

A Corpus-based Study on the Use and Syntactic Functions of Lexical Bundles in Applied Linguistics Research Articles in Two Contexts of Publications

Rajab Esfandiari^{1*}, Mohammad Ahmadi², Edward Schaefer³

¹ Associate Professor of applied linguistics, Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran

² Visiting Professor of TEFL, Department of English Language, Faculty of Humanities, Imam Khomeini International University, Qazvin, Iran

³ Professor Emeritus, Graduate School of Humanities and Science, Ochanomizu University, Tokyo, Japan and Adjunct Instructor, Dokkyo University, Saitama, Japan

Received: 2021/10/02

Accepted: 2021/10/27

Abstract: The present study investigated the use of lexical bundles (LBs) in research articles authored by English L1 and Persian L1 academic writers, with a special focus on the syntactic roles of LBs in a larger context of sentence level. Four-word bundles were retrieved and classified structurally. The use of identified LBs was compared in two writer groups. The syntactic roles and relative complexity of the bundles' structures were analyzed in relation to Biber, Gray, and Poonpon's (2011) hypothesized stages of writing development. The results indicated different patterns of reliance on LBs, with Persian writers making greater use of LBs at a higher frequency. In addition, Persian academic writers tended to use high-frequency bundles differently from native-speaker academic writers. The results of the syntactic analysis of LBs reflected more frequent use of LBs functioning as compressing lexico-grammatical structures in a native English-speaker corpus, which is indicative of a more complex academic register compared to that of a Persian L1 corpus. The pedagogical implications of the findings for the explicit instruction of syntactically complex corpus-driven LBs for discipline-specific genre writing and suggestions for future research are discussed.

Keywords: Lexical bundles, Syntactic Complexity, Research Articles.

* Corresponding Author.

Authors' Email Address:

¹ Rajab Esfandiari (esfandiari@hum.ikiu.ac.ir), ² Mohammad Ahmadi (ahmadi.m8362@gmail.com), Edward Schaefer (schaefer@dokkyo.ac.jp)

Introduction

The study of multiword sequences (MWS) has drawn the attention of researchers over the past few years. This interest has its roots in the pervasiveness of MWSs and psycholinguistic explanations which suggest a processing advantage for MWSs compared with the sequences of words that are processed individually (Conklin & Schmitt, 2008). This processing advantage is attributed to the “holistic nature of formula” in both L1 and L2 (Jiang & Nekrasova, 2007, p. 433). The psycholinguistic validity of MWSs has been strengthened in different studies (e.g., Ellis & Simpson-Vlach, 2009), where formulas have been found to have a processing advantage as well as clearly defined functions, particularly in English for academic purposes (EAP).

The function of MWSs has been specifically investigated in EAP. The bulk of the studies has documented that academic writing relies, to a great extent, on formulaic sequences (e.g., Ruan, 2017; Wei & Lei, 2011). This line of research mainly used MWSs as a linguistic means to analyze different text types produced by native/nonnative or expert/novice academic writers. While the findings of these studies broaden our knowledge of the construction of MWSs by different writer groups, they are, by no means, conclusive, as many of them have confounded ‘register/discipline’, L1, genre, audience, and topic “with the difference between groups of writers (e.g., comparing general essays written by students to research articles written by professionals)” (Pan, Reppen, & Biber, 2016, p. 62).

A particular type of formulaic sequence is LBs, which are defined as the combination of words that recur most commonly in a given register (Biber, Johansson, Leech, Conrad, & Finegan, 1999). They are of special importance in academic writing as they fulfill important discourse functions and are a hallmark of advanced academic writing (Pan & Liu, 2019). Previous studies mainly drew on a structural and functional framework of lexical bundles following Biber et al. (1999), Biber, Conrad, and Cortes (2004), and Hyland (2008). However, the syntactic function of lexical bundles within the unit of sentence length has received little attention in previous literature. This is particularly important because lexical units do not stand alone; rather, they are parts of larger units embedded within a sentence. As Shin (2018) pointed out, previous studies largely analyzed LBs within phrases and clauses; however, these units might not always be appropriate because “a bundle’s last word is often the first word of another structure” (p. 116). Shin further calls for the extension of the scope of the structural unit of LBs to the sentence level in order for researchers to be able to examine different syntactic roles of bundles within a sentence, as the same LBs which have been determined on the basis of frequency can occur in different syntactic units which function differently.

There has been surprisingly little research investigating the syntactic functions of LBs in academic writing. One of the few existing relevant studies was conducted by Shin (2018), who explored LBs situated in the texts produced by native and nonnative-speaker freshman university students. However, the present study is different from that of Shin. Although both studies investigated LBs in the academic genre, the present study employed published journal articles to construct the corpus while the study by Shin made use of argumentative essays written by university freshmen. A research article (RA) is a completely different sub-genre from those produced by student writers, “with a different purpose, audiences, and repertoire of rhetorical features” (Hyland, 2008 p. 57). RAs are the most important sub-register of professional academic writing (Biber & Gray, 2010).

Conventional analysis of LBs within phrasal or clausal units will result in a list of fragmented bundles which provide very little information with regard to their syntactic properties. Bundles do not stand alone; rather, they are incorporated into larger structures, so understanding the ways in which they are used to form larger units can help learners produce texts that read more target-like (see Garner, Crossley, & Kyle, 2019). Accordingly, the results obtained from the present study may offer more insights into the way syntactic roles of LBs contribute to the construction of expert academic registers in native and nonnative contexts. Therefore, the present study aimed at filling the gap in the literature by extending the structural unit of LBs to sentence level so that their syntactic properties will be appropriately analyzed.

In addition, previous studies have been inconclusive with regard to native versus non-native speaker contrast of LBs in academic writing with some studies showing native speakers’ heavier reliance on bundles for constructing the texts (e.g., Atai & Tabandeh, 2015) while the others showing the opposite (e.g., Esfandiari & Barbary, 2017; Rahimi Azad & Modarres Khiabani, 2018). As a consequence, more studies are required to investigate the role of native speaker status in the frequency distribution, overuse, and underuse of formulaic language in advanced academic writing, as the results could build up a clearer picture of academic formulaicity in the important sub-register of RAs. Moreover, previous studies did not provide clear evidence as to whether different distributional patterns of LBs will result in a more/less complex discourse style in relation to existing taxonomies of academic writing development. Accordingly, the purpose of the present study is to provide more understanding of the way native and nonnative academic writers employ LBs in applied linguistics RAs with a special focus on the syntactic roles of the structures in which the bundles occur.

Literature Review

LBs are understood to be semantically transparent combinations of words that are identified as “simply the most frequently recurring sequences of words” (Biber & Barbieri, 2007, p. 264). Due to their pervasive nature, a frequency threshold has been chosen for the identification of LBs, which has the great advantage of being methodologically straightforward and having face validity (Ellis, 2012). Previous studies normally used the frequency threshold of 10 times per million words (e.g., Ellis & Simpson-Vlach, 2009), 20 times per million words (e.g., Csomay, 2013), 25 times per million words (Chen & Baker, 2010), or 40 times per million words (Biber & Barbieri, 2007). In order to get round the problem of idiosyncrasies from individual writers, the criterion of dispersion is also used, which determines the number of texts in which a linguistic feature occurs (Gries & Ellis, 2015). This is to ensure that the identified bundles are typical of the entire corpus (Pan et al., 2016). Frequency distribution of LBs provides evidence for the description of register variation such that frequent language features that typify a particular register are prioritized (Grabowski, 2015).

An important register for the investigation of variations in LBs is academic writing. LBs are important building blocks of coherent discourse in academic writing because they serve as an effective discriminator of the register which employs distinct sets of LBs that are tailored to its communicative purposes (Wang & Zhang, 2021). Hyland (2008) holds that the investigation of LBs is of particular importance in EAP, as there is mounting evidence that LBs have important functions in academic writing (Staples, Egbert, Biber, & McClair, 2013). Similarly, Cortes (2004) argues that the appropriate use of formulaic expressions is the marker of proficiency in a register, including academic writing.

Recently, there has been a growing number of studies exploring fixed expressions within academic writing by L2 writers, compared with native-English speaking writers (e.g., Adel & Erman, 2012; Pan et al., 2016; Salazar, 2014; Esfandiari & Barbary, 2017). For example, Chen and Baker (2010) investigated LBs in L1 and L2 academic writing. Two corpora of published academic writing and student writing were used to be explored in terms of types and tokens of LBs both qualitatively and quantitatively. The results indicated that published academic texts used the widest range of LBs, whereas L2 Chinese student writing exhibited the smallest range. Another finding of their study was that L2 students overused certain LBs which native-speaker academics rarely used. Similarly, Adel and Erman (2012) compared the use of LBs by L1 speakers of Swedish advanced learners and their English native-speaker counterparts who were

all undergraduate students in the discipline of applied linguistics. Four-word lexical bundles were extracted from the corpora, and they were analyzed both quantitatively and qualitatively in terms of the functions they served. The results of their study showed that native speakers used more varied and a larger number of lexical bundles in comparison to L2 writers. Their findings supported previous native/non-native research traditions focusing on MWSs in general and LBs in particular. Recently, Lu and Deng (2019) explored the use of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. Four-word bundles were extracted from 13,596 and 4,755 abstracts of doctoral dissertations. The identified bundles were categorized according to their functional and structural attributes. The results of his study revealed that Chinese students used lexical bundles in a fundamentally different way with regard to functional and structural features of LBs. They also exhibited incomplete knowledge of LBs, indicating L1 transfer. The other finding of their study was that LBs that were used by Chinese learners did not meet the conventions of academic writing in hard sciences.

While the results from previous studies on LBs produced by native versus non-native language speakers are valuable in revealing the role of nativeness in academic writing proficiency (See Romer & Arbor, 2009), what is less clear is the effect of methodological issues, such as comparability of corpora and frequency/distribution thresholds, on the extracted bundles from the corpora to be compared. In a study on methodological issues in contrastive lexical bundle research, Pan, Reppen, and Biber (2020) revealed that “the difference in the number of words and number of texts across sub-corpora can have a strong effect on claimed differences in bundles across groups even when the corpora are closely matched for their register and topic” (p. 215). Pan et al. (2020) conducted a similar study on the effect of identification threshold on lexical bundle research, and it was found that “different identification thresholds applied to the same pair of corpora may yield conflicting results” (p. 336). Accordingly, it is suggested that researchers base their bundle analysis on structural and functional characteristics, rather than comparing lists of specific bundles (Pan et al., 2016).

In order to arrive at a clearer picture of the pattern of LBs associated with certain groups, and to get round the problem of long lists of produced LBs by native/non-native groups, which were of little pedagogical value, some scholars have categorized LBs through structural and functional taxonomy. Two commonly cited classifications are those of Biber et al. (1999) and Hyland (2008). The former classifies LBs based on their structural attributes, which include verb phrase (VP) bundles, noun phrase (NP) bundles, and prepositional (PP) bundles. The latter, however, takes a functional perspective on LBs, which fall into three categories:

research-oriented bundles, text-oriented bundles, and participant-oriented bundles.

Although structural and functional classifications of LBs act “as alternative formulas [which] emerged as a matter of inquiry in the language teaching field” (Güngör & Uysal, 2016, p.177), identified LBs do not reflect the developmental path to use discourse conventions appropriately (Shin, 2018). The same bundles may occur in different syntactic positions for which structural and functional classifications do not capture the complexity of the language unit within which the LBs occur. For example, the bundle *one of the most* can be used in different syntactic roles such as subject (e.g., *One of the most notable findings of the present research is...*), subject predicative (e.g., *...balance of power as being one of the most crucial elements...*), or direct object (e.g., *The software identified one of the most...*).

In a series of studies, Biber and Gray (2010, 2013, 2016), and Biber et al. (2011) have documented that academic prose is structurally more compact than conversation. This argument ran counter to previous assumptions that academic writing is maximally explicit in meaning. These researchers have shown that a compressed discourse style in academic writing is at odds with explicitness, arguing that traditional clausal measures of syntactic complexity cannot gauge the grammatical complexity of academic texts because of their poor theoretical foundations. In order to characterize the development in academic writing, Biber et al. (2011) hypothesized the developmental sequences of grammatical complexity along two grammatical parameters: grammatical form and syntactic function. Accordingly, three grammatical types were identified: finite dependent clauses, non-finite dependent clauses, and dependent phrases. These grammatical stages progress from finite dependent clauses through intermediate stages of non-finite dependent clauses and finally to the last stages of dependent phrases (Biber et al., 2011). Although the hypothesized stages of writing development did not specifically investigate lexical bundles, they “paved the way for the exploratory use of this approach in the production of other linguistic features such as lexical bundles” (Shin, 2018, pp. 119-120).

Different studies have tried to provide empirical evidence to support the hypothesized stages of writing development proposed by Biber et al. (2011). For instance, Parkinson and Musgrave (2014) explored the syntactic complexity of academic texts produced by MA and undergraduate students. With a special focus on noun phrase modifiers, the authors confirmed the developmental stages of writing complexity in the sense that undergraduate writers relied heavily on premodifiers, which are supposed to be acquired at earlier stages of writing development. On the other hand, noun modifiers employed by MA writers better approximated those of published academic prose. Similarly, Lan and Sun (2019) examined the quality of

student papers across three tiers of first-year L2 students. The results revealed that low-rated papers demonstrated lower complex nominal densities, lower mean length of clauses, and lower mean length of T-units, providing further evidence that development in academic writing moves from clausal embedding to phrasal embedding.

The current study intends to extend the structural analysis of LBs in the existing literature by analyzing the identified bundles within the framework of Biber et al.'s (2011) hypothesized stages of academic writing. To this end, we identified and examined LBs in two corpora of the RAs authored by L1-Persian and L1-English academic writers. Specifically, the study is guided by the following two research questions:

1. What are the patterns of use of lexical bundles in the writing of L1-Persian and L1-English academic writers?
2. How do L1-Persian writers and L1-English writers in applied linguistics use lexical bundles in academic writing in terms of syntactic functions?

Methodology

Corpus Construction

The present study drew on native and nonnative corpora of RAs in applied linguistics from leading journals in the field. We chose applied linguistics based on the following considerations: First, "it is an interdisciplinary field of study which represents a wide landscape of academic territories" (Shirazizadeh & Amirfazlian, 2021, p. 2). Second, the study of LBs in applied linguistics has become an increasingly important area in recent years (Wang & Zhang, 2021). Accordingly, the present study intended to extend the existing literature on the use of LBs in applied linguistics by approaching the issue from a different perspective.

The native corpus (NC) was composed of 103 texts extracted from published RAs in national English-medium journals in Iran. The nonnative corpus (NNC) was comprised of 106 texts from highly prestigious international English-medium journals. Descriptive statistics of the corpora are presented in Table 1.

Table 1. *Description of the Corpora*

Corpora	Number of Texts	Mean Length of Texts (Words)	Total Corpus Size (Words)
NC	103	9929.04	1,022,692
NNC	106	9660.80	1,024,999

The inclusion of the journals in this study was based on the two criteria of publication history and *h* index, which is defined as the number of publications of a certain author (*h*) with a citation number of at least *h* times (Hirsch, 2005). In other words, a researcher who has published 15 research papers, each with at least 15 citations, would have an *h* index of 15. The advantage of the *h* index over the traditional journal impact factor (JIF) is that it is less affected by over-citation because it is not based on mean scores (Harzing & Van der Wal, 2008). Journals with a higher *h*-index (more citations in more articles) represent a model of empirical research articles in the field of applied linguistics and language education because they impact the field through a high number of highly cited articles. Table 2 presents descriptive information of the journals from which the articles have been extracted.

Table 2. *Overview of Journals Included in Native and Nonnative Corpora*

Journal	Years of Publication	H factor
Language Learning	1948-1953, 1955-1956, 1958-ongoing	38
Applied Linguistics	1980-ongoing	38
TESOL Quarterly	1981-ongoing	36
Modern Language Journal	1916-1996, 1998-2001, 2005-ongoing	36
English for Specific Purposes	1980-1981, 1986-ongoing	25
Iranian Journal of Applied Language Studies	2009- ongoing	—
Journal of Teaching Language Skills	2009- ongoing	—
Journal of English Language Teaching and Learning	2010- ongoing	—
Journal of Language and Translation	2010- ongoing	—
Journal of Research in Applied Linguistics	2010- ongoing	—
Issues in Language Teaching	2012-ongoing	—
Applied Research on English Language	2012-ongoing	—
Iranian Journal of Language Teaching Research	2013-ongoing	—
Iranian Journal of English for Academic Purposes	2015-ongoing	—

In order to identify native and nonnative English academic writers, we followed the identification method suggested by Wood (2001), who took into account the names and affiliations of authors. To determine the L1 status of the authors in NNC, we simply deduced the names and affiliations were indicative of Persian writers. As for native English writers in NC, after checking the Anglophone origin of the names, we made sure if the authors were affiliated with any institution in English-L1 speaking countries. Texts authored by multiple authors were excluded from the study if the authors had differing native and nonnative English status.

All research articles followed the IMRD format and were published between 2018 and 2020. The collection of recently published research articles characterizes ‘the present day’ trends in academic writing (Biber & Gray, 2016). Special issues were excluded, as special issues varied both in article type (in having synthesis or review articles) and in communicative functions. Only research studies representing empirical studies were included so that rhetorical and linguistic variations could be controlled for. “Non-empirical and theoretical review articles often have varied rhetorical organization, which may result in writers’ divergence in making linguistic choices” (Ruan, 2018, p. 6). Accordingly, articles were excluded if their functions and organizational structures differed from those of empirical research articles, which included meta-analyses, position papers, forum discussions, and book reviews. All tables, appendices, diagrams, graphs, titles, captions, and footnotes were removed from the papers so as to ensure the reliability of the data.

Identification of Lexical Bundles

In order to identify LBs, the authors needed to decide on the length of word sequences as the first step in the analysis. It was an important decision because different identification thresholds may result in different lists of bundles (Pan et al., 2016). Biber et al. (1999) argued that three-word bundles are extremely common, while “four-word, five-word, and six-word bundles are more phrasal in nature and correspondingly less common” (p. 992). Given that the retrieved bundles in this study have been manually checked through concordance lines for determining the syntactic functions of each bundle, the frequency threshold of three-word bundles would generate a long list of word sequences whose analysis would be very labor-intensive. On the other hand, four-word bundles “are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles” (Hyland, 2008, p. 8). As a result, we investigated four-word bundles in this study. Frequency and dispersion are two main criteria

for the selection of LBs in literature. However, there seems to be little consensus among researchers regarding the determination of the cut-off point. In this study, we followed Cortes (2008) and set the frequency criterion of 20 times per million words across at least five or more texts.

Data Analysis

The bundles were identified using a concordance tool called *AntConc* version 3.5.9 (Anthony, 2020). Discipline-specific bundles (those which are more frequently found in a given discipline e.g., *students of other languages*) and overlapping bundles (those that are part of larger bundles) were excluded so as not to inflate the number of bundles (See Chen & Baker, 2010). Following Biber and Barbieri (2007), we normalized identified bundles to 1,000,000 words. This practice has at least two advantages: first, it allows for the comparability of the results obtained from the current study to those of others (Biber & Barbieri, 2007), and second, it allows for employing parametric tests which could otherwise be wasteful of data (Biber et al., 2011). In order to check for the significance of the differences with regard to the frequency distribution of the LBs between the two corpora, log-likelihood tests were performed. The next step for the researchers was to categorize the retrieved bundles based on Biber et al.'s structural taxonomy, which involved identifying the type of internal structural unit (verb phrase bundles, noun phrase bundles, and prepositional bundles). Drawing on Biber et al.'s (2011) hypothesized stages of writing development, and syntactic classification of phrasal bundles (Cortes, 2015; Shin, 2018), we subsequently analyzed the retrieved bundles in terms of the syntactic roles they played in the sentence. Concordances surrounding the occurrences of LBs were examined qualitatively to determine their discursive and rhetorical functions within a broader context. This allowed us to analyze the construction of LBs produced by Persian writers and compare them to those of native-speaker writers from the perspective of L1 transfer, overuse, or misuse.

Results

The analysis of the lexical bundles revealed that L2 academic writers employed more types and tokens of LBs than L1 academic writers. This suggests that L2 writers relied more heavily on LBs than L1 writers. The final lists of four-word bundles produced by L1 and L2 academic writers are presented in the Appendix. These bundles have been identified after excluding topic-dependent and discipline-dependent bundles. Table 3 presents the number of types and tokens of LBs in the two writer groups.

Table 3. *Number of Types and Tokens of Lexical Bundles in Two Pairs of Corpora*

Writer groups	Types	Tokens
Native-speaker academic writers	54	2004
Nonnative academic writers	103	4079
Total	157	6083

Closer analysis of bundles revealed that 27 bundles were found to have occurred in both corpora. Table 4 shows the bundles with the normalized token frequency of occurrences in NC and NNC. As Table 4 illustrates, nearly 55% of the retrieved LBs are PP-based bundles, 39% are NP-based bundles, and only 6% of shared LBs are VP-based bundles. These bundles were used with different frequencies in the two corpora.

Table 4. *Shared Bundles with Normalized Frequency per 1,000,000 Words*

	Rank (NC)	Token (NC)	Rank (NNC)	Token (NNC)
on the other hand	1	86.93	2	155.8
the extent to which	3	71.59	18	54.32
as well as the	4	61.36	13	57.4
in the context of	5	60.34	7	78.92
at the same time	6	59.32	64	26.65
in the present study	7	59.32	9	72.77
on the basis of	8	59.32	60	29.72
the results of the	9	59.32	1	218.32
in the current study	11	54.2	17	55.35
in the case of	12	53.18	21	52.27
at the time of	15	49.09	72	24.6
on the role of	16	42.95	53	31.77
in the field of	17	41.93	19	53.3
in the form of	20	39.88	41	36.9
with respect to the	23	36.82	61	28.7
as a result of	24	35.79	12	57.4
in addition to the	25	34.77	83	23.57
in terms of the	26	34.77	23	49.2
the students in the	28	32.73	57	30.75
the nature of the	30	31.7	97	21.52
a wide range of	31	29.66	100	20.5
the meaning of the	34	28.64	96	21.52
to be able to	36	27.61	67	26.65
on the one hand	37	26.59	77	24.6
in line with the	39	25.57	6	85.07
on the part of	53	20.45	84	23.57
the participants in the	54	20.45	47	34.85

LBs in each group were classified structurally using Biber et al.'s (1999) taxonomy. Accordingly, three are broad categories of VP-based bundles, NP-based bundles, and PP-based bundles have been distinguished. Table 5 presents the structural distribution of bundle types in both corpora.

Table 5. *Structural Distribution of LBs in NC and NNC*

Structural subcategories		Native-English writers (%)	Persian writers (%)
NP-based bundles	NP with of-phrase fragment	450(0.22)	1016(0.25)
	NP with other post-modifier fragments	117(0.06)	371(0.09)
	Other noun phrase	45(0.02)	164(0.04)
	Total	612 (0.31)	1551(0.38)
PP-based bundles	PP phrase with embedded of-phrase fragment	469(0.23)	780(0.19)
	Other prepositional phrase fragment	501(0.25)	542(0.13)
	Total	970 (0.48)	1322(0.32)
VP-based bundles	Copular be + NP/Adj. phrase	45(0.02)	216(0.05)
	Anticipatory it + VP/Adj. phrase	75(0.04)	162(0.04)
	Passive verb + prepositional phrase fragment	32(0.02)	133(0.03)
	VP + that-clause fragment	27(0.01)	140(0.03)
	Verb/adjective + to-clause fragment	24(0.01)	49(0.01)
	Verb phrase with active verb	23(0.01)	46(0.01)
	Adverbial clause fragment	39(0.02)	74(0.02)
	Pronoun/noun phrase + be + (...)	22(0.01)	10(0)
	Total	287 (0.14)	830(0.20)
	Other expressions	135 (0.07)	376(0.09)
Total	2004	4079	

VP-based bundles comprised the least proportion of identified bundles in both corpora in this study (NC: 14%, NNC: 20%). These bundles were subsequently categorized based on their syntactic roles in relation to a subset of Biber et al.'s (2011) hypothesized stages of writing development. Table 6 presents the syntactic roles of VP bundles as well as the frequency of the occurrence of each type, which are compared between two writer groups by means of a log-likelihood test.

Table 6. *Distribution of Syntactic Roles of VP-based Bundles in NC and NNC*

Stage	Syntactic Roles	NC	NNC
1	Finite complement clause (CC) controlled by common verbs*	20(0.07)	78(0.09)
	Finite CC controlled by wider set of verbs	25(0.09)	62(0.07)
2	Finite adverbial clauses	60(0.21)	185(0.22)
	Nonfinite CC, controlled by common verbs	23(0.08)	135(0.16)
	Finite CC controlled by adjectives	11(0.04)	63(0.08)
3	Nonfinite CC Controlled by wider set of verbs	45(0.16)	96(0.12)
	<i>That</i> relative clauses, especially with animate head nouns	50(0.17)	113(0.14)
	Nonfinite CC controlled by adjectives	15(0.05)	26(0.03)
4	Extraposed CC	3(0.01)	13(0.02)
	Nonfinite relative clauses	17(0.06)	31(0.04)
5	CC controlled by nouns	4(0.01)	11(0.01)
	Other	14(0.05)	17(0.02)
Total		287 (100%)	830 (100%)

Table 6 presents the syntactic functions of VP-based bundles which are compared based on the number of tokens. The findings revealed that *finite adverbial clauses* were the most frequent category of VP-based bundles used in NC. They were followed by *that relative clauses* and *nonfinite complement clauses*. NNC, similarly, showed the heaviest reliance on *finite adverbial clauses* which were followed by *nonfinite complement clauses controlled by common verbs*, and *that relative clauses*. The results of log-likelihood showed that none of the syntactic categories showed a significant difference between the two writer groups.

Persian academic writers demonstrated a greater reliance on NP-phrase bundles than native academic writers. On the whole, NP-phrase bundles comprised 31% of LBs in NC, while for NNC the figure is 38%, a substantially, and statistically significant difference. Table 7

shows the subcategories of the syntactic roles with the results obtained from the log-likelihood test for each role.

Table 7. *Distribution of Syntactic Roles of Noun-phrase bundles in NC and NNC*

Syntactic Role	NC	NNC
Subject**	112(0.18)	482(0.31)
Subject predicative*	97(0.16)	381(0.25)
Direct object*	139(0.23)	202(0.13)
Indirect object	12(0.02)	23(0.01)
Agent in passive voice	6(0.01)	77(0.05)
of-phrase as postmodifier**	195(0.32)	264(0.17)
Relative clause	12(0.02)	35(0.02)
Other	39(0.06)	87(0.06)
Total	612 (100%)	1551 (100%)

Note. **significant at $p < 0.001$. * = Significant at $p < 0.05$

As presented in Table 7, both corpora have a different proportion of NP-based bundles, with NC relying mostly on of-phrase as post-modifiers, and NNC on the subject, which accounted for 32% and 31% of all NP bundles, respectively. In NC, of-phrase as post-modifiers was followed by direct object, subject, subject predicative, indirect object, relative clause, and agent in passive voice. Other bundles accounted for 6% of all NP-based bundles in NC. However, different patterns of results were observed in NNC, where the second most frequent bundles were found to be subject predicative, followed by of-phrase as post-modifiers, direct object, agent in passive voice, indirect object, and relative clause. Other bundles made up 5% of all NP-based bundles. The results obtained from the log-likelihood test revealed that significant differences were found in the frequency of the four syntactic roles of subject, subject predicative, direct object, and of-phrase as postmodifier. NNC made greater use of subject and subject predicative bundles than NC did, while NC relied more heavily on the direct object, and of-phrase as postmodifier than NNC.

PP-based bundles constituted the largest proportion of all bundle types in NC (48%), while for NNC they were the second-largest proportion (32%) after NP-based bundles. As shown in Table 8, LBs as adverbials were a more frequent type of PP-based bundles in NNC. In NC, 23% of PP-based bundles were adverbials, while for NNC the figure is 77%, a substantial and statistically significant difference. Native-speaker writers relied more heavily

on LBs such as post-nominal modifier (65%) than nonnative writers (23%). This suggests that a larger number of PP-based bundles in NC occur in syntactically more complex units (post-nominal modifiers as opposed to adverbials) compared to those of NNC (see Biber et al.'s (2011) hypothesized stages of writing development).

Table 8. *Distribution of Syntactic Roles of PP-based Bundles in NC and in NNC*

Syntactic Role	NC	NNC
Adverbial*	340 (0.35)	1021 (0.77)
Post-nominal modifier*	630 (0.65)	305 (0.23)
Total	970 (100%)	1322 (100%)

Note. * = Significant at $p < 0.05$

Discussion

The purpose of the present study was to compare lexical bundles used by L1 Persian and L1 English academic writers. The results of the study indicated that Persian academic writers made greater use of LBs at a higher frequency than English academic writers. Structural analysis of LBs revealed that PP-based bundles made up the greatest proportion of all bundle types in NNC, followed by NP-based bundles, and VP-based bundles. However, NC showed different patterns of use where PP-based bundles constituted the largest proportion, followed by NP-based bundles, and VP-based bundles. Retrieved bundles were also examined in terms of the syntactic roles of the units in which they occurred. Significant differences were found for the syntactic roles of NP-based and PP-based LBs between the two writer groups. The syntactic roles of VP-based bundles, however, showed no significant differences between the groups.

The finding that VP-based bundles were the least favored bundles in the entire corpus is not surprising given that clausal bundles are more extensively used in the spoken register than academic writing. This finding supports that of Biber et al. (1999), who argued that the majority of the bundles in academic writing are phrasal bundles. Similarly, Hyland (2008) noted that “most bundles in academic writing are parts of noun or prepositional phrases” (p. 9). The writers’ reliance on phrasal bundles reveals that both groups are aware of the way information is densely packed into phrasal groups (see Fang, Schleppegrell, & Cox, 2006; Staples, Egbert, Biber, & Gray, 2016). However, PP-based bundles were the most frequent bundles in NC, while NP-based bundles comprised the largest group of bundles in NNC. This finding supports that of Chen and Baker (2010), who found that expert writers tend to use more NP/PP-based bundles and fewer VP-based bundles.

The fact that Persian L1 writers made greater use of LBs at a higher frequency than L1 English writers is notable, suggesting that the former group drew on their lexicalized knowledge to construct academic research articles to a greater extent than the latter group did. “Although greater use of the target bundles may indicate L2 phraseological development, learners may also develop their competence in RMCs [recurrent multiword combinations] that do not pass the strict corpus-based distributional criteria for bundles” (Chen, 2019, p. 6). The findings of the present study are consistent with those of Ahmadi, Esfandiari, and Zarei (2020), who revealed that Persian writers used significantly more lexical bundles of all types as noun modifiers compared to native writers. In the same vein, Shahmoradi, Jalali, and Ghadiri (2021) have revealed that L1 Persian writers used more LBs in RAs in applied linguistics and information technology than did their native-speaker counterparts. Similarly, Lu and Deng (2019) found that Chinese doctoral students used LBs more frequently than their native-speaker counterparts, although they “exhibited incomplete knowledge of some aspects of the English lexico-grammatical system” (p. 1).

Analysis of shared bundles in our study revealed that they have been used with different frequencies in both corpora. However, four PP-based bundles (i.e. *in the current study*, *in the case of*, *to be able to*, *for example in the*) show a similar pattern of use in NC and NNC. Previous research has suggested that these LBs are among the most common bundles in the academic register, and RAs in particular (e.g., Bychkovska & Lee, 2017; Chen & Barker, 2010; Hyland, 2012; Pan & Liu, 2019). Out of 53 shared bundles, 30 were used more frequently in NC, and 23 were used more frequently in NNC (See the Appendix).

As noted above, certain bundles were overused in NNC, while the LBs which are commonly used in academic writing were either underused or were nonexistent in NC. In addition, a great number of LBs were used differently in terms of syntactic roles or discursive features in NNC compared to those of NC. The following examples show how two groups of writers used *in the process of*. In NC, the bundle was often employed as a subject predicative after copula *be*-verb, or as the post-modification of an NP, whereas in NNC the bundle often occurred in the sentence-initial position functioning as the premodification of an NP.

- (1) All participating youth are **in the process of** learning English. (NC)
- (2) The ‘framing’ power of metaphor constitutes this bias **in the process of conceptualization**. (NC)
- (3) It appears that **in the process of** EFL teacher recruitment and selection there should be a variety of selection stages and methods. (NNC)

Similarly, the bundle *on the other hand*, which was found to have been far more common in NNC than in NC, was not actually used appropriately by Persian L1 writers. Native writers generally use the bundle “to introduce a contrary view of the previous sentence” (Pan & Liu, 2019, p. 153). However, a closer investigation of concordance lines revealed that Persian writers seemed to employ *on the other hand* as a text-linking bundle for joining any types of ideas (especially additive markers) irrespective of any contrasting links between them. A considerable proportion of all the occurrences in NNC were found to be inappropriate. Examples 4 and 5 show the use of this bundle in NNC and NC, respectively.

- (4) Considering native speakers, this paper tries to tentatively develop a PP which contributes to the way of utilizing metadiscourse units in spoken genres. **On the other hand**, the current study aims to apply the PP and its maxims to the analysis of three spoken genres. (NNC)
- (5) Much of the contribution of LP to multiple-documents comprehension is mediated via impacting single-text comprehension. **On the other hand**, a smaller share of the contribution of PK to multiple-texts comprehension is mediated through single-text comprehension and a larger share of it is unmediated. (NC)

An important finding of the current study is that PP-based bundles were employed proportionally less frequently in NNC than in NC. The most frequent bundles in both corpora were the sequences of *preposition + NP + of* (e.g., *in the case of*). Such structures are hallmarks of advanced academic writing because they “are highly productive in sentence framing” (Ruan, 2017, p. 9). L2 writers’ underuse of prepositional phrases in general and overuse of particular common academic structures (such as *in the context of*) suggest that they may be familiar with their functions in academic writing, but they “cling to words or phrases with which they feel comfortable using” (Appel & Wood, 2016, p. 66).

As for syntactic roles of NP-based bundles, Persian L1 writers were found to have used significantly more LBs in *subject* and *subject predicative* positions than English L1 writers. On the other hand, English L1 writers relied more heavily on LBs as *direct object* and *of-phrase as postmodifier* than L1 Persian writers. Persian L1 writers’ greater use of LBs in the subjective position indicates their tendency to overuse sentence-initial bundles. As Grabowski (2015) pointed out, a great number of high-frequency bundles in the sentence-initial position are

typical of non-academic spoken discourse. Similar to the results of the present study, Shin (2018) and Li, Franken, and Wu (2019) have found that nonnative academic writers tend to use LBs in the sentence-initial position. In their study of Chinese postgraduate students' sources of sentence-initial bundles in their thesis writing, Li and her colleagues found that such reasons as interlingual transfer, literal transfer, semantic transfer, and transfer of training accounted for the sources of a major proportion of the LBs used in the subjective position. The following examples demonstrate how the same LB is used in sentence-medial and sentence-initial positions in NC and NNC, respectively.

- (6) The revised principles informed **the design of the** second-year ELA curriculum and enabled us to propose new instructional theories. (NC)
- (7) **The design of the** present study was both quantitative and qualitative; therefore, mixed method is applied. (NNC)

The more frequent use of *of-phrase as postmodifier* in NC compared to NNC indicates that L1 English writers are more attuned to these constructions as important academic writing conventions. The following examples indicate how LBs are used in syntactic units functioning as *of-phrase as postmodifier* in NC (8) and NNC (9).

- (8) For example, the plural marker at **the end of the** verb is redundant because number is expressed by the subject. (NC)
- (9) Both learners and their instructors were asked to provide information **on the content of the** courses, particularly as related to pronunciation. (NNC)

In comparison, English native writers often used NP-based bundles within *of-phrase postmodifiers* functioning as nominal modifiers, while Persian native writers often employed them as adverbials. The former contributes to a compressed discourse style, whereas the latter results in an elaborated discourse style (See Biber & Gray, 2010; Biber et al., 2011; Biber & Gray, 2016). The following examples from NC and NNC show how NP-based bundles are used to function as adverbials.

- (10) By alternating learning and test trials, we were able to examine how cue use and relative strength changed over **the course of learning**. (NC)

- (11) They commented on **the design of the** semi-structured interview, adequacy and usefulness of the questions, and adjustments were made accordingly. (NNC)

According to Biber et al. (2011), *prepositional phrases as adverbials* are acquired at earlier stages of writing development compared to *prepositional phrases as post-nominal modifiers*. The more frequent use of these structures in *postnominal prepositional phrases* in NC suggests that English L1 academic writers used a greater proportion of NP-based bundles in more complex syntactic units than Persian L1 academic writers did. This different pattern of reliance may be due to dissimilar amounts of exposure to these structures. Persian writers may still need more exposure to compressing lexico-grammatical features required for academic research writing.

Similar differences could also be observed in PP-based bundles where English L1 writers used post-nominal modifiers significantly more frequently than Persian L1 writers. As Biber et al. (1999) put it, postmodifying prepositional phrases are the most common type of postmodifier in the written register in general and in academic writing in particular. They further argue that many of the most common frequent LBs in academic writing include *of-phrases prepositional phrases* because they mark abstract/logical/physical relations. Examples 12 and 13 demonstrate how two groups of writers used PP-based bundles functioning as *postnominal prepositional phrases* to show meaning relationships.

- (12) It led participants to form a predictive strategy such that they might have predicted to produce regulars **in the absence of** irregulars in the experimental list. (NC)
- (13) In the literature on teacher candidates' identity, reflection is widely considered as a critical process **in the development of** teacher professional identity. (NNC)

Biber and Gray (2010) asserted that the recurrent use of *post-modifying prepositional phrases*, and *of-phrases* inter alia, indicates the less explicit and more complex nature of academic writing in which a great deal of meaning is embedded in phrasal expressions. Accordingly, we can safely argue that the more frequent use of LBs in PP-based syntactic units adds to the complexity of the texts. This finding is in line with that of Shin (2018), who found that native academic writers used more than four times as many *postnominal prepositional phrases* as nonnative academic writers did.

Phrasal embedding as postmodifiers has been proposed as the most complicated feature in Biber et al.'s (1999) hypothesized stages of writing development. Several studies have documented that advanced academic writing relies heavily on phrasal features, many of which are *postnominal prepositional phrases* as opposed to *post-modifying prepositional phrases* functioning as adverbials (e.g., Parkinson & Musgrave, 2014; Staples et al., 2016; Taguchi, Crawford, & Wetzel, 2013). Postnominal prepositional phrases contribute to the complexity of clauses. Fang et al. (2006) argued that expanded nominal groups (e.g., *postnominal prepositional phrases*) can compress information that could otherwise take different clauses to convey into a single clause. These compressing elements are central features of advanced academic writing, as they facilitate the flow of information and the development of a complex discourse style.

Conclusion

The present study has examined the use of LBs in RAs authored by English L1 and Persian L1 academic writers in applied linguistics, compiled from two corpora of RAs from leading international journals and Persian English-medium journals. Four-word LBs in both corpora were retrieved and their frequency distribution and syntactic roles in the clause were compared between writer groups. The findings revealed that Persian writers made greater use of LBs at a higher frequency than English academic writers.

Identified bundles were subsequently categorized based on Biber et al.'s (1999) taxonomy. It was found that VP-based bundles were the least frequently used structural category in both NC and NNC. PP-based bundles constituted the largest proportion of all bundles in NC, followed by NP-based bundles. NP-based bundles, however, accounted for the most common structure in NC followed by PP-based bundles. The analysis of syntactic roles of LBs in the clause indicated that Persian writers tended to use NP-based bundles in the sentence-initial position, whereas English writers often used the expressions in sentence medial position. As for PP-based bundles, adverbials made up the greatest proportion of all PP-based bundles in NNC, while postnominal prepositional phrases were the largest sub-category in NC.

Given that VP-based bundles constituted the smallest proportion of LBs and that no significant differences were found between L1 Persian and L1 English academic writers in terms of syntactic functions of VP-based bundles, it seems that Persian writers are already familiar with the structural/distributional/functional features of VP-based bundles in the academic register and know how to use them in the same way as expert native English

academic writers do. However, based on Biber et al.'s (1999) hypothesized stages of writing development where progression starts from clausal features to phrasal features, particularly multiple prepositional phrases which are the most advanced level of developmental category, L1 English writers in our study, who predominantly employed LBs as PP-based bundles mostly functioning as *post-modifying prepositional phrases*, appeared to rely on syntactically more complex bundles than did L1 Persian writers.

The findings of the current study have several pedagogical implications. In addition to structural and functional classifications of LBs, syntactically developmental classifications of LBs can also be developed, and LBs generated on the basis of these classifications could be integrated into academic writing courses. The explicit instruction of syntactically complex LBs seems necessary, as an increasing number of studies have shown that advanced lexicogrammatical features in writing, particularly LBs, are not naturally acquired in the same way as complex language features in spoken register (Biber et al., 2011; Cortes, 2004; Staples et al., 2016; Wei & Lei, 2011). Accordingly, L2 writers need to be explicitly aware of the way complex ideas are embedded in compressing language features through the use of LBs. This study has also shown that native academic writers tended to use certain bundles in particular positions in the sentence which differed from those of nonnative academic writers. Therefore, it seems that instruction in LB usage may benefit from corpus-based learning approaches for exploring, comparing, and analyzing the positional distribution of bundles to resolve any discrepancies in the rhetorical conventions of LBs in advanced academic writing (see Li et al., 2019).

Although corpus-based studies provide invaluable insight into patterns of L2 writers' language use and guide researchers in hypothesizing sources of deviations from target norms, corpus data does not explain why language users opt for particular features while writing (Hyland, 2012). Accordingly, future contrastive analyses of LBs could carry out qualitative analysis such as interviews to complement quantitative methods and to elicit L2 writers' "interpretation of their own bundle choices" (Li et al., 2019, p. 3).

Declaration of Interests

The authors of this study declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *Journal of English for Specific Purposes*, 31(2), 81-92.
- Ahmadi, M., Esfandiari, R., & Zarei, A. A. (2020). A corpus-based study of noun phrase complexity in applied linguistics research article abstracts in two contexts of publication. *Iranian Journal of English for Academic Purposes*, 9(1), 76-94.
- Anthony, L. (2020). *Antconc: A freeware corpus analysis toolkit for concordancing and text analysis*. Retrieved from: <http://www.laurenceanthony.net/software.html>
- Atai, M. R., & Tabandeh, F. (2015). Lexical bundles in applied linguistics articles: Exploring writer, sub-discipline, and sub-genre variations. *Journal of ESP across Cultures*, 11, 33-56.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *Journal of English for Specific Purposes*, 26(3), 263-286.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20.
- Biber, D., & Gray, B. (2013). Nominalizing the verb phrase in academic science writing. In B. Aarts, J. Close, Leech, G., & S. Wallis (Eds.), *The English verb phrase: Corpus methodology and current change* (pp. 99-132). Cambridge: Cambridge University Press.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.
- 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., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?. *TESOL Quarterly*, 45(1), 5-35.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- 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.
- Chen, A. C. H. (2019). Assessing phraseological development in word sequences of variable lengths in second language texts using directional association measures. *Language Learning*, 69(2), 440-477.

- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Journal of Language Learning & Technology*, 14(2), 30-49.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?. *Applied linguistics*, 29(1), 72-89.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *Journal of 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.
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes, & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 197-218). John Benjamins.
- Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, 34(3), 369-388.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17-44.
- Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Journal of Corpus Linguistics and Linguistic Theory*, 5, 61-78.
- 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.
- Fang, Z., Schleppegrell, M. J., & Cox, B. E. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, 38(3), 247-273.
- Garner, J., Crossley, S., & Kyle, K. (2019). N-gram measures and L2 writing proficiency. *System*, 80, 176-187.
- Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *Journal of English for Specific Purposes*, 38, 23-33.
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228-255.
- 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.

- Harzing, A. W. K., & Van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Journal of Ethics in Science and Environmental Politics*, 8(1), 61-73.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569-16572.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Jiang, N. A., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433-445.
- Lan, G., & Sun, Y. (2019). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes*, 38, 14-24.
- Li, L., Franken, M., & Wu, S. (2019). Chinese postgraduates' explanation of the sources of sentence initial bundles in their thesis writing. *RELC Journal*, 50(1), 37-52.
- 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.
- Pan, F., & Liu, C. (2019). Comparing L1-L2 differences in lexical bundles in student and expert writing. *Southern African Linguistics and Applied Language Studies*, 37(2), 142-157.
- 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.
- Pan, F., Reppen, R., & Biber, D. (2020). Methodological issues in contrastive lexical bundle research: The influence of corpus design on bundle identification. *International Journal of Corpus Linguistics*, 25(2), 215-229.
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48-59.
- Rahimi Azad, H., & Modarres Khiabani, S. (2018). Lexical bundles in English abstracts of research articles written by Iranian scholars: Examples from Hhumanities. *Iranian Journal of Applied Language Studies*, 10(2), 149-174.
- Romer, U., & Arbor, A. (2009). English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies*, 20(2), 89-100.

- Ruan, Z. (2017). Lexical bundles in Chinese undergraduate academic writing at an English medium university. *RELC Journal*, 48(3), 327-340.
- Ruan, Z. (2018). Structural compression in academic writing: An English-Chinese comparison study of complex noun phrases in research article abstracts. *Journal of English for Academic Purposes*, 36(1), 37-47.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. UK: John Benjamins Publishing Company.
- 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. (2018). The construction of English lexical bundles in context by native and nonnative freshman university students. *English Teaching*, 73(3), 115-139.
- 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, 100946.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Journal of Written Communication*, 33(2), 149-183.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225.
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *Tesol Quarterly*, 47(2), 420-430.
- Wang, M., & Zhang, Y. (2021). 'According to...': The impact of language background and writing expertise on textual priming patterns of multi-word sequences in academic writing. *Journal of English for Specific Purposes*, 61, 47-59.
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, 42(2), 155-166.
- Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 71-83). Cambridge University Press.

The Complete List of Lexical Bundles in NC and NNC with Normalized Frequency per 1,000,000 Words

Rank	NC	Token	Type	NNC	Token	Type
1	on the other hand	86.93	44	the results of the	218.32	64
2	the end of the	72.61	34	on the other hand	155.8	64
3	the extent to which	71.59	30	of the present study	124.02	45
4	as well as the	61.36	39	the findings of the	105.57	50
5	in the context of	60.34	34	significant difference between the	87.12	31
6	at the same time	59.32	35	in line with the	85.07	46
7	in the present study	59.32	22	in the context of	78.92	41
8	on the basis of	59.32	27	at the end of	72.77	39
9	the results of the	59.32	20	in the present study	72.77	39
10	as a function of	54.2	19	the first research question	62.52	39
11	in the current study	54.2	24	as shown in table	60.47	34
12	in the case of	53.18	30	as a result of	57.4	30
13	it is important to	53.18	36	as well as the	57.4	31
14	the ways in which	53.18	23	the results indicated that	57.4	31
15	at the time of	49.09	24	the second research question	57.4	37
16	on the role of	42.95	8	in the process of	56.37	35
17	in the field of	41.93	21	in the current study	55.35	33
18	in relation to the	40.91	28	the extent to which	54.32	28
19	at the beginning of	39.88	20	in the field of	53.3	31
20	in the form of	39.88	29	the participants of the	53.3	29
21	in this study we	38.86	20	in the case of	52.27	23
22	there was a significant	37.84	17	is one of the	50.22	33
23	with respect to the	36.82	19	in terms of the	49.2	24
24	as a result of	35.79	21	with regard to the	49.2	25
25	in addition to the	34.77	25	it was found that	48.17	29
26	in terms of the	34.77	24	the reliability of the	48.17	30
27	it is possible that	34.77	25	the purpose of the	45.1	32
28	the students in the	32.73	10	as one of the	44.07	28
29	the fact that the	31.7	21	in other words the	44.07	30
30	the nature of the	31.7	21	on the development of	44.07	15
31	a wide range of	29.66	20	the present study was	43.05	28
32	one of the most	29.66	23	to the fact that	43.05	30

33	over the course of	29.66	12	descriptive statistics of the	42.02	20
34	the meaning of the	28.64	18	the analysis of the	42.02	23
35	the use of the	27.61	21	it can be claimed	41	8
36	to be able to	27.61	20	in the use of	39.97	8
37	on the one hand	26.59	20	the following research questions	39.97	39
38	the onset of the	26.59	6	the results showed that	39.97	22
39	in line with the	25.57	19	of the three groups	38.95	7
40	in the absence of	25.57	15	development and validation of	36.9	5
41	were more likely to	24.54	12	in the form of	36.9	24
42	a main effect of	23.52	10	the beginning of the	36.9	24
43	as can be seen	23.52	14	the content of the	36.9	19
44	as the dependent variable	23.52	11	be attributed to the	35.87	21
45	can be used to	22.5	14	can be concluded that	35.87	27
46	the results of this	22.5	14	in this study the	34.85	26
47	as a measure of	21.48	15	the participants in the	34.85	20
48	as part of the	21.48	16	theory and practice in	34.85	27
49	at the level of	21.48	13	they were asked to	33.82	21
50	for each of the	21.48	15	test for equality of	32.8	10
51	than those in the	21.48	8	the mean score of	32.8	17
52	the number of words	21.48	9	on the role of	31.77	64
53	on the part of	20.45	11	can be seen in	30.75	18
54	the participants in the	20.45	11	of the control group	30.75	20
55				the results revealed that	30.75	8
56				the students in the	30.75	21
57				used in this study	30.75	17
58				it should be noted	29.72	25
59				on the basis of	29.72	20
60				with respect to the	28.7	17
61				a systematic review of	27.67	17
62				are presented in table	27.67	6
63				at the same time	26.65	16
64				in the control group	26.65	17
65				it can be argued	26.65	11
66				to be able to	26.65	12
67				a large number of	25.62	17

	experimental and control		
68	groups	25.62	20
69	in the course of	25.62	8
70	as indicated in table	24.6	11
71	at the time of	24.6	13
	immediate and delayed		
72	posttests	24.6	19
73	in a similar vein	24.6	5
74	of the fact that	24.6	21
75	on the acquisition of	24.6	16
76	on the one hand	24.6	8
77	the descriptive statistics of	24.6	17
78	this study aimed to	24.6	16
79	was an attempt to	24.6	20
80	for the sake of	23.57	18
81	in a way that	23.57	14
82	in addition to the	23.57	18
83	on the part of	23.57	19
84	a comparative study of	22.55	19
85	as far as the	22.55	14
86	as the most important	22.55	12
87	be due to the	22.55	5
88	in the experimental group	22.55	15
89	investigate the effect of	22.55	6
90	items of the questionnaire	22.55	14
91	on the use of	22.55	10
92	to analyze the data	22.55	14
93	to participate in the	22.55	18
94	the majority of the	21.52	20
95	the meaning of the	21.52	14
96	the nature of the	21.52	10
97	the needs of the	21.52	12
98	a case study of	20.5	7
99	a wide range of	20.5	17
100	one of the main	20.5	13
101	so that they can	20.5	16
102	the impact of the	20.5	14
103	was found to be	20.5	14