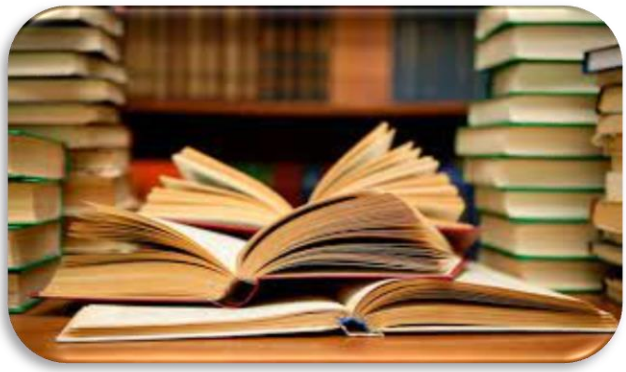
Journal of New Trends in English Language Learning (JNTELL)

Research Paper



A Comparative Corpus-driven Study of Lexical Bundles in Applied Linguistics Research Articles between Native (English) and Non-native (Iranian) Authors

Ali Alizadeh[1], Davud Kuhi[2]*, Yaser Hadidi[3]

[1] Ph.D. Candidate in ELT, Maragheh Branch, Islamic Azad University, Maragheh, Iran
alizadehali9228@gmail.com

[2] Assistant Professor of TEFL, Department of English Language and Literature, Maragheh Branch, Islamic Azad University, Maragheh, Iran.
davudkuhi@iau-maragheh.ac.ir

[3] Assistant Professor of TEFL, Department of English Language and Literature, Faculty Of Literature and Foreign Languages, University of Tabriz, Tabriz, Iran.
hadidiy@hotmail.com

Received: 28 October, 2024 **Accepted: 31 December, 2024**

ABSTRACT

This study examined the use of structures and functions of lexical bundles in scientific research articles written by native English writers and non-native English writers. To this end, three-, four--, and five-word lexical bundles were contrastively investigated in the corpora, which contained 954,615 words. The RAs were then examined structurally by Biber et al. (1999) and functionally by Hyland's model (2008). Based on the results, native authors used more lexical bundles regarding structural classification than non-native authors. Furthermore, based on the detailed analysis, the authors, no difference in whether they were native or non-native, used noun phrases with of-phrase fragments with high frequency and high rate. Regarding LBs' functions, the results indicated that the highest concentration of native and non-native corpus is on research-oriented bundles. In general, there was no significant difference in the frequency of using lexical bundles in terms of functional classification between native and non-native researchers of soft science. Regarding LBs' functions, the highest concentration of native and non-native corpus is on text-oriented bundles.

Keywords: *Lexical Bundles, Native English writers, Non-native Iranian writers, Research Articles*

این مطالعه به بررسی استفاده از ساختارها و کارکردهای بسته‌های واژگانی در مقالات علمی پژوهشی نوشته شده توسط نویسندگان بومی انگلیسی و نویسندگان غیربومی انگلیسی پرداخته است. برای این منظور، بسته‌های واژگانی سه، چهار و پنج کلمه‌ای به طور متضاد در مجموعه‌ها مورد بررسی قرار گرفتند که شامل ۹۵۴۶۱۵ کلمه بود. سپس RAها توسط Biber و همکارانش از نظر ساختاری مورد بررسی قرار گرفتند. (۱۹۹۹) و از نظر عملکردی با مدل هایلند (۲۰۰۸). بر اساس نتایج، نویسندگان بومی از بسته‌های واژگانی بیشتری در رابطه با طبقه‌بندی ساختاری نسبت به نویسندگان غیربومی استفاده کردند. علاوه بر این، بر اساس تجزیه و تحلیل دقیق، نویسندگان، بدون تفاوت در بومی یا غیر بومی بودن، از عبارات اسمی با قطعات خارج از عبارت با فراوانی و نرخ بالا استفاده کردند. با توجه به کارکردهای LBs، نتایج نشان داد که بیشترین غلظت بدنه بومی و غیربومی بر روی بسته‌های پژوهش‌محور است. به طور کلی، در فراوانی استفاده از بسته‌های واژگانی از نظر طبقه‌بندی عملکردی بین محققان بومی و غیربومی علوم نرم تفاوت معناداری وجود نداشت. با توجه به توابع LBها، بیشترین غلظت بدنه بومی و غیر بومی در بسته‌های متن‌گرا است.

کلیدواژه‌ها: مجموعه‌های واژگانی، نویسندگان بومی انگلیسی، نویسندگان غیربومی ایرانی، مقالات پژوهش

INTRODUCTION

Formulaic language refers to groups of words that work together to do specified tasks and are called up more or less automatically by native speakers (Schmitt & Carter, 2004). In other words, the formulaic language is an umbrella term that refers to multi-word units in a foreign language and elements that form the basis of a language. These structures, which take their place in the literature as multi-word expressions, are also essential to fluent linguistic production. We can say that these lexical structures are linguistic regulators that follow each other frequently, provide integrity within the text, and ensure the distinguishability of contextual structures. In other words, as Hyland (2008a) mentions, to gain control of a new language, users need to achieve sensitivity in using certain lexical structures more than others. Formulaic language structures, which can be seen as a part of phraseology, include multi-word units, idiomatic structures, repetitive sequences, collocations, and lexical bundles (LBs). Since formulaic sequences are the most often recurrent multiword sequences in a register, independent of their idiomaticity, and regardless of their structural state, LBs are of utmost significance in studying formulaic language (Biber et al., 1999).

This study compares the academic writing of native and non-native English-speaking researchers of different academic fields using frequently recurring word sequences and lexical bundles (Biber et al., 1999). Recent years have seen an increased interest in LBs, with research demonstrating that specific sets of bundles are widely used in academic genres and that their competent use is essential to academic fluency (e.g., Perez-Llantada, 2014). Many studies have examined the rate, frequencies, and structures of LBs in different academic fields and various levels of first language (L1) and second language (L2) academic writing (e.g., Breeze, 2013; Chen & Baker, 2010; Durrant, 2017; Paquot, 2017).

While useful, these previous studies have used different types of academic texts for their comparative analysis (Wei & Lei, 2011). The findings of these previous studies, therefore, may blur the differences due to the characteristics of the groups, i.e., native versus nonnative/novice versus expert, and the confounding influences of register differences, an issue that recent studies have strongly argued about (Pan et al., 2016). Such differences are expected to affect the choice and usage of lexical bundles. In addition, although there have been a few studies on LBs used by natives and non-natives (Güngör & Uysal, 2016; Latif et al., 2022), actually, there is a paucity of studies on both structural and functional analysis of LBs by natives and nonnative researchers who published scientific research articles in high tire ISI journals. Furthermore, extra research is needed. One reason for further research into the LBs, besides their importance and high frequency in texts, is the discrepancies in the results. For instance, Aziz's (2022) results showed that native speakers rely more on formulaic language than non-native users (Aziz, 2022). Contrarily, it was found in Estaji and Montazeri's study (2022) that native speakers and non-native speakers share several characteristics in their usage of bundles, including a propensity for VP-based bundles, stance-expression bundles, idiomatic PP bundles, and informal quantifying bundles. Hence, another reason to study bundles as used by L1 and L2 English writers, in particular, researchers are to learn more about features of their production, "not just with respect to the communities they are born into, but those they choose to join or hope to change or decide to create" (Belcher, 2014, p. 66). Based on the discrepancies in the results of the studies and due to the importance of research in LBs as a powerful tool for creating an understanding of the role of nativity/non-nativity, the current study was an attempt to investigate, analyze, compare, and contrast the frequency, structure,

and functions of lexical bundles in the research articles authored by native and non-native speakers of English.

REVIEW OF THE LITERATURE

Formulaic language research focusing on LBs, particularly, has flourished since the late 1990s. The term "lexical bundle" was first coined by Biber et al. (1999, ch. 13) in a chapter of the Longman Grammar of Spoken and Written English (LGSWE). LBs are characterized by how they are determined based solely on their frequency in a corpus. They are thus simply the most frequent recurrent multiword sequences in a register, "regardless of their idiomaticity, and regardless of their structural status" (Biber et al., 1999, p. 990). It should be emphasized that lexical bundles are much more than sequences of individual words; these sequences have pragmatic functions in discourse and meet recurrent communicative needs (e.g., Panthong & Poonpon, 2020). The fact that lexical bundles serve pragmatic purposes in conversation and satisfy recurring communication demands emphasizes that they are much more than just collections of individual words (Wood, 2015).

As Durant (2015) stated, due to some crucial characteristics of LBs, such as their automatically detected nature, their fulfillment of definite functional tasks, and their sensitivity to variations in text kinds, researchers can identify linguistic variations across large linguistic samples and to describe those variations in functional terms. Scholars have used these characteristics to study differences between various group types, such as L1 vs. L2 writing, academic disciplines, and student vs. professional academic writers (Zhang et al., 2021; Latif et al., 2022). It should be emphasized that several researchers (e.g., Biber, Conrad, & Cortes, 2004; Salazar, 2014) have investigated the functional typologies of bundles in discourse. Biber et al. (2004) classified bundles into three primary categories: stance expressions (e.g., it is crucial to), discourse organizers (e.g., on the one hand), and referential expressions (e.g., as indicated in the flowchart). The specific multiword sequences of formulaic language vary according to context; for instance, face-to-face conversation uses more lexical bundles that express stance; academic writing uses more referential bundles; classroom teachers regularly use all three types (Shin, 2018). Previous studies have also documented the internal structures of LBs, which vary according to register (Abdollahpour & Gholami, 2018; Hyland, 2008b; Salazar, 2014). These studies have shown that many high-frequency sequences are structurally incomplete, yet specific bundles are strongly correlated with specific types of grammatical structures. For example, Biber et al. (1999) grouped bundles into several basic structural types based on their occurrence in the Longman Grammar of Spoken and Written English. They found that conversation uses more clausal bundles, which consist of the verb phrase (VP) fragments (e.g., is going to be) and dependent clause fragments (e.g., if you attempt to). In contrast, academic prose uses more phrasal bundles, as in noun phrases (NPs) (e.g., the essence of the) or prepositional phrases (PPs) (e.g., in the field of).

Lexical bundles are common in academic genres and exhibit qualities specific to academic writing because they form the fundamental building block of discourse in academic writing (Biber et al., 2004). Researchers and postgraduate students of any major must submit research articles in their fields because admission to higher education programs may be contingent on the number of articles they can publish in reputable publications. Research articles that have been published can be used as academic

writing. Scholars can advance their careers and achieve success by publishing in renowned publications. As a result, one of the crucial competencies for academics across all disciplines is the ability to write research articles for publication that are strengthened by the accurate use of LBs (Güngör & Uysal, 2020). According to Yoon and Choi (2015), the use of lexical bundles in academic writing gives language learners a clear indication of the appropriate and fluent written form required in academic topics. This study and previous studies on lexical bundles in academic writing have supplied significant evidence supporting the significance of prepared phrases in this type of discourse. These studies show that using lexical bundles frequently and effectively is a crucial component of fluent language production in academic settings, helping to create meanings in specific contexts and contributing to our sense of text coherence (Hyland, 2008a).

The current research focuses on nativity/non-nativity, comparing the use of LBs by the authors who were different in first language. Some studies have been done so far; for instance, Tribble (2011) compared LBs in native-speaker student academic writing (MA-level assignments and dissertations) and published research articles in equivalent fields. The findings showed that the student writers used only a small number of bundles common in research articles. Tribble argued that the bundles absent in student writing (for instance, in terms of the, in the case of) were often those that functioned as “framing markers” (p. 94), which suggests that students may need specific training in how to employ the bundles that give textual coherence to academic writing. Furthermore, Adel and Erman (2012) investigated bundles used by non-native undergraduate students (L1 Swedish) in the discipline of linguistics. They found that Swedish students’ writing displayed fewer English bundles than comparable native writing, especially regarding hedge devices. Comparing the structural and functional aspects of the use of lexical bundles in L1 and L2 research articles in English, similar to this work, was done by Güngör and Uysal (2016). The findings showed how non-native English speakers' utilization of lexical bundles varied from that of native speakers. Furthermore, the findings showed that Turkish researchers overused verb- or clausal-based vocabulary bundles in their research publications. However, their native counterparts favored noun- and prepositional-based lexical bundles over clausal-based ones. The use of lexical bundles in the works of two groups of university students, including native English and Chinese speakers, was examined by Bychkovska and Lee (2017). The study, which involved 206 article studies, found that second-language learners used lexical bundles more frequently than first-language learners. However, there were differences in how often structural and functional lexical bundles were used. Akbulut (2020) sought to compare the features of LBs in terms of function, structure, and frequency used by native and non-native academic writers of English. AntConc 3.2.4w was used to arrange a corpus of 257 academic articles published by native and non-native writers and to identify multi-word LBs. According to the research, non-native writers tend to repeat themselves more but employ more lexical bundles overall. In addition, there are also noticeable variations in the frequency of LB structural and functional utilization. In conclusion, it is said that the distinguishing aspects of LB usage were revealed by the lexico-grammatical differences between the two languages and the NNSs' propensity to create result-oriented articles.

Khamkhien (2021) aimed to explore how Thai L2 undergraduate students use lexical bundles in their academic papers written in English and to compare the use of lexical bundles with that in two written corpora: the British Academic Written English (BAWE) and Cambridge Academic English (CAE). The results show that native and non-native writers use lexical bundles differently. Thai L2 English students

overused some patterns that did not occur in the reference corpus. More recently, Azizi (2022) compared the use of lexical bundles in academic writing in Applied Linguistics across three corpora: expert writers, native students, and non-native students. The findings revealed that the expert writers were different from native and non-native students in their use of structural and functional bundles. Interestingly, most of the differences between expert and student writers in using bundles applied to both sets of students. This suggests that the main challenge for all students is learning the conventions of academic writing rather than any problems linked to non-nativeness. Therefore, the appropriate use of bundles in academic writing might need to be taught more explicitly to both native and non-native students. Besides, Estaji and Montazeri (2022) examined the use of lexical bundles in the results and discussion sections of public health research articles by comparing native English writers with Iranian non-native English writers. The corpus contained 496,985 words, and each sub-corpora included 100 RAs. The findings highlight that Iranian non-native English authors employed more four-word, five-word, and six-word lexical bundles than native English authors. The descriptive and overall findings also suggested some differences in the two groups' functional and structural patterns of lexical bundles. In contrast, statistically insignificant differences were identified in the structural patterns of bundles in the groups.

As it is axiomatic, there is a bulk of research on lexical bundles in general and the utilization of LBs by different groups, including natives and non-natives. This research focus is needed as it would be interesting to investigate whether different paradigms and different mother tongues may affect the use of lexical bundles and to compare and contrast the frequencies and functions of LBs by two groups of authors that are being native or non-native of English to facilitate the teaching-learning of the bundles to those groups while writing highly prestigious articles. Hopefully, this research will add to the theoretical and practical application of lexical bundles in training academic writing skills. Specifically, this research would assist students, especially non-natives, improve their writing fluency and precision in academic settings. As Cortes (2023) stated, a study on the lexical bundles might improve students' academic writing fluency and precision. Furthermore, this research aids EFL instructors, students, and material authors by highlighting the significance of lexical bundles in teaching materials, curricula, and classroom instruction. Based on the objectives of the study, the following research questions were formulated:

RQ1: Is there any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from the structural point of view between research articles written by native and non-native authors?

RQ2: Is there any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from the functional point of view between research articles written by native and non-native authors?

METHODOLOGY

Corpus

This study analyzed lexical bundles used in the research papers authored by native and non-native speakers of English. The research articles were from different fields and disciplines, such as psychology, sociology, applied linguistics, computer science, chemistry, and medicine. The research articles published in top-tier ISI journals from Sage, Elsevier, Taylor and Francis, Springer, and Wiley Online

Library between the years 2010 to 2020 were randomly selected. The papers were randomly selected from different journals, with special attention to the authors' affiliations. Regarding nativity and non-nativity, the researcher tried to select just the articles written by American and England authors (as native speakers of English) based on their affiliations, following Yılmaz and Özdem Ertürk (2017). Furthermore, non-native English writers were Iranian researchers. This research describes how lexical bundles are distributed across each corpus. A total of 180 articles, 90 by native writers and 90 by non-native writers, were selected and analyzed. The data of the study was small and specialized. Justification for the use of small size specialized corpus can be found in the writings of several authors like Flowerdew and Forest (2009). They suggest that a corpus that includes the same genre and discipline texts may produce sufficient data for the analysis regardless of their size. Besides, a small corpus enables some analyses requiring the hand-coding of LBs in terms of functions, which cannot be handled manually within large data (Flowerdew & Forest, 2009).

Procedure

At the onset of the study, first, the 180 articles were downloaded from the key ISI databases, including Elsevier (Science Direct), Sage, and Cambridge publications. Then, they were extracted based on convenient and purposive sampling (nativity/non-nativity). The articles were selected from journals, published between 2010 to 2020, and indexed in reliable databases. Only papers published in various issues from 2010 to 2020 with Modified Impact Factors (MIFs) ranging between 0.5 and above were included in the corpus to reflect better the genre aspects of the most recently published research articles. To this end, 180 research articles from online versions of selected journals were downloaded and converted to Word document files. The files were cleaned of headers, footers, figures, pictures, titles, references, irregular capitalizations, and paragraph breaks to ensure smooth and accurate data processing. The lexical bundles were calculated using Biber et al.'s (1999) structural classification model and Hyland's (2008) functional classification model. Biber et al. (1999) were the first to develop the structural taxonomies of LBs.

Concerning the structure of LBs, Biber et al. (1999) divided LBs into three major structural types. Type 1 LBs embrace verb phrase fragments (e.g., *is underlying to be*). Type 2 LBs include dependent clause fragments in addition to simple verb phrase fragments (e.g., *what I need to*), and type 3 LBs incorporate noun phrase and prepositional fragments (e.g., *the mean of the, of the elements that*). These grammatical features are considered to appear differently depending on the register. That is, bundles in conversation tend to be clausal (e.g., *it is going to be*), while in academic prose, most bundles are commonly phrasal (e.g., *as a consequence of*) (Biber et al., 1999). In terms of the functions of LBs, this study used Hyland's model (2008a) to analyze the functions of the LBs. The subsequent scientific categorization allocates every bundle to one of three general research classes, text, and participants, which are further isolated into different subcategories (see Table 1).

Table 1

Functional Classification of Lexical Bundles in Academic Writing (taken from Hyland, 2008a, pp. 13-14)

Research-oriented bundles Help writers to structure their activities and experiences of the real world	**Text-oriented bundles** Concerned with the organization of the text and its meaning as a message or argument	**Participant-oriented bundles** Focused on the writer or reader of the text
Location	**Transition signals**	**Stance features**
Indicating time/place	Establishing additive or contrastive links between elements	Convey the writer's attitudes and evaluations
at the beginning of, at the same time, in the present study	*on the other hand, in addition to the, in contrast to the*	*are likely to be, may be due to, it is possible that*
Procedure bundles	**Resultative signals**	**Engagement features**
the use of the, the role of the, the purpose of the, the operation of the	Mark inferential or causative relations between elements	Address readers directly
Quantification	*as a result of, it was found that, these results suggest that*	*it should be noted that, as can be seen*
the magnitude of the, a wide range of, one of the most	**Structuring signals**	
Description	Text-reflexive markers which organize stretches of discourse or direct the reader elsewhere in text	
the structure of the, the size of the, the surface of the		
Topic	*in the present study, in the next section, as shown in figure*	
related to the field of research	**Framing signals**	
in the Hong Kong, the currency board system	Situate arguments by specifying limiting conditions	
	in the case of, with respect to the, on the basis of, in the presence of, with the exception of	

All elements were calculated per 10,000 words for two types of research articles. To test the study's hypotheses, the frequency of each lexical bundle in native- and non-native-authored articles in different academic fields was calculated and compared. This study examined 3-to-5-word structures. This was done to capture the full range of formulaic language used by each group of writers and to determine if lexical bundle length was related to assessed proficiency. AntConc sorted lexical frequency and range.

Lexical bundles must meet the method's frequency and range criteria. The list was also excluded based on five pedagogical criteria. All master list word sequences must meet frequency and range criteria, but that does not mean they all fall within the scope of the study or benefit learners equally. Five additional criteria were established to narrow the list of lexical bundles. Titles, figures, pictures, tables, charts, formulas, acknowledgments, reference lists, bio data, appendices, and authors' information were manually removed to eliminate possible factors affecting data analysis and to ensure the corpus texts were computer-readable.

The researcher used frequency to identify the most lexical bundles in the corpus. AntConc (2012) was used to identify the most common lexical bundles. AntConc is a freeware concordance program that offers word lists, n-grams, collocates, and clusters (Anthony, 2012). The concordancer scans the corpus word by word for multiple-word bundles. Proper nouns, like institution names, were omitted from the lists because they did not help the research. Mathematical variables and symbols detected as lexical bundles, like u v u v, were also removed.

To ensure the reliability of the analysis in the process of data categorization, 20% of the data was rechecked and reanalyzed independently for LBs d by a second researcher (a Ph.D. graduate of TEFL) who was briefed about the purpose of the study by the researcher. Also, the field of study of this expert was discourse analysis, and she was familiar with the data analysis phase. The second rater coded 20% of the data, taken randomly from the corpus, and finally, the inter-rater reliability was estimated and reported. The inter-rater agreement, measured using Cohen's Kappa formula, was found to be Kappa = 0.929, p = 0.000. After all of these procedures, the gathered data was analyzed via Chi-square data analysis to explore the existence of any significant difference between native and non-native authors in the use of lexical bundles.

RESULTS

To find the existence of any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from the structural point of view between research articles written by native and non-native authors in different academic fields (the first research question), the frequencies and use of lexical bundles from the structural point of view between research articles written by native and non-native authors were gathered and reported. To this end, a corpus of 180 research articles of approximately954,615 words from recently published Institute for Scientific Education-indexed journals that were authored by non-native of English (94814 for applied linguistics; 109830 for sociology; 109830 for psychology; 75825 for computer; 69355 for medicine; 26525 for chemistry) and native speakers (101327 for applied linguistics; 97437 for sociology; 105672 for psychology; 86334 for computer; 71869 for medicine; 24988 for chemistry) was compiled and analyzed. Table 2 shows the structural categorizations of LBs in the research articles written by native and non-native speakers of English.

Table 2

Structural Classification of Lexical Bundles Written by Native and Non-Native Authors

STRUCTURE	EXAMPLES

		Native Authors		Non-Native Authors	
		Frequency	Percentage	Frequency	Percentage
Noun phrase with *of*-phrase fragment	the end of the, the beginning of the, the base of the, the point of view of	121	12%	110	11.7%
Noun phrase with other post-modifier fragments	the way in which, the relationship between the, such a way as to	73	7.2%	67	7.1%
Prepositional phrase with embedded *of*-phrase fragment	about the nature of, as a function of, as a result of the, from the point of view of	65	6.3%	64	6.8%
Other prepositional phrase (fragment)	as in the case, at the same time as, in such a way as to	52	5.1%	43	4.6%
Anticipatory *it* + verb phrase/adjective phrase	it is possible to, it may be necessary to, it can be seen, it should be noted that, it is interesting to note that	59	5.8%	70	7.4%
Passive verb + prepositional phrase fragment	is shown in figure/fig., is based on the, is to be found in	137	13.6%	122	13.6%
Copula *be* + noun phrase/adjective phrase	is one of the, may be due to, is one of the most	108	10.7%	102	11%
(Verb phrase +) *that*-clause fragment	has been shown that, that there is a, studies have shown that	89	8.8%	82	8.7%
(Verb/adjective +) *to*-clause fragment	are likely to be, has been shown to, to be able to	76	7.5%	61	6.5%
Adverbial clause fragment	as shown in figure/fig., as we have seen	75	7.1%	65	6.9%
Pronoun/noun phrase + *be* (+…)	this is not the, there was no significant, this did not mean that, this is not to say that	68	6.5%	69	7.3%
Other expressions	as well as the, may or may not, the presence or absence	96	9.1%	79	8.4%
Total		1008	100%	934	100%

Based on what is presented in Table 2, it can be claimed that the native authors in different disciplines used more LBs in comparison to the non-natives; furthermore, both of the authors, with no difference in their being native or non-native ones, used noun phrase with *of*-phrase fragment with high frequency than lexical bundles as they form 12% and 11.7% by native and non-native writers respectively of the whole corpus. As it is clear from the above table, two groups of authors used the other sub-categories related to the structural aspect of LBs with similar rates and percentages. However, there is a need to run a Chi-square data analysis to see the significant differences between the two groups. The results of the Chi-Square analysis, x^2 = .016, p = 1.000, revealed that differences in relation to the frequencies of LBs in terms of structural classification across native and non-natives were not statistically significant. In other words, although proportionately more LBs are used by native speakers and although in some sub-categories, natives used more structural sub-categories in comparison to the non-natives or vice versa, actually, the differences are not found to be statistically different. As a result, the hypothesis of the research question was retained. Since there was no significant association, there was no need to report Cramer's V as the strength of association.

The second research question posed in the present study aimed to investigate whether there is any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from a functional point of view between research articles written by native and non-native authors (See Table 3).

Table 3

Functional Classification of Lexical Bundles Written by Native and Non-Native Authors

Major Functions	**Sub-Categories**	Native Authors		Non-Native Authors	
		Frequency	Percentage	Frequency	Percentage
Research-oriented	**Location-** indicating time and place, e.g. *in the present study*.	84	12.8%	74	12.3%
	Procedure- indicating methodology or purpose of research, e.g. *the purpose of this*.	64	9.8%	60	10%
	Quantification- describing the amount or number, e.g. *is one of the*.	62	9.4%	34	5.6%
	Description- detailing qualities or properties of material, e.g. *in the control group.*	35	5.3%	49	8.1%
	Topic- related to the field of research, e.g. *in the United States.*	50	7.5%	45	7.5%
Text-oriented	**Transition signals-** establishing additive or contrastive links between elements, e.g. *on the other hand, as well as the.*	77	11.6%	73	12.1%

	Resultative signals- mark inferential or causative relations between elements, e.g. *the results of the*.	50	7.5%	54	9.2%
	Structuring signals- text-reflexive markers which organize stretches of discourse or direct readers elsewhere in the text, e.g. *as shown in fig*.	78	11.8%	63	10.5%
	Framing signals- situate arguments by specifying limiting conditions, e.g. *in the presence of.*	86	13.4%	60	10.3%
Participant-oriented:	**Stance features-** convey the writers' attitudes and evaluations. According to Cortes (2002), this category includes attitude markers, epistemic-certain, epistemic-uncertain and intention bundles, e.g. *were more likely to*.	37	5.6%	37	6.2%
	Engagement features- address readers directly, e.g. *it should be noted*.	38	5.5%	50	8.4%
Total		659	100%	599	100%

As shown in Table 3, the highest concentration of native speakers in the corpus is on Framing signals- (situate arguments by specifying limiting conditions, e.g., *in the presence of*), with 13.4% at the first place and 12.8% in the Location sub-category (indicating time and place, e.g., *in the present study*) that is under the category of Research-oriented, at the second place. Moreover, non-native speakers used the same category and sub-category with a high percentage of 12.3%. To see the existence of any significant difference in the functions of LBs, a Chi-square data analysis was run. The findings indicated that x^2 = .017, p = .092, revealed that differences in relation to the frequencies of LBs' in terms of functional classification across native and non-natives were not statistically significant. In other words, there is no significant difference in the frequency of lexical bundles from functional classification lenses written by native and non-native authors. Additionally, the difference is not statistically significant because the *p*-value (.092) of the Chi-square test is not lower than 0.05.

DISCUSSION

This study aimed to explore lexical bundles in terms of structural and functional categories in research articles written by native and non-native authors in different academic fields. Biber et al.'s (1999) model for structural classification and Hyland's (2008) model for functional classification were very successful not only in describing the overall framework of the research papers in both corpora but also in the detailed definition and description of the individual steps and sub-steps realizing the lexical bundles. The results indicated no significant difference in the frequency of lexical bundles from the structural lenses written

by native and non-native authors. Although native authors used more lexical bundles in terms of structural classification compared to non-native authors, there was no significant difference between the two authors. Furthermore, based on the detailed analysis, both of the authors, with no difference in their being native or non-native ones, used noun phrases with of-phrase fragments with higher frequency than lexical bundles as they form 12% and 11.7% by native and non-native writers, respectively, of the whole corpus. As it is clear from the above table, two groups of authors used the other sub-categories related to the structural aspect of LBs with similar rates and percentages.

In this study, too many instances of noun phrases could be due to the fact that academic writing is informational in nature, and informational integration requires using noun phrases (Halliday & Hasan, 1989; Pan et al., 2016). The results are in line with Abdollahpour and Gholami's study that investigated frequently-used four-word general and technical lexical bundles (LBs) in the abstract sections (ASs) of research articles (RAs) in medical sciences. Like the current study, they found that medical abstracts structurally contain a wider range of noun phrase bundles (e.g., *this study was designed*) than clausal phrases. Their justification for the abundant use of phrasal structures in abstracts is that abstracts, as an important piece of academic discourse, are more compressed than elaborated. Therefore, this complexity leads to phrasal embedding rather than a clausal one.

The focus of the second question was again on authors with differences in L1s but in terms of LBs' functions. The results indicated that the highest concentration of native speakers in the corpus is on framing signals- with 13.4% at the first place and 12.8% in the Location sub-category, which is under the category of Research-oriented at the second place. Moreover, non-native speakers used the same category and sub-category with a high percentage of 12.3%. To see the existence of any significant difference in the functions of LBs, a Chi-square data analysis was run. The findings indicated that differences in relation to the frequencies of LBs in terms of functional classification across natives and non-natives were not statistically significant. The justification for the result can be the fact that the non-native authors (similar to the native ones) are familiar with the principles of academic writing in general and the functions of LBs in particular. Hence, the difference was not significant between the two groups of authors. This is a hunch; its validity can be measured via author interviews. The results are in line with Amirian, Ketabi, and Eshaghi's (2013) study that, like this study, studied the nature of lexical bundles in native and non-native post-graduate students' writing. The differences between the above-mentioned study and the current one are the differences between the two groups regarding LBs' functions. Based on their results, Iranian students show more concern about mentioning the procedures of completing their research by using more research-oriented bundles. In contrast, native students try to organize their writing by using more text-oriented bundles.

The results of this study again prove that lexical bundles are constructing elements in academic discourse. Their importance is proved by their high frequency in the different corpora used in this study.

However, what is remarkable about this piece of research is that for the first time, research articles of native and non-native speakers of English are compared not just in one major and academic field, such as Applied Linguistics, but in different academic fields. It is true that previous studies such as Jalali (2008) and Hyland (2008a) investigated lexical bundles in post-graduate writing. However, the focus in those studies was not on whether the writers were native speakers of English or not and whether they had different majors.

As it is clear crystal, the native authors used exactly the same passive verb + prepositional phrase fragment (13.6%) in the first place and more noun phrases in the second place. These findings corroborate the fact that verbs, especially passive verbs, are the most common forms in academic discourse, and in most cases, they are integrated into propositional phrases and presented in scientific papers. The results are in harmony with a study conducted by Cortes (2008) on comparing lexical bundles in published history articles in English and Spanish. The findings revealed that most of the bundles in both corpora were prepositional phrases. She found noun phrases as her study's second most frequent structural category. Her findings also approved that most of the bundles in academic writing, in general, and research articles, in particular, are essentially phrasal. To sum up, it is worth noting that it seems crucial that novice researchers in different academic fields be trained to establish a territory in their research papers by providing background information, referencing previous research, and emphasizing the topic's centrality. They should also include a research gap in their research articles by citing a lack of research or a limitation in previous research.

CONCLUSION

This study considers the lexical bundles and compares and contrasts them in two groups of native and non-native authors. To this end, the employed LBs were analyzed both structurally and functionally by two well-known frameworks postulated by Biber et al.'s (1999) model for structural classification and Hyland's (2008) model for functional classification. Based on the results, there was no significant difference in terms of the frequency of lexical bundles from the structural lenses written by native and non-native authors. Interestingly, the same result was obtained regarding the functions of LBs, and the findings showed that the authors overthought different functions of LBs at various rates. Generally, there was no meaningful difference between the two groups of authors in the functions of LBs. the results empowered our hunch of the non-native researchers' high familiarity with both structures and functions of LBs. The results have implications for different key stakeholders. This study may help explain the importance of lexical bundles in academic writing, especially academic articles. This study can empower Iranian writers by teaching them about word clusters, lexical bundles, and how experts use these resources to write more specialized texts, making their writing stand out from others. Researchers and educators can help non-native English writers use lexical bundles preferred by community members in different fields of science by highlighting their strengths and weaknesses in using them in research articles. The study also educates writers of different L1 majors about lexical bundles community members use. In terms of the study's limitations, the most important one is that, although the biographies of all authors in the study have been examined, some authors may have different backgrounds. Namely, although an Iranian author wrote the article, this author may have lived for many years in a country where English is spoken as the mother tongue or can have acquired enough competence in the second language like a native speaker or may be bilingual or multilingual. This study used two groups of researchers to analyze their writings in terms of LBs' structures and functions in different academic fields; for further research, it is suggested that LBs will be analyzed via move sequences and the frequency of occurrence of moves, steps, and sub-steps. It is also suggested that this study be replicated by interested researchers to use different frameworks or integration of models to empower the results. Considering the results of

the current study, some qualitative studies can also be carried out to analyze the usage of lexical bundles with a small and specialized corpus since corpus-based studies are likely to contribute to the development of writers and the design of academic writing courses. To conclude, it is worth noting that lexical bundles are acquired incrementally, just like single words. Based on this fact, learners are in need of a large number of repeated exposures to acquire lexical bundles. In this aspect, noticing, retrieval, and generative activities such as rephrasing, substitution tasks, writing activities, techniques, and tasks are some of the many ways that writing instructors can benefit in the EFL classroom to enhance learners' successful acquisition and retention of these multiword combinations.

REFERENCES

Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: Frequency, functions, and structures of lexical bundles. *Asian ESP Journal*, *14*(1), 82–110.

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31, 81-92.

Alise, M.A. & C. Teddlie. (2010). A continuation of the paradigm wars? Prevalence rates of methodological approaches across the social/behavioral sciences. *Journal of Mixed Methods Research* 4 (2), 103–26.

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.

Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 105–127.

Amirian, Z. Ketabi, S., & Eshaghi, H. (2013). The use of lexical bundles in native and non-native post-graduate writing: The case of applied linguistics MA theses. *Two Quarterly Journal of English Language Teaching and Learning University of Tabriz*, *5*(11), 1–29.

Ang, L. H., & Tan, K. H. (2018). Specificity in English for Academic Purposes (EAP): A corpus analysis of lexical bundles in Academic Writing. *3L, Language, Linguistics, Literature*, *24*(2).

Aziz, S. (2022). *Use of lexical bundles in academic writing in English by expert writers, native students, and non-native students in Applied Linguistics* (Doctoral dissertation, University of Essex).

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Johns Benjamins Publishing Company.

Biber D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistic*, *14*(3) 275-311

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written register. *English for specific purposes*, *26*(3), 263-286.

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., Johnsson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Breeze, R. (2013). Lexical bundles across four legal genres. *International journal of corpus linguistics*, *18*(2), 229-253.

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

Cao, F. (2021). A comparative study of lexical bundles across paradigms and disciplines. *Corpora*, *16*(1), 97-128.

Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language and 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 , 43–58.

Cortes, V. (2023). Lexical Bundles in EAP. In *The Routledge Handbook of Corpora and English Language Teaching and Learning* (pp. 220-233). Routledge.

De Chazal, E., Rogers, L., & McCarter, S. (2013). *Oxford EAP: a course in English for academic purposes.* Oxford University Press.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Estaji, M., & Montazeri, M. R. (2022). Native English and non-native authors' utilisation of lexical bundles: A corpus-based study of scholarly public health papers. *Southern African Linguistics and Applied Language Studies*, *40*(2), 177-199.

Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *Tesol Quarterly*, *37*(3), 489-511.

Flowerdew, J., & Forest, R. (2009). Schematic structure and lexico-grammatical realization in corpus-based genre analysis: The case of research in the PhD literature review. *Academic writing: At the interface of corpus and discourse*, 15-36.

Friginal, E., & Mustafa, S. S. (2017). A Comparison of U.S.-Based and Iraqi English Research Article Abstracts Using Corpora. *Journal of English for Academic Purposes*, 25, 45-57.

Fuertes-Olivera, P. A. (2009). The function theory of lexicography and electronic dictionaries: Wiktionary as a prototype of collective free multiple-language internet dictionary. *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*, *90*, 99.

Goh, C. L., & Lepage, Y. (2019). Extraction of lexical bundles used in natural language processing articles. In *2019 International Conference on Advanced Computer Science and information Systems (ICACSIS)* (pp. 223-228). IEEE.

Granger, S. (2018). Formulaic sequences 1 in learner corpora: Collocations and Lexical Bundles. *Understanding Formulaic Language*, 228-247.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, *52*(3), 229-252.

Güngör, F., & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, *9*(6), 176-188.

Güngör, F., & Uysal, H. H. (2020). Lexical bundle use and crosslinguistic influence in academic texts. *Lingua*, *242*, 102859.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of pragmatics*, *34*(8), 1091-1112.

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

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

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

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58-69.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL quarterly*, *41*(2), 235-253.

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information Science and Management*, *13*(1), 51-69.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes*, *24*(3), 269-292.

Khamkhien, A. (2021). Functional patterns of lexical bundles and limitations in academic writing by Thai L2 English learners. *Journal of English Language Teaching and Linguistics*, *6*(3), 607-632.

Lake, W. M., & Cortes, V. (2020). Lexical bundles as reflections of disciplinary norms in Spanish and English literary criticism, history, and psychology research. *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*, 183-204.

Latif, F., Dodhy, S., & Tajammul, M. (2022). A Comparative analysis of lexical bundles used by native and non-native novel writers. *Pakistan Journal of Social Research*, *4*(2), 1216-1231.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the Discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, *39*, 45-61.

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.

Moynie, J. (2018). *Lexical bundles within English for academic purposes written teaching materials: A canadian context*. (MA thesis) Carleton University, Ottawa, Ontario.

Narkprom, N., & Phoocharoensil, S. (2022). Lexical Bundles in Native English Speakers' and Thai Writers' Dissertations. *GEMA Online Journal of Language Studies*, *22*(3).

Omidian, T., Siyanova-Chanturia, A., & Durrant, P. (2021). Predicting Parameters of Variation in the Use of Academic Multiword Expressions in University Student Writing. *Vocabulary Theory, Patterning and Teaching*, *152*, 141.

Pang, W. (2010). Lexical bundles and the construction of an academic voice: a pedagogical perspective. *Asian EFL Journal*, 47, 2-13.

Panthong, P., & Poonpon, K. (2020). Lexical bundles in Thai medical research articles. *Journal of Studies in the English Language*, *15*(1), 59-106.

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

Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, *10*(2), 231-250.

Rajaeian, P., & Rabbani Yekta, R. (2021). Comparing Lexical Bundles in Hard Science Lectures; A Case of Native and Non-Native University Lecturers. *International Journal of Research in English Education*, *6*(4), 106-127.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual review of applied linguistics*, *31*, 205-225.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. *Formulaic sequences: Acquisition, processing and use*, 1-22.

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

Shin, Y. K. (2018). The Construction of English Lexical Bundles in Context by Native and Nonnative Freshman University Students. *English Teaching*, *73*(3), 115-139.

Toulmin, S. (1972). Rationality and scientific discovery. In *PSA: Proceedings of the Biennial meeting of the philosophy of science association* (Vol. 1972, pp. 387-406). Cambridge University Press.

Ucar, S. (2017). A Corpus-Based Study on the Use of Three-Word Lexical Bundles in the Academic Writing by Native English and Turkish Non-Native Writers. *English Language Teaching*, *10*(12), 28-36.

Wingate, U. (2012). 'Argument!'helping students understand what essay writing is about. *Journal of English for academic purposes*, *11*(2), 145-154.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463-489.

Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 590–597). Oxford: Elsevier.

Wood, D. (2015). Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon?. *Journal of Personality and Social Psychology*, *108*(2), 317-329.

Yakut, I., Yuvayapan, F., & Bada, E. (2021). Lexical bundles in L1 and L2 English doctoral dissertations. *Journal of Teaching English for Specific and Academic Purposes*, 475-493.

Yilmaz, M., & Özdem Erturk, Z. (2017). A Contrastive Corpus-Based Analysis of the Use of Reporting Verbs by Native and Non-Native ELT Researchers. *Novitas-ROYAL (Research on Youth and Language)*, *11*(2), 112-127.

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), 100-115.

Yoon, C., & Choi, J. M. (2015). Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education*, *16*(3), 47-69.

Zhang, S., Yu, H., & Zhang, L. J. (2021). Understanding the sustainable Growth of EFL students' writing skills: differences between novice and expert writers in their use of lexical bundles in academic writing. *Sustainability*, *13*(10), 5553.