 409210 running words. Since there are generally around 2000 high-frequency words in English (Ghadessy et al., 2001), the overall number of the corpora in this study is reliable for extracting the most frequently used lexical types in MRA abstracts of foreign and Iranian journals. For developing the corpus of foreign journals, 600 abstracts with 209173 running words were selected randomly from foreign journals of New England Journal of Medicine (NEJM), Journal of American Medical Association (JAMA), Science, and Journal of Clinical Investigations (JCS). The selected journals are English's most prestigious medical journals for students seeking professional advancement (Mbodj-Diop, 2016). In the case of the Iranian corpus, 628 article abstracts with a total number of 200037 words were gathered from the Iranian journal of Advances in Medical Education and Professionalism (JAMP), Journal of Caring Science (JCS), Journal of Cardiovascular and Thoracic Research (JCVTR), and Journal of Nephropathology (JNP) being published at top universities of Iran according to US News Education, including Tehran, Shahid Beheshti, Isfahan, and the Tabriz University of Medical Science in English. The medical articles of both groups were original studies picked from subject matters of allergy, asthma, cancer, types 1 and 2 diabetes, cardiology, HIV, infection, and hypertension.

To answer the research questions, both quantitative and qualitative methods were used. The most frequently used lexical bundles were extracted quantitatively by AntConc3.5.7 concordance program, a freeware concordance program capable of automatically counting the frequency of bundles based on the applied criteria. For this purpose, the bundles were extracted according to the three criteria of cut-off frequency, dispersion, and mutual information (MI). As previously mentioned, the actual frequency cut-off for lexical bundle identification is fairly arbitrary (Biber, Conrad, & Cortes, 2004). While some research based on large corpora (Hyland, 2008b; Biber et al., 2004) has used the frequency cut-off of 20–40, those studies using small size corpora have selected the frequency cut-off of 1-10 (e.g., Altenberg, 1998; De Cock, 1998; Shahmoradi, Jalali, & Ghadiri, 2021). Therefore, because the present study has been built upon mini-corpora, the cut-off frequency was set at 3 times per million words and the dispersion at 2% of the articles. The identified bundles' mutual information score was also calculated to discern free word combinations with the exact lexical bundles. Only 4-word lexical types were the case of study because many four-word bundles hold three-word bundles in their structures (Cortes, 2004; Nesi & Basturkmen, 2006), and in comparison to five- and six-word strings, four-word bundles are more common (Hyland, 2008a). Then, using the qualitative method, the structures and functions of the identified bundles were examined utilizing AntConc3.5.7 in their context. Biber

et al.'s (1999) structural classifications and Hyland's (2008a) functional taxonomies were applied to investigate the bundles structurally and functionally.

RESULTS

BUNDLE IDENTIFICATION

To answer the research questions, the corpora of both groups were analyzed by AntConc3.5.7 based on the stated criteria. Because of the frequency cut-off of 3, many four-word lexical types were extracted from both groups. However, the final lists of the bundles were determined based on their functions. That is, bundles that could not be classified functionally were excluded from the list despite their high frequency. 317 and 253 four-word lexical bundles were identified in the MRA abstracts of foreign and Iranian journals, respectively.

COMPARING THE STRUCTURES OF LEXICAL BUNDLES

Comparing structural differences of the bundles in MRA abstracts of foreign and Iranian journals was among the primary purposes of this study. Thus, the lexical bundles of both groups were classified structurally according to Biber et al.'s (1999) structural taxonomy, and the results were compared. Table 1 compares the distribution of lexical bundles structurally in MRA abstracts of foreign and Iranian journals.

TABLE 1. Comparing the structures of lexical bundles (%)

Structure	% Foreign Journal		% Iranian Journal	
Noun Phrase with *of*-phrase fragment	20.36	28.87	13.58	25.92
Noun phrase with other post-modifier fragments	8.51		12.34	
Prepositional phrase with embedded *of*-phrase fragment	13.06	26.73	8.64	20.57
Other prepositional phrase fragments	13.67		11.93	
Be + noun/ adjective phrase	3.03		2.057	
Passive verb + prepositional phrase fragment	13.98		17.28	
Anticipatory *it* + verb/ adjective phrase	0.91		1.23	
(verb phrase) + *that*-clause fragment	0.91		3.70	
(Verb/ adjective) + *to*-clause fragment	1.82		2.46	
Adverbial clause fragment	1.82		0.82	
Pronoun/ noun phrase + be (+)	2.12		8.23	
Other expressions	19.75		18.93	

According to table 1, bundles with noun structures are the most frequently used lexical types in MRA abstracts of both groups. As can be seen in table 1, 28.87% of the extracted bundles from MRA abstracts of foreign journals, as well as 25.92% of the overall bundles in MRA abstracts of Iranian journals, are noun structures (see examples 1 and 2). Such evidence confirms the findings of previous researchers, such as Swales (2008), who concluded that academic writing is "noun-centric" (Swales, 2008, p. v; as Cited in Salazar, 2011). In addition, as the data in table 1 show, bundles with prepositional phrase fragments are another most frequently used four-word

lexical type in MRA abstracts of both groups (see examples 3 and 4). This includes 26.73% of the identified bundles in foreign medical journals and 20.57% in Iranian journals. On the other hand, there are differences in the distribution of other structural sub-categories. For example, as can be seen in table 1, while noun phrases with embedded *of*-phrase fragments with 20.36% are the most frequently used structures in MRA abstracts of foreign journals, another expression (18.93%) is higher than other structural patterns in MRA abstracts of Iranian journals. For comparing other structural patterns, refer to table 1.

(1) (MRA abstract of the foreign journal): in this 6-month trial involving patients with type 1 diabetes, **the use of a** closed-loop system was associated with a greater percentage of time spent in a greater glycemic range than **the use of a** sensor-augmented insulin pump.
(2) (MRA abstract of Iranian journal): there was **a significant increase in** the median scoring from 1 to 2 (P= 0.027) in overall self-reported preparedness for meeting the level 1 milestone included in the elective, as well as a significant increase in sub-categories across competencies 1-4 outlined by the ACGME.
(3) (MRA of the foreign journal): Bariatric surgery results in weight loss and health improvements in adults and adolescents. However, whether outcomes differ according to the age of the patient **at the time of** surgery is unclear.
Example 4 (MRA of Iranian journal): **in the control group,** the students underwent the current format of the Fixed Prosthesis curriculum.

COMPARING THE FUNCTIONS OF LEXICAL BUNDLES

After functional classification of bundles in both groups, the results were compared to find similarities and differences. However, because of the multi-functional characteristics of lexical bundles, like previous research (e.g., Salazar, 2011), this study decided on the functions of bundles according to their most common use and their co-text—the words that are coming before and after a lexical bundle. Therefore, those bundles which could play different functions in the same context were included in more than one functional category. Table 2 below compares the overall percentage of the functions of bundles between MRA abstracts of foreign and Iranian journals according to Hyland's (2008a) functional taxonomies.

TABLE 2. Comparing the functions of lexical bundles (%)

Category & Sub-category of Functions		% Foreign Journal	% Iranian Journal
Research Oriented Bundles	location (in time/ place)	9.78	17.02
	Procedure	14.06	28.36
	Quantification	27.52	21.27
	Description	10.39	2.48
	Topic	5.50	0.70
		67.25	69.83
Text Oriented Bundles	Transition Signals (Additive /Contrastive Links)	0.30	1.06
	Resultative Signals (Inferential/causative relations)	7.64	10.63
	Structuring signals	3.97	10.28
	Framing Signal	13.14	2.83
		25.05	24.8

Participant Oriented Bundles	Stance	5.19	2.83
	Engagement	2.44	2.48
		7.63	5.31

According to the obtained results in table 2, among the three main functional categories of bundles, that is, research-oriented, text-oriented, and participant-oriented bundles, research-oriented bundles are the main functional patterns in MRA abstracts of foreign and Iranian journals. As the information in table 2 indicates, research-oriented bundles cover over 60% of the identified four-word bundles in MRA abstracts of both classes. The findings in this regard are similar to the findings of Hyland (2008a). To be more specific, Hyland (2008a) concluded that hard-science articles are "laboratory-focused" and concerned with things related to the "real world." Therefore, research-oriented bundles in MRA abstracts of foreign and Iranian journals are mainly utilized for describing research contexts and objects, indicating methodology and purposes, and measuring and showing the numbers or properties and changes (see examples 5, 6, and 7). Nevertheless, by close examination of the sub-functions presented in table 2, it was revealed that there are differences in the distribution of sub-categories of research-oriented bundles between the two groups. For example, as table 2 illustrates, the bundles with functions of quantification (27.52%), procedure (14.06), and description (10.39%) are respectively the first, second, and third frequently used lexical bundles in MRA of foreign journals. On the other hand, bundles with the functions of procedure (28.36%), quantification (21.27%), and location (17.02%) are in the first, second, and third place in MRA of Iranian journals. For the distribution rate of other lexical types, refer to table 2.

(5) (MRA of the foreign journal, location): **at the end of** the primary trial, all the participants to avoid peanuts for 12 months. The primary outcome was the percentage of participants with peanut allergy **at the end of** the 12-month period when the participants were 72 months of age.
(6) (MRA of the foreign journal, quantification): protein-truncating variants in 5 genes (ATM, BRCA2, CHEK2, AND PALB2) were associated with a risk of breast cancer overall with **a p-value of** less than 0.0001.
(7) (MRA of Iranian journal, procedure): all observational studies reporting the prevalence of MetS among people with and without asthma **were included in the** study.

SHARED LEXICAL BUNDLES IN MRA ABSTRACTS OF BOTH CORPORA

Table 3 shows the frequency of shared four-word lexical types in MRA abstracts of both classes.

TABLE 3. The frequency of shared lexical bundles in MRA abstracts of Foreign and Iranian Journals

Lexical Bundle	Frequency in Foreign Journal	Frequency in Iranian Journal
Were randomly assigned to	81	17
In the control group	53	43
Was associated with a	49	4
Were included in the	43	13
With the use of	43	3

On the basis of	34	5
In the intervention group	31	46
Patients were randomly assigned	31	3
Did not differ significantly	26	3
In the two groups	24	8
There was no significant	24	33
At the time of	23	7
For the treatment of	18	5
Was not associated with	17	4
Between the two groups	15	34
No significant difference in	13	11
These findings suggest that	13	4
Were more likely to	13	8
No significant difference between	11	23
In the context of	11	3
The use of a	10	3
The end of the	9	11
With an increase in	9	4
An increase in the	9	7
In the absence of	8	3
In the treatment of	8	4
The results of the	8	23
At the end of	8	18
A p-value of	6	3
The results of this	6	28
In this study we	6	17
The aim of this	5	84
Aim of this study	5	77
More likely to be	5	4
It is important to	5	3
The follow-up period	5	3
Did not differ between	5	3
As well as the	5	10
There was a significant	4	47
A significant reduction in	4	3
Differences between the groups	4	3
The effect of the	4	4
Results of this study	4	31
The majority of the	4	11
Was found to be	4	9
Was not significantly different	3	6
Can be used to	3	10
Has been shown to	3	4
Studies are needed to	3	4

As shown in Table 3, 8.55% (n=49) of the identified bundles are commonly shared by MRA of foreign and Iranian journals. However, as can be seen in Table 3, the frequencies of the shared bundles differ in the two corpora. For example, while bundles like *were randomly assigned to* (n=81), *in the control group* (n=53), *was associated with a* (n=49), and *with the use of* (n=43) are the most frequently used lexical types in MRA abstracts of foreign journals, in MRA of Iranian

journals, *the aim of this* (n=84), *aim of this study* (n=77), *there was a significant* (n=47), and *in the intervention group* (n=46) are the most frequently utilized types. For structural and functional classifications of the above-mentioned shared bundles, refer to tables 4 and 5.

TABLE 4. Structural classification of the shared bundles

Noun Phrase with *of*-phrase fragment	The use of a, the end of the, a p-value of, the majority of the, the results of the, results of this study, aim of this study, the effect of the, the results of this, the aim of this,
Noun phrase with other post-modifier fragments	No significant difference in, no significant difference between, differences between the groups, an increase in the, a significant reduction in
Prepositional phrase with embedded *of*-phrase fragment	At the time of, for the treatment of, in the context of, in the absence of, in the treatment of, at the end of, on the basis of, with the use of
Other prepositional phrase fragments	In the control group, in the intervention group, in the two groups, with an increase in, between the two groups, in this study we
Be + noun/ adjective phrase	Were more likely to
Passive verb + prepositional phrase fragment	Were randomly assigned to, was associated with a, were included in the, was not associated with, was found to be, has been shown to
Anticipatory *it* + verb/ adjective phrase	It is important to
(verb/ noun phrase) + *that*-clause fragment	These findings suggest that
(Verb/ adjective) + *to*-clause fragment	More likely to be, can be used to, studies are needed to
Adverbial clause fragment	-----
Pronoun/ noun phrase + be (+)	There was no significant; there was a significant
Other expressions	Was not significantly different, as well as the, patients were randomly assigned, did not differ significantly, the follow-up period, did not differ between

Again, as the information in table 4 indicates, noun structures (n= 15) and bundles beginning with prepositional phrases (n=14) are leading structural patterns. Especially coinciding with the findings of the previous researchers (e.g., Biber et al., 1999; Hyland, 2008a; Salazar, 2011), the findings in Table 4 show that bundles with the structural patterns of Noun Phrase with *of*-phrase fragments with 20.40% are the most frequently shared types. For more information about other structural patterns of shared bundles, refer to table 4.

TABLE 5. Functional classification of the shared bundles

Research Oriented Bundles	location (in time/ place): the end of the, at the time of, at the end of, in the control group, in the intervention group, in the two groups, between the two groups, the follow-up period
	Procedure: the use of a, for the treatment of, in the treatment of, were randomly assigned to, were included in the, was found to be, patients were randomly assigned, with the use of
	Quantification: a p-value of, the majority of the, no significant difference in, no significant difference between, an increase in the, a significant reduction in, with an increase in, there was no significant, there was a significant, was not significantly different, did not differ significantly, differences between the groups, did not differ between
Text Oriented Bundles	Transition Signals (Additive /Contrastive Links): as well as the
	Resultative Signals (Inferential/causative relations): results of this study, the effect of the, the results of this, was associated with a, was not associated with, these findings suggest that, the results of the
	Structuring signals: aim of this study, the aim of this, in this study we
	Framing Signal: in the context of, in the absence of, on the basis of
Participant Oriented Bundles	Stance: we're more likely to, more likely to be, can be used to
	Engagement: has been shown to, it is important to, studies are needed to

As table 5 indicates, many identified shared bundles are included within research-oriented bundles (n=29). As it is clear from table 5, bundles in this category are used to function as location (n=7), procedure (n= 8), and quantification (n=13). In addition to this, as the results in table 5 demonstrate, text-oriented bundles, which are used to organize text and meanings (Hyland, 2008a), and participant-oriented bundles, which are used to focus on writers or readers are in the second (n=14) and third places (n=6) respectively. Look at table 5 for detailed information regarding the functions of the shard bundles.

DISCUSSION

This study attempted to compare the most frequently used structural and functional patterns of four-word lexical bundles in MRA abstracts of foreign and Iranian medical journals. Extracting the most commonly shared four-word lexical types in MRA abstracts of both groups was also among the primary purposes of this research. In agreement with the earlier research findings (Biber et al., 1999, Cortes, 2004; Hyland 2008a; Salazar, 2011; Mbodj-Diop, 2016; Shirazizadeh, Amirfazlian, 2021), comparing the structures of four-word lexical types indicated that noun

structures and prepositional phrases are the main structural patterns in MRA abstracts of foreign and Iranian journals. Together they accounted for over half of the extracted bundles in MRA abstracts of foreign and Iranian journals. In terms of the functional patterns, consistent with the results of the previous researchers (e.g., Hyland, 2008a; Rezoug & Vincent, 2018; Wachidah, Fitriati, & Widhiyanto, 2020), research-oriented bundles were predominant in MRA abstracts of both classes. However, as clearly illustrated, significant differences were seen in the distribution of sub-structures, sub-functions, and frequencies of the shared bundles in MRA abstracts of foreign and Iranian journals. While bundles with noun phrases with embedded *of*-phrase fragments were in the first place in MRA abstracts of foreign journals, other expression bundles were more frequent in MRA of Iranian journals. Furthermore, while bundles with the function of quantification were the primary sub-category of the research-oriented bundle in MRA abstracts of foreign journals, bundles with the function of procedure were in the first place in MRA abstracts of Iranian journals. Regarding the frequency of the shared bundles, as shown, many highly frequently used lexical types in MRA abstracts of the foreign journal were less frequent in MRA abstracts of Iranian journals and vice versa.

Although authors' identities were not considered as strict criteria in corpus development, a close examination of the available data revealed that Iranian authors wrote many gathered data from Iranian medical journals. However, no clear-cut information is at hand for writers of foreign journals. So, to determine whether such differentiations result from being a native or non-native writer of research articles, further studies can examine the main features of MRA abstracts of foreign and Iranian journals by considering the authorship factor as the main criteria. According to the findings of the previous researchers, in comparison to native speakers of English, non-native speakers of English use fewer lexical bundles (Howarth, 1998); or "underuse native-like collocations and use typical word combination" (Granger, 1998, p. 6). Therefore, studying lexical bundles from this perspective is worthwhile. In line with this, the findings of the previous researchers have also highlighted the different use of lexical bundles in the discourse of native and non-native writers (Pan, Reppen & Biber, 2016; Esfandiari and Barbary, 2017; Shin, 2019; Lu & Deng, 2019; Akbulut, 2020, Anwar, Ali Malik, Jamshid, 2020). Therefore, replicating this study by considering native and non-native elements as the main criterion can give fruitful results in the case of obtained differences.

Since every discourse community uses a specific set of lexical bundles associated with that discipline's discourse community, the findings of this study are promising for both syllabus designers and novice EFL Iranian students in the field of medicine. Using the findings of this study, Iranian novice researchers of medicine can publish their research papers in foreign indexed journals. In addition, the findings can help Iranian medical journals written in English to get indexed in reputable information databases such as Web of Science, Medline, or PubMed (Aminpour & Kabiri, 2009) by summarizing the content and results of the study in abstract parts in a more effective way. On the other hand, course developers in the field of medicine can also design the course materials based on the needs of the EFL students. They can predict the bundles' types, structures, and functions, which would be very helpful for Iranian EFL students in writing the abstract part of the research.

CONCLUSION

LIMITATIONS AND FUTURE RESEARCH

Considering the size of the corpora and the arbitrariness nature of cut-off frequency for identifying bundles, the results of this study might have been influenced. That is, because of the low-frequency cut-off of 3 and dispersion of 2% of the texts, some extracted four-word bundles might be mistakenly considered as fixed prefabricated bundles. So, more reliable results would be obtained by replicating the study based on large corpora and a frequency cut-off of 20-40 (Hyland, 2008b; Biber et al., 2004). Besides, as stated, whether the authors of the articles are native or non-native was not considered as the main criteria in the data collection process. Therefore, repeating the study by considering native-English and non-English writers as the main factors can give deep comprehension of the extracted differences in MRA abstracts of foreign and Iranian journals.

REFERENCES

Akbulut, F. D. (2020). A bibliometric analysis of lexical bundles usage in native and non-native academic writing. *Journal of Language and Linguistic Studies*, *16*(3), 1146-1166.

Anwar, N., Malik, M. A., & Jamshaid, S. (2020). Lexical Bundles In Academic Discourse: A Comparative Study Of Functional Use Of Lexical Bundles In Native And Non-Native Learner Corpora. *Epistemology*, *7*(3), 28-43.

Ahmadi, H. S., Ghonsooly, B., & Fatemi, A. H. (2013). An analysis of lexical bundles in research article abstracts by Iranian and native English-speaking authors of applied linguistics articles. *The Asian ESP Journal*, *9*(1), 5-25.

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English. *English for specific purposes*, *31*(2), 81-92.

Allen, D. (2009). Lexical bundles in learner writing: an analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education, 1*, 105-127.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie: Oxford University Press.

Aminpour, A., & Kabiri, B. (2009). Science production in Iran: the scenario of Iranian medical journals. *Journal of research in* medical *sciences, 14*(15), 313-322

Anderson, K., & Maclean, J. (1997). *A Genre Analysis Study of 80 Medical Abstracts*.

Andrade, C., (2011). How to write a good abstract for a scientific paper or conference presentation. *Indian Journal of Psychiatry*, *53*(2), p.172.

Al-Khasawneh, F. M. (2017). A genre analysis of research article abstracts written by native and non-native speakers of English. *Journal of Applied Linguistics and Language Research, 4*(1), 1–13.

Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers.* Amesterdam: Benjamin.

Banerjee, S. & Pedersen, T. (2003). Extended gloss overlap as a measure of semantic relatedness. *In Proc. Of IJCAI-03,* (PP. 805-810).

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Biber, D., Conard, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In: A. Wilson, P. Rayson, & T. McEnery, (Eds), *corpus linguistics by the lune: A festschrift for Geofferey Leech.* Frankfurt: Peter Lang.

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

Balhara, Y. P. (2012). Indexed journal: What does it mean. *Lung India*, *29*(2), 193.

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

Bychkovshka, 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.

Charles, M. (2003). 'This mystery…': a corpus-based study of the use of nouns to construct stance in theses from two constructing desciplines. *Journal of English for Academic Purposes, 2*(4), 313-326.

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

Cortes, V., Jones, J. K., & Stoller, F. (2002). Lexical bundles in ESP reading and writing. Paper presented at the TESOL

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

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

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Darabad, A. M. (2016). Move Analysis of Research Article Abstracts : A Cross-Disciplinary Study. *International Journal of Linguistics*, *8*(2), 125–140.

De Cock, S. (1998). A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English. *International Journal of Corpus Linguistics, 3*(1), 59–80.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer (*pp. 67-79). London: Longman.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Esfandiari, R., & Barbary, F. (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.

Fattani, R. (2018). *Profiling lexical bundles in an EAP pre-sessional course: A corpus-based study on textbooks and instructors' materials* (phd). the university of Sheffield.

Fryer, D. L. (2012). Analysis of the generic discourse features of the English-language medical research article: A systematic-functional approach. *Functions of Language, 19*(1), 5-37.

Ghadessy, M., Henry, A., & Roseberry, R. (2001). *Small corpus studies and ELT*. Philadelphia: John Benjamins.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: An interdisciplinary perspective* (pp. 145-160). Amsterdam: John Benjamins.

Gungor, F., & Uysal, H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, *9*(6), 176.

Heng, C. S., Kashiha, H., & Tan, H. (2014). Lexical Bundles: Facilitating University "Talk" in Group Discussions. *English Language Teaching, 7*(4).

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *phraseology: Theory, analysis, and applications* (pp. 161-168). Oxford: Oxford University Press.

Hsu, W. (2015). The Most Frequent Formulaic Sequences In College Engineering Textbooks. *Corpus Linguistics Research*, *1*, 109-149.Hyland, K. (2006). *English for Academic Purposes: An advanced resource book.* Routledge, London.

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), 1-9.

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

Jalali, Z. S., & Moini, M. (2018). A Corpus-Based Study of Lexical Bundles in Discussion Section of Medical Research Articles. *Iranian Journal of Applied Linguistic Studies, 10*(1), 94-124.

Jalali, Z. S., & Moini, M. R. (2014). Structure of Lexical Bundles in Introduction Section of Medical Research Articles. *Procedia - Social And Behavioral Sciences*, *98*, 719-726.

Jalali, H., Eslami Rasekh, A., Tavangar Rizi, M. (2009). Anticipatory 'it' lexical bundles: A comparative study of student and published writing in applied linguistics. *Iranian Journal of Language Studies, 3*(2), 177-194.

Jeblonaki, R. (2009). 'IN THE LIGHT OF: A corpus-based analysis of lexical bundles in two EU-related registers. *WoPLP,* 3, 1-27.

Kanoksilapatham, B. (2009). Generic structure of research article abstracts in sciences. *Journal of Studies in the English Language*, *4*, 95-111.

Kim, E. (2014). An Analysis of Move Patterns in Abstracts of Social Sciences Research Articles. *Journal of Korean Library and Information Science Society*, *45*(2), pp.283-309.

Kwary, D., Ratri, D. and Artha, A., (2017). Lexical Bundles In Journal Articles Across Academic Disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 131.

Kim, H. O., & Lee, H. K. (2018). Identifying the vocabulary size for Korean EFL students in linguistics. Paper presented at the 4th Asia Pacific Corpus Linguistics Conference. Takamatsu, Japan.

Kim, H. O., & Lee, H. K. (2019). The academic vocabulary list and amount in linguistics for EFL university students. *Korean Journal of English Language and Linguistics, 19*(1), 27-53.

Lee, H. K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research*, *37*(1), 121-145.

Liu, C. Y., & Chen, H. J. H. (2020). Functional variation of lexical bundles in academic lectures and TED talks. *Register Studies*, *2*(2), 176-208.

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, *39*, 21-36.

Maher, J. (1986). The development of English as the international language of medicine. *Applied Linguistics, 7*(2), 206-218

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

Mbodj-Diop, N. (2016). *Lexical Bundles in Medical Research Articles: Structures and Functions* (MA). Micigan State University.

Muşlu, M. (2018). Use of Stance Lexical Bundles by Turkish and Japanese EFL Learners and Native English Speakers in Academic Writing. *Gaziantep University Journal of Social Sciences*, *17*(4), 1568-1585.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics, 11*(1), 283-304.

O'Kee*ff*e, A., McCarthy, M. J. and Carter, R. A. (2007) From Corpus to Classroom. Cambridge: Cambridge University Press.

Panthong, P., & Poonpon, K. (2020). Functional analysis of lexical bundles in doctor talks in the medical TV series Grey's Anatomy. *LEARN Journal: Language Education and Acquisition Research Network, 13*(2), 335-353.

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.

Pho, P. D. (2008). Research Article Abstracts in Applied Linguistics and Educational Technology: A Study of Linguistic Realizations of a Study of Rhetorical Structure and Authorial Stance. *Discourse Studies, 10*(2), 231-250.

Rezoug, F., & Vincent, B. (2018). Exploring Lexical Bundles in the Algerian Corpus of Engineering. *Arab Journal of Applied Linguistics*, *3*(1), 47-77.

Subramaniam, R., & Kaur, S. (2021). A Corpus-driven Study on the Use of Passive Verb Bundles in Academic Writing: A Comparison between L1 and L2 English Speakers. *GEMA Online Journal of Language Studies*, *21*(4).

Salazar, D. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing.* University de Barcelona.

Schmitt, N., & Crater, R. (2004). Formulaic sequences in action: an introduction. In: N. Schmitt (Ed.), *formulaic Sequences* (pp. 1-22). Amsterdam/ Philadelphia: John Benjamins.

Shahmoradi, N., Jalali, H., & Ghadiri, M. (2021). Lexical bundles in the abstract and conclusion sections: The case of applied linguistics and information technology. *Applied Research on English Language*, *10*(3), 47-76.

Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, *40*, 1-14.

Shin, Y. K., Cortes, V., & Yoo, I. W. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing, 39*, 29-41.

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

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (Ed), *Developing linguistic A guide to good practice* (pp. 1-16). Oxford: Oxbow Books.

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. (1990). *Genre analysis: English in academic and research settings.* Cambridge (etc.): Cambridge University Press.

Swales, J. M. (2008). Foreword. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspective on L2 speaking, writing, and other media interactions* (p. v-viii). Ann Abor: University of Michigan Press.

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes, 49*.

Verdaguer, I., Comelles, E., Laso, N. J., Giménez, E., & Salazar, D. (2009). SciE-Lex: Anelectronic database for the Spanish medical community.

Wachidah, W. D. N. A., Fitriati, S. W., & Widhiyanto, W. (2020). Structures and Functions of Lexical Bundles in Findings and Discussion Sections of Graduate Students’ Thesis. *English Education Journal*, *10*(2), 131-142.

Wisberge, R., & Buker, S. (1990). *Writing Up Research: Experimental Research Report writing for students of English* (10th ed.). United States of America: Prentice-Hall.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review, 57*, 573-589.

Yang, Y. (2017). Lexical Bundles in Argumentative and Narrative Writings by Chinese EFL Learners. *International Journal of English Linguistics, 7*(3), 58.

ABOUT THE AUTHORS

Elshan Varghaei is a Ph.D. candidate of TEFL at Islamic Azad University, Ardabil Branch, Ardabil, Iran. Her research interests mainly include areas of EAP, (critical) discourse analysis, psycholinguistics, and complex, dynamic systems. She has published a paper in the *International Journal of All Research Education and Scientific Methods (IJARESM)* and has worked on Lexical Bundles in her MA thesis. At the moment, she is doing research in the sphere of psycholinguistics, the dynamics of foreign language learning, and discourse analysis. Address for correspondence: the English department, Islamic Azad University, Ardabil Branch, Ardabil, Basij Square, Ardabil, Iran, 5615731567. Email: varghaeielshan@gmail.com

Golsa Khodadadi received her MA in Applied Linguistics from the University of Tabriz. Her research interests include Psycholinguistics, Ellipsis, Syntax, Semantics, Bilingualism/Multilingualism, and (Critical) Discourse Analysis. She has worked on Verb Phrase Ellipsis (VPE) in different genres. Currently, she is examining the processing of VPE in bilingual settings. She has articles published in a variety of journals, including IJELS and IJARESM. Address for correspondence: The English Department, University of Tabriz, Jam-e-Jam Street, 5166616471, Tabriz, Iran. Email: Golsakhodadadi25@gmail.com