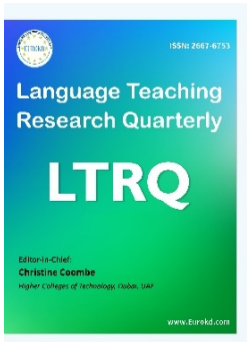


Language Teaching Research Quarterly

2024, Vol. 40, 1–28

Lexical Bundles in the Discussion Sections of Medical Sciences Articles: Frequencies, Syntactic Structures, and Discourse Functions

Roya Goodarzi[1], Javad Gholami[2]*, Zeinab Abdollahpour[3]

[1]English Language Department, Sarab Branch, Islamic Azad University, Sarab, Iran
[2]English Language Department, Faculty of Literature and Humanities, Urmia University, Iran
[3]English Language Department, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran

Received *11 August 2023* ***Accepted*** *12 December 2023*

Abstract

Lexical bundles (LBs) are frequent groups of words that appear repeatedly in different academic texts. A plethora of research has explored their distribution and usage in general, particularly in academic texts. However, to our knowledge, the extent of research investigating LBs in the discussion sections of Medical Research Articles (MRAs) is scant. The present study examined the diversity and density of four-word LBs in the discussions of 1400 MRAs. Four-word bundles totaling 413, including general and subject-bound LBs, were extracted using the freeware *concordance software program* AntConc and categorized based on their syntactic structures and discourse functions. The findings revealed that discussions structurally rely heavily on phrasal LBs (i.e., prepositional phrases and noun phrases) in general and subject-bound LBs compared to clausal bundles, which include VP-based and Clause-based LBs. Regarding functional categories, the general referential bundles with their subcategories were found to have the most considerable proportion in the medical RA genre. Given the importance of LBs in disciplinary writing and academic discourse, the findings could be instrumental in crafting suitable pedagogical materials and activities on general and subject-specific LBs for academic writing in English for Medical Purposes.

Keywords: *Discourse Functions, Lexical Bundles, Discussion Section, Medical Sciences, Research Article, Syntactic Structure*

Introduction

Formulaic structures such as lexical bundles (LBs) have been categorized in relation to academic oral and written outputs based on frequency orientations in a North American context (Pérez-Llantada, 2014). As a type of multi-word expressions, LBs are often employed to examine general aspects applicable to several disciplines (Yin & Li, 2021). As Biber and

* Corresponding author.
E-mail address: j.gholami@urmia.ac.ir
https://doi.org/10.32038/ltrq.2024.40.01

Barbieri (2007) put it, LBs are considered as part and parcels that frequently occur in speech and writing. Besides, many scholars believe that lexical bundles are pivotal in academia (Cortes, 2013; Hyland, 2008a; Simpson-Vlach & Ellis, 2010). According to Hyland (2008b), in English for Academic Purposes (EAP) and English for Specific Purposes (ESP), LBs present noticeable differences at graduate levels, and academic writing programs must incorporate such discipline-specific variety into syllabi and expose students to a wide range of LBs across disciplines (Cunningham, 2017). In corpus linguistics, LBs occupy an important place in realizing discipline-specific academic features, particularly in writing (Biber et al., 1999; Hyland, 2008a).

A substantial body of research has been done on LBs in different research articles (RAs) sections, including applied linguistics abstracts, introductions, total RAs, and results (Abdollahpour & Gholami, 2018; Amnuai & Wannaruk, 2013; Basturkmen, 2009; Brett, 1994; Cortes, 2013; Farnia & Barati, 2017; Hassanzadeh & Tamleh, 2023; Henry & Roseberry, 1997; Kanoksilapatham, 2003; Malmir et al., 2019; Omidian et al., 2018; Samraj, 2002; Tessuto, 2015; Varghaei & Khodadadi, 2022). However, there still seems to be a scarcity of such genre analysis in discussions of medical RAs. A handful of studies targeted RA discussions in terms of rhetorical features (Basturkmen, 2009; Ruiying & Allison, 2003; Sadeghi & Alinasab, 2020). Given this gap, in this study, we identify LBs employed by medical writers in RA discussions. Studies such as ours could help deepen the disciplinary vision of LBs in the academic genre. Indeed, novice writers and academic writing instructors may benefit from disciplinary LB knowledge of RA sections and improve specific language uses in particular disciplines, such as medicine.

In this study, we focus on the discussion section due to its critical role in forming RAs, which holds true in the case of medical texts. In other words, to develop an effective discussion, writers need to restate their findings based on theoretical grounds and competent justification potentials. Research findings are amalgamated with meaning in discussion sections, and writers provide solid evidence of the relevance and contribution of their study to a general field (Le & Harrington, 2015). In addition, a well-developed discussion section provides new insights into field knowledge (Basturkmen, 2009).

Lexical Bundles Importance in Academic Writing

In academic fields, LBs are groups of words that are normally used together to make a whole, known as multi-word structures, for a specific purpose in discourse and pragmatics (Chen & Baker, 2010). It is typically clear to discern the meaning of LBs based on their constituting words (e.g., *it is possible that*) (Biber et al., 1999). In academic articles, LB usage indicates high linguistic capability within a particular discourse community (Biber et al., 2004). Thus, different academic genres encompass a diverse set of LBs specific to particular disciplines (Salazar, 2014).

For disciplinary writers and readers, LBs seem straightforward and foster active involvement in a discourse community. In addition, academic writers can gain fluency in writing utilizing frequent LBs, which fit readership needs and expectations well (Coxhead & Byrd, 2007). On the other hand, lack of LB use may put forward the idea that writers do not hold the expert knowledge of an academic figure or are not adequately fluent in disciplinary writing (Bamberg, 1983; Hyland, 2008a; Li & Schmitt, 2009; Wray, 2002). Consequently, such

shortcomings in academic writing in terms of LB use encourage a negative impression among L2 readers (Jones & Haywood, 2004; Lewis et al., 1997; Li & Schmitt, 2009).

Expert academic writing requires the use of frequent language structures by non-Anglophone writers across academic registers and communities (Ellis & Simpson-Vlach, 2009; Martinez & Schmitt, 2012). According to established evidence, 52.3% of written discourse consists of LBs and formulaic sequences (Erman & Warren, 2000). Fruitful academic communication involves shared knowledge of content and recurrent language structures, such as collocations and chunks (Lewis et al., 1997).

It was observed in an EAP research, frequent LBs exist across disciplines and should be learned and practiced to gain profound disciplinary knowledge (Ellis et al., 2008). In other words, students must master academic lexis to enhance their expert and specific language repertoire and effectively express meaning (Coxhead & Byrd, 2007). As Wingate and Tribble (2012) strongly argue, disciplinary writing features require explicit focus in advanced writing courses for experienced and novice writers in English as Additional Language contexts, which can help improve disciplinary genre knowledge among students (Wingate & Tribble, 2012).

Syntactic Structures and Discourse Functions of LBs

The Longman Grammar of Spoken and Written English has provided categorization for the most commonly used LBs (Cortes, 2002). In this definition, LBs consist of grammar features specific to registers. Given this structural variation, studies focused on LBs of oral and written outputs and reported phrasal and clausal differences in terms of LBs. Phrasal and clausal categories are diverse concerning their sub-components, such as *noun phrases* (e.g., '*the course of the'*) and *prepositional phrases* (e.g., '*in the present study'*) as phrasal LBs and *verb phrases* (e.g., '*is the first study*') and *dependent clause bundles* (e.g., '*when compared to the'*) as clausal LBs (Shin, 2019).

Studies on the lexical bundle structures have revealed that certain bundle structures may be more frequently employed in a certain register or genre. Nonetheless, it is vital to investigate the functional features of lexical bundles since they are functional units that act as the building blocks of various discourses (Oktavianti & Prayogi, 2022). In another classification, LBs are separated and identified depending on their functionality. In this vein, Biber et al. (2004) probed the discoursal function of LBs in academic communities and reported *referential expressions* (e.g., *an important role in*), discourse *organizers* (e.g., *the current study was*), and *stance expressions* (e.g., *more likely to be*) as three main functions with specific sub-components. They referred to referential bundles as contextualized information and interpretations, discourse structures representing idea interrelationships, and stance bundles that targeted writers' subjective judgments about information.

Drawing insights from Biber et al. (2004), Hyland (2008a) adapted the scheme for scholarly writing and suggested three LB functionalities, including *research-oriented bundles*, *text-oriented bundles*, and *participant-oriented* bundles. Using the first function, writers manage to structure real-world issues. Alternatively called discourse markers, text-oriented bundles are associated with textual organization and respective meanings. Finally, participant-oriented bundles (so called stance expressions) deal with writers or readership.

The recent literature on academic writing has adequately documented LBs and their uses and functions across fields and languages. For example, Lake and Cortes (2020) focused on

the differences between English and Spanish RAs in history, which L1 writers crafted. The findings showed the superiority of Spanish RAs in terms of LB manifestations, and both RA types were comparable concerning function and structure. Similarly, Pan et al. (2016) provided a grammatical categorization of LB functions, phrasal or clausal functions dependent on nouns, prepositions, or verbs.

Additionally, another study examined LB realizations in English argumentative essays written by first and second-language writers (Bychkovska & Lee, 2017). Based on their findings, L2 writers dominantly employed stance bundles and discourse organizers (e.g., *on the one hand*), while L1 English writers did not perform significantly in this regard. Pan and Liu (2019) compared native and non-native writers considering LB counts in RAs and theses, and LBs were frequent in theses compared to RAs developed by experienced writers. Moreover, clausal LBs were commonly used in published RAs, while phrasal LBs outshined in MA theses. However, both RAs and theses were similar regarding LB functionality and included a wide variety of text-oriented bundles, though stance bundles were rarely used.

The use of LBs across four major rhetorical sections: Introduction, Method, Results, and Discussion (IMRD) of the public health RAs was investigated by Szczygłowska (2022). According to the results, the Method section was the most formulaic. The sections varied in how they met their specific communicative demands by utilizing the different structural and functional categories of common bundles. Nekrasova-Beker and Becker (2020) evaluated five distinct engineering disciplines and revealed cross-disciplinary variance patterns in bundles' frequency, form, and function. Nasrabady et al. (2020) identified several novel functional categories of LBs employed in published RAs in applied linguistics that were not included in the functional taxonomies. The results of those investigations suggest that lexical bundle variations can occur inside a discipline, signifying its linguistic features, in addition to being prone to differences across fields.

As noted, LBs have been widely studied in the RA genre across fields. However, LB realizations in individual RA sections have remained under-investigated in academic writing research. Therefore, in the present study, we concentrate on RA discussions in medical fields to investigate syntactic structures and discourse functions of LBs. Our study can act as a springboard to disciplinary attempts on LB research, particularly in the case of medical writers. The current study addresses three questions:

RQ$_1$: How frequently are four-word LBs employed in the discussions of medical RAs?
RQ$_2$: What syntactic structures do four-word LBs have in the discussions of medical RAs?
RQ$_3$: What discourse functions do four-word LBs have in the discussions of medical RAs?

Method

Corpus

The present study utilized 1400 MRAs, totaling 1,575,125 words, selected from Sage, Elsevier, Springer, Wiley, and Taylor & Francis databases. The chosen RAs were published between 2015 and 2020. In the selection of this corpus, we ensured that there was a proportionate number of native as well as non-native writers. A brief description of the corpus is given in Table 1. All of the journals are considered accredited in their respective areas. On average, the IFs ranged between 2 and 6 for the journals. We used whole texts to build the corpus, leaving out the tables, figures, and footnotes.

Table 1

Corpus Description of Discussion Sections in Medical RAs

Journal Title	No. of Words	% of Corpus	No. of Discussions	MIFs	Ave. length
Journal of Cerebral Blood Flow & Metabolism	74,718	4.74	47	6.96	1589.7
Journal Inherit Metabolism Disease	87,961	5.58	62	4.75	1418.7
Journal of Parenteral and Enteral Nutrition	116,147	7.37	95	3.89	1222.6
Cancer Genetic Journal	30,904	1.96	28	2.16	1103.7
Nutrition Research Journal	108,157	6.86	98	3.87	1103.6
Radiotherapy and Oncology	70,870	4.49	63	6.28	1124.9
Pediatric obesity	67,133	4.26	59	3.91	1137.8
International Journal of Cardiology	76,400	4.85	66	3.99	1157.5
Cardiovasc Intervent Radiol	325,629	20.67	265	2.79	1228.7
AUTISM	117,987	7.49	90	6.68	1310.9
American Journal of Alzheimer's Disease & Other Dementias	63,057	4.00	154	2.63	409.4
Otolaryngology–Head and Neck Medicine	152,472	9.67	145	2.65	1051.5
Drug ad alcohol review	135,759	8.61	107	4.04	1268.7
Sleep & Breath	147,904	9.38	121	2.94	1222.3
Total	*157,5125*	*99.93*	*1,400*	-	*1089.02*

Note: Ave. length. average text length; MIFs = Median Impact Factors

Data Analysis

Using the freeware *concordancer software program* AntConc, we retrieved four-word LBs in the corpus (Anthony, 2019). The study concentrated on 4-word LBs since they perform a broader spectrum of uses, and many 3- and 5-word LBs contain 4-word bundles (Cortes et al., 2004). Four-word clusters are also easier to categorize and verify in their respective contexts (Chen & Baker, 2010).

Several methods have been developed to determine the total number and average frequency of bundle sequences. Biber et al. (2004) devised a frequency cut-off of forty frequencies per million words (pmw) to extract LBs in instructional textbooks (Biber et al., 2004). For a four-word statement to be regarded as an LB, Adel and Erman (2012) and Chen and Baker (2010) established a cut-off of at least 25 frequencies in pmw (Ädel & Erman, 2012; Chen & Baker, 2010). The present study adopted a cautious approach by setting the threshold at 25 occurrences in the corpora.

Afterwards, we used the syntactic structures and discourse functions taxonomies generated by Biber et al. (2004) to categorize LBs. The researchers provided a thorough categorization using the structural correlates of LBs, and we primarily used the framework in the structural analysis in this study. Numerous studies have adopted this method (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2013). Identifying clausal and phrasal structural units was considered essential for the structural classification.

Using a scheme of classification created by Biber et al. (2004), we categorized the discourse functions of the LBs into three major groups based on their meanings in the texts: stance expressions (such as *it is important to*), discourse organizers (such as *on the other hand*), and referential expressions (such as *one of the most*). To verify the accuracy of the data coding and its categorization, another coder was invited to classify the entire corpus, both functionally and structurally. Both raters coded 15% of the whole data in the corpus to test the interrater

reliability of their work. The findings were then compared. The initial agreement rate of structural and functional classifications was 95.2% and 93.6%, respectively. The researchers attained an almost full agreement with further discussions.

Results

This part outlines and describes the distribution of the four-word LBs across the discussion sections of medical RAs. Table 2 shows the distribution of LBs used in the discussions. We found that a corpus of 1.5 million words included 413 distinct LBs. General bundles with 388 bundle types and 18,329 tokens comprised 1.16 % of the words in the whole corpus. Twenty-five different subject-bound bundles (types) with 1,038 tokens (occurrences) make up about 0.06 % of the whole corpus (see Appendix). To assign the retrieved LBs to general and subject-bound, consultation with eight medical specialists at Urmia University of Medical Sciences was conducted. To calculate the intercoder reliability, the Cohen's kappa was used to evaluate the congruence between researchers and medical specialists. A satisfactory kappa coefficient rate of 0.87 was found between them.

Table 2

Number of Types and Frequency of Lexical Bundles in the Corpus

Corpus	Number of words	Total No. of Tokens (pwm)		Discipline	Number of texts
		General	Technical		
Discussion sections of RAs	1,575,125	18,329(1.16)	1,038(0.06)	Medical Sciences	1400
Total		19,367			

Note: pmw. per million words.

Frequency and Syntactic Structure of Lexical Bundles

The research undertaken by Biber et al. (2004) was used as a guide to categorize bundles structurally in this study. As Table 3 describes, our bundle data revealed four main structural categories, including NP-based, PP-based, VP-based, and Clause-based bundles with different subcategories. VP-based LBs include sequences of words containing a verb constituent (e.g., *plays an important role*). Clause-based bundles comprise clause fragments (e.g., *we found that the*) and initiate with a main clause plus a complementizer (e.g., *to, if*) or a Wh-word that introduces a dependent clause. In contrast to PP-based LBs, which consist of a preposition plus an NP fragment (e.g., *in the present study*), NP-based bundles include nominal phrases with of-phrase fragments (e.g., *and the presence of*) and post-modifier fragments (e.g., *the fact that the*). Distinctive structural features were shown with an asterisk (*) in Table 3.

Table 3

Structural Categories of Lexical Bundles (Adopted from Biber et al., 2004)

Categories	Subcategories	Sample bundles
VP-based	(connector +) 3rd person pronoun + VP fragment	*it is not possible*
	Copula be + noun phrase/adjective phrase*	*was no significant difference*
	Verb phrase (with non-passive verb)	*plays an important role*
	Verb phrase with passive verb	*be explained by the*
	That-clause fragments	*should be noted that*
Clause-based	Wh-clause fragments	*when compared to the*
	(verb/adjective+) to-clause fragment	*appears to be a*
	Pronoun/noun phrase + be (+ . . .) *	*this study is the*
	Adverbial clause fragment *	*as measured by the*
	(noun phrase/pronoun) +V+(complement) *	*we found that the*
NP-based	(connector +) Noun phrase with *of*-phrase fragment	*and the presence of*
	Attributive adjectives as premodifiers*	*the small sample size*
	Noun phrase with post-nominal clause fragment *	*the extent to which*
	Noun phrase with prepositional phrase fragment*	*significant difference in the*
	Prepositional phrase expressions	*in the present study*
PP-based	Comparative expressions/ other expressions	*higher than that of*

Note: NP= Noun Phrase; PP= Prepositional Phrase; VP= Verb Phrase

Table 4 presents the structural categories and tokens of LBs in the corpus. Four-hundred and thirteen LBs that appeared across discussion genres, comprising 19,367 tokens. Table 4 shows that medical academic writers use more general types (93.86%) and significantly greater number of tokens (94.57%) of LBs than subject-bound bundles (i.e., 6.14% of types vs 5.43% of tokens).

Table 4

Number of Types and Frequency of Structural Lexical Bundles in the Corpus

Genre	No. of Types (%)		No. of Tokens (%)	
	General	Discipline-specific	General	Discipline-specific
Medical Sciences RAs Discussions	388(93.86)	25 (6.14)	18,329 (94.57)	1,038 (5.43)
Total		*413*		*19,367*

The most common bundles used by medical sciences RAs writers were *in the present study, it is possible that, in the current study,* and *on the other hand,* which occurred 557, 310, 271, and 214 times, respectively. The top ten frequently used four-word LBs were phrasal (see Table 5).

Table 5

The Top 10 Most Frequent Lexical Bundles

Lexical bundles	Tokens
in the present study	557
it is possible that	310
in the current study	271
on the other hand	214
it is important to	163
of the present study	154
has been shown to	149
the results of this	138
as well as the	135
the results of the	134

Table 6 displays the distribution of the syntactic structures and subcategories of LBs in the discussions. Overall, medical science writers used significantly more PPs with prepositional phrase expressions, which comprise over 24% of the general bundles and over 4% of subject-bound bundles. These bundles made up almost 29.14% (120) of types and 35.15% (6,749) of the tokens, respectively, found in the corpus (see Appendix B for the complete list of structural distribution of LBs in the corpus).

Among PP-based bundles, the subcategory of *Prepositional phrase expressions* is notable. Table 6 demonstrates that medical science writers use a wider variety of types and tokens of this categorization. These bundles serve as a guide within the texts (e.g., *'in the present study', 'in the current study'*), link elements and ideas together (e.g., *'on the other hand'*), and function as discourse frameworks to connect to new material or as interactive tools to illustrate the reader/ writer's commitment. (e.g., *'it is possible that'*) (Hyland, 2008a; Pan & Liu, 2019). Some studies have demonstrated that LBs used in academic writing exhibit disciplinary variances (Biber & Barbieri, 2007; Hyland, 2008a). Therefore, it can be argued that these LBs are crucial in medical sciences and need to be covered in medical sciences writing courses (see Table 6).

VP-based fragments are the second important fragments. Compared to Clause-based and NP-based fragments, they contain various bundle types and tokens. They comprise more than 25 % (105) types and over 24% (4,740) of tokens of the total general bundles. It was found that the writers used more passive structures (8.17%) (e.g., *be explained by the)*. The cornerstones of assertion are built using formulaic passive structures. To imply that the outcomes would be the same regardless of who conducted the research can assist in minimizing the personal influence played by the scientist in interpreting findings (Hyland, 2008a).

Both Clause-based fragments (91 types and 3,893 tokens) and NP-based fragments (97 types and 3,940 tokens) used comparable percentages of structural distributions of types and tokens. Regarding Clause-based fragments, the learners overused Pronoun/noun phrase + be (+ . . .) bundles (e.g., *this study is the*) (1,104 tokens) which is a distinctive structural feature that is not in Biber et al.'s (2004) scheme. Writers also favored an excessive amount of types (60) and tokens (2,513) of the (connector +) Noun phrase with *of*-phrase fragment, which Hyland (2008a) claims that they typically imply indicate obvious connections between components of the propositions. They constituted roughly 15 % of the LBs in the corpus.

Table 6
Distribution of Structural Subcategories of General and Technical LBs in Medical RAs Discussions

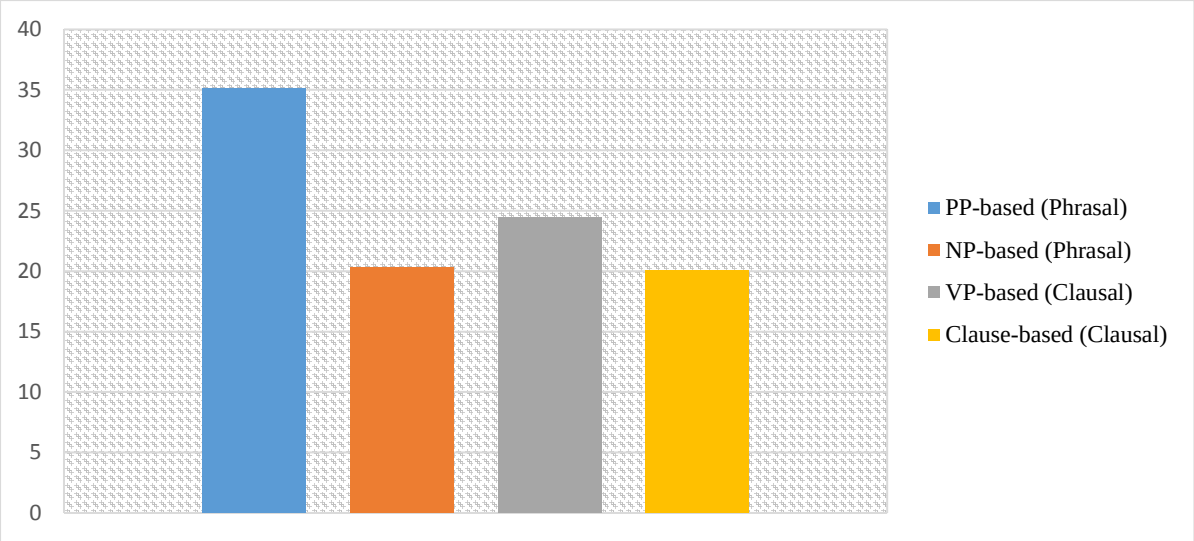
Categories	Subcategories	No. of Types (%)		No. Tokens (%)	
		General	Technical	General	Technical
VP-based-Clausal	(connector +) 3rd person pronoun + VP fragment	20(4.84)	N.O.	1,230(6.35)	N.O.
	Copula be + noun phrase/adjective phrase*	33(7.99)	N.O.	1,431(7.38)	N.O.
	Verb phrase (with non-passive verb)	15(3.63)	N.O.	495(2.55)	N.O.
	Verb phrase with passive verb	37(8.95)	N.O.	1,584(8.17)	N.O.
Total		*105(25.41)*	-	*4,740(24.45)*	-
Clause-based	That-clause fragments	24(5.81)	2(0.48)	1000(5.16)	78(0.4)
	Wh-clause fragments	3(0.72)	N.O.	91(0.46)	N.O.
	(verb/adjective+) to-clause fragment	17(4.11)	N.O.	918(4.74)	N.O.
	Pronoun/noun phrase + be (+ . . .) *	26(6.29)	N.O.	1,104(5.7)	N.O.
	Adverbial clause fragment *	2(0.48)	N.O.	61(0.32)	N.O.
	(noun phrase/pronoun) +V+(complement) *	17(4.11)	N.O.	641(3.3)	N.O.
Total		*89(21.52)*	*2 (0.48)*	*3,815(19.68)*	*78(0.4)*
NP-based - Phrasal	(connector +) Noun phrase with *of*-phrase fragment	60(14.52)	N.O.	2,513(12.97)	N.O.
	Attributive adjectives as premodifiers*	4(0.96)	3(0.72)	197(1.01)	99(0.51)
	Noun phrase with post-nominal clause fragment *	3(0.72)	N.O.	153(0.79)	N.O.
	Noun phrase with prepositional phrase fragment*	26(6.29)	1(0.24)	952(4.91)	26(0.13)
Total		*93(22.49)*	*4(0.96)*	*3,815(19.68)*	*125(0.64)*
PP-based- Phrasal	Prepositional phrase expressions	97(23.48)	19(4.7)	5,702(29.44)	835(4.39)
	Comparative expressions/ other expressions	4(0.96)	N.O.	257(1.32)	N.O.
Total		*101(24.44)*	*19*(4.7)	*5,959(30.76)*	835*(4.39)*
Overall		*388(93.86)*	*25 (6.14)*	*18,329 (94.57)*	*1,038 (5.43)*

Note: Distinctive structural features are shown with an asterisk (*)

Figure 1 presents the percentages of the structural distribution, including NP-based, PP-based, VP-based, and Clause-based bundles. Medical sciences writers use more phrasal bundles (i.e., prepositional and noun phrases) (i.e., 29.14% vs. 23.45% types; 35.15% vs. 20.32% tokens), in comparison to Clausal bundles which include VP-based and Clause-based bundles (i.e., 25.41% vs. 22% types; 24.45% vs. 20.08% tokens).

Figure 1

Distribution of Structural Categories



Distinctive Structural Features

NP-based, VP- based, and Clause-based are distinctive structural subcategories not included in Biber et al.'s (2004) taxonomic framework. Distinctive structural features were shown with an asterisk (*) (see Table 6). As Table 6 illustrates, NP-based comprised *attributive adjectives as premodifiers* (e.g., *the small sample size*) made up 0.96% and 0.72% of general and technical bundle types and 1.01% and 0.51% of tokens, respectively. These descriptive adjectives help create a logical, well-organized, and reader-friendly professional paper with a firm foundation for its claims in the relevant literature (Salazar et al., 2014).

Noun phrases with post-nominal clause fragments (0.72% types and 0.79% tokens) (e.g., *the extent to which),* and *noun phrases with prepositional phrase fragments* (e.g., *significant difference in the*) which constituted 6.29% and 0.24% of general and technical bundle types and 4.91% and 0.13% of tokens, respectively were the other distinctive structural subcategories. Three distinct types of clause-based bundles composing of clause fragments were discovered in this study: *adverbial clause fragments* (2 types, 0.48% and 61 tokens 0.32%) (e.g., *as measured by the*), *pronoun/noun phrase + be (+…)* (26 types, 6.29% and 1,104 tokens 5.7%) (e.g., *this study is the*), and *(noun phrase/pronoun) +V+ (complement)* (17 types, 4.11% and 641 tokens 3.3%) (e.g., *we found that the*). VP- based bundles also included a different pattern, copula be + n phrase/adj phrase (33 types, 7.99% and 1,431 tokens 7.38%) (e.g., *was no significant difference*).

Functional Classification of the Lexical Bundles

The discourse functions taxonomic framework proposed by Biber et al. (2004) was used to evaluate the function played by LBs. The goal was to determine how frequently and for what purposes medical sciences writers employ LBs. Three categories were applied: *stance bundles, discourse organizers*, and *referential bundles*, and we classified each one into specific subcategories (see Table 7).

Table 7

Functional Categories of Lexical Bundles (Adapted from Biber et al., 2004)

Categories	Subcategories	Examples
Stance bundles	Epistemic	*were more likely to*
	Attitudinal/modality stance	*this study was to*
Discourse organizers	Topic introduction/focus	*aim of this study*
	Topic elaboration/ clarification	*as well as the*
Referential bundles	Identification/focus	*is one of the*
	Quantity specification	*small number of patients*
	Framing attributes	*in the absence of, the extent to which*
	Time/place/text-deixis/multi-dimensional reference	*at the time of/ in the setting of/ in this study we/ at the end of*

The third research question was concerned with the examination of the discourse functions and distributions of the highly frequent LBs in the corpus. Using Biber et al.'s (2004) taxonomic framework, two raters independently classified the 413 bundles extracted into several functional groups and reached 92% interrater agreement. Following that, a group discussion helped to settle every remaining issue. Table 8 shows the number of bundle types and token frequencies of each functional category identified in the corpus. It also presents the 413 most frequent four-word LBs found in the corpus along with their different functions, including 185 referential expressions, 105 stance expressions, 76 discourse organizers, and 47 other bundles (see Appendix C).

Referential Expressions

Among all the functional categories, the referential bundles (45%) with their subcategories have the biggest proportion in medical sciences academic writing. Referential bundles include four subcategories in the corpus: *identification/focus, imprecision, specification of attributes (quantity specification, tangible framing attributes, intangible framing attributes), and time/place/text reference.* Two new categories had to be created to categorize some of the bundles in this corpus because the corpora used in earlier LBs research did not contain these bundles. These categories were *contrast and comparison* and *referential subject-bound bundles* (Cortes et al., 2004; Simpson-Vlach & Ellis, 2010) (see Table 8).

Table 8

Distribution of Bundle Types and Tokens of Each Subcategory in Referential Expressions

Subcategory	No. of Types (%)	No. Tokens (%)
Identification/focus	16(3.87)	796(4.11)
Specification of attributes		
quantity specification	45(10.89)	1715(8.85)
tangible framing attributes	16(3.87)	607(3.13)
intangible framing attributes	26(6.29)	1498(7.73)
Time/place/text-deixis bundles		
Time	14(3.38)	1411(7.28)
place	7(1.69)	359(1.85)
deixis bundles	9(2.17)	525(2.71)
Subject-bound bundles*	25(6.14)	1098(5.43)
Contrast/Comparison *	27(6.53)	952(4.91)
Total	*185(44.83)*	*8,961(46.32)*

As Table 8 shows, medical sciences writers use significantly more *specification of attributes* bundles (e.g., *little is known about*). They constitute approximately 21 % of types and 20% of tokens. Among its subcategories, the subcategory of *quantity specification* is noteworthy. As Table 8 presents, medical sciences writers apply more types (10.89%) and tokens (8.85%) of the *quantity specification* subcategory than other subcategories of referential expressions bundles. Next come *Time/place/text-deixis* bundles (e.g., *at the time of/ in the setting of/ in this study we/ at the end of*), which comprise 7.24% of types and 11.84% of all the tokens. Medical sciences writers show similar percentages of both Contrast/Comparison phrasal bundles (6.14% of types and 5.43/% of tokens) and Subject-bound bundles (6.53% of types and 4.91% of tokens).

Stance Expressions

When one proposition is framed by a set of certain attitudes or judgments, it becomes known as a "stance bundle" (Biber et al., 2004). Expressions like “desire”, "intention", and "ability" were used to communicate the speakers' feelings about the issues being discussed (Kashiha & Chan, 2015). As can be seen in Table 9, among stance expressions subcategories, the subcategory of epistemic stance (e.g., *it is possible that*) and other stance bundles (e.g., *has been associated with*) are noteworthy. Regarding stance expressions, as shown in Table 9, the second most common function, *stance expressions*, was discovered to be used predominantly by medical sciences writers. They used more types (55 and 22 respectively) and significantly more tokens (2,591 and 840 respectively) of these subcategories.

Table 9

Distribution of Bundle Types and Tokens of Each Subcategory in Stance Expressions

Subcategory	No. of Types (%)	No. Tokens (%)
Epistemic stance	55(13.31)	2,591(13.37)
Other stance bundles	22(5.32)	840(4.33)
Attitudinal/modality stance		
desire	1(0.24)	27(0.13)
obligation/directive	14(3.38)	800(4.13)
Intention/prediction bundles	8(1.93)	337(1.74)
ability	5(1.21)	198(1.02)
Total	*105(25.39)*	*4,793(24.72)*

Next comes the attitudinal/modality stance, which includes four subcategories (e.g., *desire, obligation/directive, Intention/prediction bundles, and ability*). Obligation/directive expresses the academic writer’s view about the proposition (e.g., *it should be noted*). Attitudinal/modality stance, along with its subcategories (e.g., *desire, obligation/directive, Intention/prediction bundles, and ability*), characterize almost 7% of all LBs in the corpus (0.24%, 3.38%, 1.93%, and 1.21%, respectively).

Discourse Organizers

The purpose of a discourse organizer is to illustrate the connection between the current and pre-discussed subjects (Biber et al., 2004). Discourse organizers involve the following subcategories: topic introduction/focus (e.g., *the current study was*) and topic elaboration/ clarification (e.g., *on the other hand*). As shown in Table 10, medical science writers employ

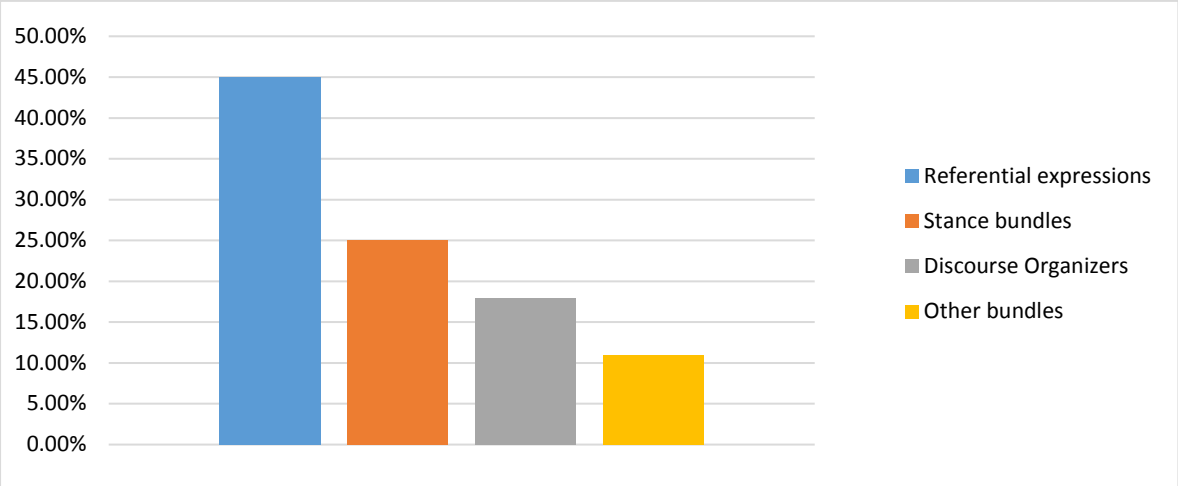
the lowest proportion of discourse organizer bundles (18.39% of types vs 20% of tokens) in comparison to referential and stance expressions.

Table 10
Distribution of Bundle Types and Tokens of Each Subcategory in Discourse Organizers Expressions

Subcategory	No. of Types (%)	No. Tokens (%)
Discourse organizers		
Topic introduction/focus	15(3.63)	751(3.87)
Topic elaboration/ clarification	61(14.76)	3125(16.13)
Other bundles	47(11.39)	1737(8.96)
Total	*123(29.78)*	*5,613(28,96)*

Some lexical bundles detected from the corpus were distinct from the original taxonomy and categorized under *other LBs* (Biber et al., 2004). As shown in Table 10, these bundles had the lowest proportion of total bundles (*11.39*% of types vs 9% of tokens) and exhibited a low level of diversity. As Figure 2 illustrates, discussions of medical RAs heavily rely on referential expressions, accounting for roughly half of the LBs. *Stance expressions, discourse organizers*, and *other bundles*, however, have proportions as high as 25%, 20%, and 11%, respectively.

Figure 2
Distribution of Main Functional Categories



Discussion

This study aimed to categorize and describe how four-word LBs were used in MRA discussions. Utilizing the AntConc concordance program and a frequency-based strategy, we extracted 413 four-word LBs from a 1.5 million-word corpus. General bundles with 388 bundle types and 18,329 tokens and twenty-five different subject-bound bundles (types) with 1,038 tokens made up the whole corpus.

As for the structural distribution of the LBs, the results revealed that the texts produced by medical sciences writers tend to include more NP- and PP-based types than VP- and Clause-based bundles, which contained approximately 56% of the tokens, thus showing their preference for phrasal over clausal LBs (i.e., verb phrases and dependent clause). This result

aligns with earlier research findings (Hassanzadeh & Tamleh, 2023; Varghaei & Khodadadi, 2022) who found that phrasal LBs are the primary structural patterns in MRA abstracts of foreign and Iranian journals as well as in native English authors' discussions. Additionally, they are the most significant grammatical feature of sophisticated academic writing (Biber et al., 2013). A large body of research shows a strong relationship between phrasal nominal modifiers and L2 writing quality (e.g., Kyle & Crossley, 2018; Taguchi et al., 2013) or L2 writing proficiency (Kim, 2020; Lan et al., 2019).

Congruent with Biber et al. (2011), with increasing expertise, academic writers switch from clausal to phrasal styles. This result supports earlier research that claimed academic writing is more phrasal than clausal (Abdollahpour & Gholami, 2019; Biber et al., 2004; Chen & Baker, 2010; Cortes et al., 2004; Hassanzadeh & Tamleh, 2023; Hyland, 2008a; Pan & Liu, 2019; Pan et al., 2016; Salazar et al., 2014; Yin & Li, 2021).

Our results appear to corroborate the findings reported by Biber et al. (2004) and Pan et al. (2016). They found that clausal types are more prevalent in spoken registers of English academic prose, whereas phrasal bundles predominate in written academic prose. Academic writers' transition from a clausal to a phrasal writing style confirms their writing proficiency development (Biber et al., 2011). Our results confirm Cortes's (2002) findings that, in contrast to some genres, such as interpersonal speaking and class instruction, scholarly writing relies primarily on phrasal rather than clausal bundles. It is argued that compressed phrasal bundles are preferable compared to clausal bundles since they are more cost-effective, enable faster, more effective reading, and are understandable to professional readers (Staples et al., 2016). Phrasal-level syntactic complexity factors have been shown to be reliable indicators of L2 academic writing quality. These phrasal patterns are significant because, although they are relatively uncommon in most other registers, they are ubiquitous in written academic discourse (McNamara et al., 2010).

Pan and Liu (2019) demonstrated that LB usage in expert authors' articles in the field of applied linguistics was less common than in MA theses written by both native and non-native writers. In contrast to the present study's findings, they also reported that published articles included more clausal and fewer phrasal bundles than MA theses. The prevalent utilization of clausal bundles in medical research papers implies that there may be intra-sectional variations in the structural application of lexical bundles in addition to register variations (Liu & Pan, 2023). Phrasal bundles appear to be used by medical writers to convey information and clarify concepts and claims. According to Siyanova et al. (2011), phrasal frequency influences how simple language is to understand and is crucial for language use and processing models. We think the same conclusion can be drawn from our results.

In reference to the third research question, functional analysis of LBs suggests that medical writers utilize more referential LBs to identify new information (Biber et al., 2009). This is consistent with the results of earlier studies on LBs (Biber et al., 2004; Cortes, 2013; Pan et al., 2016). Referential bundles serve an ideational purpose by assisting writers in organizing their experiences and determining their points of view (Cortes, 2013; Shin, 2019). Aligning with previous studies (Ädel & Erman, 2012; Appel, 2022; Biber et al., 2004; Chen & Baker, 2010; Li et al., 2023; Liu & Pan, 2023), the majority of the four-word LBs that frequently appear in academic writing are referential bundles since a greater emphasis is placed on communicating only factual information in academic writing (Conrad & Biber, 2005). According to Hyland

(2008b), there are variations between hard and soft science in terms of the bundles they employ. Hard science prioritizes "the empirical over the interpretive" (p.15), while soft science uses more referential and stance bundles. In hard sciences, referential bundles are primarily focused on the physical world, physical location, and quantification. On the other hand, in soft sciences, they are more concerned with abstract constructs and location in history or a process (Durrant, 2017).

The second most common function, *stance expressions*, was discovered to be used predominantly by medical sciences writers. One factor in the excessive use of stance expressions in scholarly writing is writers' propensity to indicate their devotion to or distance from other viewpoints (Lancaster, 2011). Hyland (2008a) argues that research papers focus on providing new knowledge and generating peer acknowledgment, necessitating additional stance bundles.

Our results corroborate those of Yang and Fang (2021), who examined essays written by EFL students in China and demonstrated that, in terms of type and frequency, research-oriented bundles (referential expressions) are the most commonly used bundles, followed by participant-oriented (or stance expressions) and text-oriented bundles (discourse organizers). Furthermore, stance expressions are the most common in pharmaceutical science RAs (Ren, 2021). Upon serious inspection, most bundles in academic writings are used to describe the writer's position or assert certainties regarding other remarks (Appel, 2022). Finally, the results of the present study pinpointed the fact that medical sciences writers employed the lowest proportion of discourse organizing bundles (18.39% of types vs 20% of tokens) compared to referential and stance expressions. Discourse organizers are less common in academic written discourse, which is consistent with the findings of some other studies (Biber & Barbieri, 2007; Chen & Baker, 2010; Oktavianti & Prayogi, 2022).

Conclusion

The current study probed into into the frequency, syntactic structures, and discourse functions of four-word LBs in discussing medical sciences RAs. To do this, 1400 RAs in medical sciences were collected. AntConc software (Anthony, 2019) was used to find the most frequent 4-word LBs in the corpus. Discussion section of RAs, ranking next to introductions in difficulty, is challenging for academic writers since it entails interpreting the results section in light of previous studies (Ferguson et al., 2011; Lim, 2010). Therefore, it is stated that ESP and EAP instructors have to provide their learners with information on their discipline-specific LBs in order to help them examine the corpora in the relevant fields (Cortes, 2013). In this respect, making L2 learners aware of the significance of certain formulaic sequences in creating strong academic prose seems to encourage their propensity to employ them (Hyland, 2008a).

Expert (native or not) academic writers are more likely to be connected with formulaicity in academic writing due to their formal education and intensive academic reading and writing rehearsals (Knight et al., 2018). Thus, formulaicity may not be an innate competence in scholarly writing. Consequently, gaining more knowledge on how formulaicity develops in academic discourse is crucial from the vantage point of native novice authors (Pérez-Llantada, 2014). Each register uses a unique collection of bundles that are connected to the typical communication goals of that register and they show affiliation with a certain discourse group (Ädel & Erman, 2012; Biber & Barbieri, 2007). In other words, learning a new language or

register necessitates being aware of the fact that skilled users prefer certain word sequences over others. EAP course designers need to understand that bundles appear and behave differently in various disciplinary contexts, with the student's unique target context serving as the best place to begin instruction (Hyland, 2008a). Thus, explicit instruction and pedagogical practice of bundles are required for novice writers to understand these linguistic elements. The findings of this study provide pedagogical conceptualization into scholarly writing instruction in EMP courses. The research findings could create instructive materials on general and subject-specific LBs for scholarly writing in EMP. The retrieved LBs can also be utilized as both learning and instructional tools for novice researchers as well as graduate students in academic writing courses.

ORCID

https://orcid.org/0000-0002-9016-8100

http://orcid.org/0000-0002-4970-1110

https://orcid.org/0000-0003-0648-2648

Acknowledgements

We extend our gratitude to Urmia University and Urmia University of Medical Sciences for grating us the permission to collect the corpus from their library sources and subscriptions to journal databases.

Funding

This study received no funding.

Ethics Declarations

None to declare.

Competing Interests

There are no conflicting interests to declare.

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.

Abdollahpour, Z., & Gholami, J. (2019). Embodiment of rhetorical moves in lexical bundles in abstracts of the medical sciences. *Southern African Linguistics and Applied Language Studies*, *37*(4), 339-360. https://doi.org/10.2989/16073614.2019.1692681

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and nonnative speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Amnuai, W., & Wannaruk, A. (2013). Investigating move structure of English Applied Linguistics research article discussions published in international and Thai journals. *English Language Teaching*, *6*(2), 1-13. https://doi.org/10.5539/elt.v6n2p1

Anthony L. (2019). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. http://www.laurenceanthony.net/software

Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System, 110*, 102899. https://doi.org/10.1016/j.system.2022.102899

Bamberg, B. (1983). What makes a text coherent? *College Composition and Communication, 34*(4), 417-429. https://doi.org/10.2307/357898

Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in language teaching. *Journal of English for Academic Purposes, 8*(4), 241-251. https://doi.org/10.1016/j.jeap.2009.07.001

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

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. https://doi.org/10.5054/tq.2011.244483

Biber, D., Gray, B., & Poonpon, K. (2013). Pay attention to the phrasal structures: Going beyond T-units—A response to WeiWei Yang. *TESOL Quarterly, 47*(1), 192-201. https://doi.org/10.1002/tesq.84

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

Biber, J., Hernando, N., Forster, I., & Murer, H. (2009). Regulation of phosphate transport in proximal tubules. *Pflügers Archiv-European Journal of Physiology, 458*, 39-52. https://doi.org/10.1007/s00424-008-0580-8

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes, 13*(1), 47-59. https://doi.org/10.1016/0889-4906(94)90024-8

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. https://doi.org/10.1016/j.jeap.2017.10.008

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

Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, *20*, 56-71. https://doi.org/10.1515/9783484604674.56

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 131-145). John Benjamins Publishing Company.

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43. https://doi.org/10.1016/j.jeap.2012.11.002

Cortes, J., Martinez, S., Karatas, T., & Bullo, F. (2004). Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation, 20*(2), 243-255. https://doi.org/10.1109/tra.2004.824698

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147. https://doi.org/10.1016/j.jslw.2007.07.002

Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes, 25*, 71–83. https://doi.org/10.1016/j.jeap.2016.11.005

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165-193. https://doi.org/10.1093/applin/amv011

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory, 5*(1), 61-78. https://doi.org/10.1515/CLLT.2009.003

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk, 20*(1), 29-62. https://doi.org/10.1515/text.1.2000.20.1.29

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

Ferguson, G., Pérez-Llantada, C., & Plo, R. (2011). English as an international language of scientific publication: A study of attitudes. *World Englishes, 30*(1), 41-59. https://doi.org/10.1111/j.1467-971x.2010.01656.x

Hassanzadeh, M., & Tamleh, H. (2023). The use of lexical bundles by native English authors in applied linguistics: A corpus-driven study. *Language Related Research, 13*(6), 541-569. https://doi.org/10.2989/16073614.2022.2043169

Henry, A., & Roseberry, R. L. (1997). An investigation of the functions, strategies, and linguistic features of the introductions and conclusions of essays. *System, 25*(4), 479-495. https://doi.org/10.1016/s0346-251x(97)00047-x

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (ed.). *Formulaic Sequences* (pp. 269- 292). John Benjamins.

Kanoksilapatham, B. (2003). *A corpus-based investigation of scientific research articles: Linking move analysis with multidimensional analysis*. Georgetown University.

Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and nonnative speakers' use of formulaic language. *Australian Journal of Linguistics, 35*(4), 297-310. https://doi.org/10.1080/07268602.2015.1067132

Kim, H. (2020). Nominal modifiers in argumentative essays as discriminators for writing course placement decisions. *English Teaching, 75*(3), 3-24. https://doi.org/10.15858/engtea.75.3.202009.3

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (2018). Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education, 28*, 1-28. https://doi.org/10.1007/s40593-016-0121-0

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal, 102*(2), 333–349. https://doi.org/10.1111/modl.1246

Lake, W. M., & Cortes, V. (2020). Lexical bundles as reflections of disciplinary norms in Spanish and English literary criticism, history, and psychology research. In U. Romer, V. Cortes & E. Friginal (Eds.), *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise (*pp. 183-204). John Benjamins. https://doi.org/10.1075/scl.95.08lak

Lan, G., Liu, Q., & Staples, S. (2019). Grammatical complexity:'What does it mean'and 'so what'for L2 writing classrooms? *Journal of Second Language Writing, 46*, 100673. https://doi.org/10.1016/j.jslw.2019.100673

Lancaster, Z. (2011). Interpersonal stance in L1 and L2 students' argumentative writing in economics: Implications for faculty development in WAC/WID programs. *Across the Disciplines, 8*(4), 1-23. https://doi.org/10.37514/atd-j.2011.8.4.22

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. https://doi.org/10.1016/j.esp.2015.03.003

Lewis, M., Gough, C., Martínez, R., Powell, M., Marks, J., Woolard, G. C., & Ribisch, K. H. (1997). *Implementing the lexical approach: Putting theory into practice*. Language Teaching Publications Hove.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85-102. https://doi.org/10.1016/j.jslw.2009.02.001

Li, M., Zhang, X., & Reynolds, B. L. (2023). Exploring lexical bundles in low proficiency level L2 learners' English writing: an ETS corpus study. *Applied Linguistics Review, 14*(4), 847-873. https://doi.org/10.1515/applirev-2020-0129

Lim, J. M.-H. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for Academic Purposes, 9*(4), 280-294. https://doi.org/10.1016/j.jeap.2010.10.001

Liu, C., & Pan, F. (2023). Connecting lexical bundles and moves in medical research articles' Methods section. *Southern African Linguistics and Applied Language Studies*, 1-17. https://doi.org/10.2989/16073614.2023.2226171

Malmir, B., Khany, R., & Aliakbari, M. (2019). Journal article highlights in Applied linguistics: An exploration into the rhetorical moves and their lexico-grammatical features. *Iranian Journal of English for Academic Purposes, 8*(4), 49-63. https://doi.org/10.33252/sih.2019.12.63.167

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320. https://doi.org/10.1093/applin/ams010

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*(1), 57-86. https://doi.org/10.1177/0741088309351547

Nasrabady, P., Elahi Shirvan, M., & Ehsan Golparvar, S. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages, 6*(3), 175-197. https://doi.org/10.1080/21698252.2020.1797992

Nekrasova-Beker, T., & Becker, A. (2020). The use of lexical patterns in engineering**:** A corpus-based investigation of five sub-disciplines. In U. Romer, V. Cortes & E. Friginal (Eds.), *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise* (pp. 137-168). John Benjamins. https://doi.org/10.1075/scl.95.10nek

Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education, 9*(2), 761-783. https://doi.org/10.24815/siele.v9i2.23995

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes, 36*, 1-14. https://doi.org/10.1016/j.jeap.2018.08.002

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. https://doi.org/10.2989/16073614.2019.1625276

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. https://doi.org/10.1016/j.jeap.2015.11.003

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84-94. https://doi.org/10.1016/j.jeap.2014.01.002

Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes, 50*, 100968. https://doi.org/10.1016/j.jeap.2021.100968

Ruiying, Y., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*(4), 365-385. https://doi.org/10.1016/s0889-4906(02)00026-1

Sadeghi, K., & Alinasab, M. (2020). Academic conflict in applied linguistics research article discussions: The case of native and non-native writers. *English for Specific Purposes*, *59*, 17-28. https://doi.org/10.1016/j.esp.2020.03.001

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.65

Salazar, L. F., Vivolo-Kantor, A., Hardin, J., & Berkowitz, A. (2014). A web-based sexual violence bystander intervention for male college students: Randomized controlled trial. *Journal of Medical Internet Research, 16*(9), e3426. https://doi.org/10.2196/jmir.3426

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes, 21*(1), 1-17. https://doi.org/10.1016/s0889-4906(00)00023-5

Shin, Y. K. (2019). Do native writers always have a head start over non-native writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes, 40*, 1-14. https://doi.org/10.1016/j.jeap.2019.04.004

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487-512. https://doi.org/10.1093/applin/amp058

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 37(3), 776–784. https://doi.org/10.1037/a0022531

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. *Written Communication*, *33*(2), 149-183. https://doi.org/10.1177/0741088316631527

Szczygłowska, T. (2022). Lexical verbs of epistemic modality in academic written English: disciplinary variation. *Linguistica Silesiana*, *43*, 91-111. https://doi.org/10.24425/linsi.2022.141219

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. Portico. https://doi.org/10.1002/tesq.91

Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes*, *37*, 13-26. https://doi.org/10.1016/j.esp.2014.06.002

Varghaei, E., & Khodadadi, G. (2022). Comparing Lexical Bundles in Medical Research Article Abstracts of Iranian and Foreign Journals. *GEMA Online Journal of Language Studies*, *22*(3), 86-102. http://doi.org/10.17576/gema-2022-2203-05

Wingate, U., & Tribble, C. (2012). The best of both worlds? Towards an English for Academic Purposes/Academic Literacies writing pedagogy. *Studies in Higher Education*, *37*(4), 481-495. https://doi.org/10.1080/03075079.2010.525630

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Yang, W., & Fang, Q. (2021). The Structures and Functions of Lexical Bundles in Argumentative Essays by Chinese EFL Students at the Tertiary Level. *International Journal of TESOL Studies, 3*(3). https://doi.org/10.46451/ijts.2021.09.03

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), 100006. https://doi.org/10.1016/j.acorp.2021.100006

Appendix A

Distribution of LBS in Discussion Sections of Medical Sciences RAs

General LBs (388 types, 19,367 tokens)			Subject-bound LBs (25 types, 1098 tokens)
in the present study 557	the course of the 46	patients in our study 31	of children with autism 111
it is possible that 310	is also possible that 45	are consistent with previous 31	of children with ASD 100
in the current study 271	in line with the 45	and the risk of 31	in patients with AD 73
on the other hand 214	these results suggest that 45	in relation to the 31	in patients with severe 48
it is important to 163	no significant difference in 45	in this study is 31	in children with autism 47
of the present study 154	a risk factor for 45	study has several limitations 31	in patients with a 46
has been shown to 149	there was a significant 44	the validity of the 31	of the patients with 44
the results of this 138	has been reported in 44	limitation of our study 30	that children with ASD 39
as well as the 135	can be used to 44	have contributed to the30	in patients with OSA 38
the results of the 134	parents of children with 44	when compared to the 30	that children with autism 39
as a result of 133	has been associated with 43	is consistent with previous 30	mini mental state examination 37
results of this study 125	the magnitude of the 43	similar to that of 30	BMI body mass index 36
in this study we 125	to the best of 43	activities of daily 1650living 30	in critically ill patients 36
this is the first 124	be attributed to the 43	may be attributed to 30	in children with ASD 34
were more likely to 116	of interest with respect 42	may not have been 30	in the patients with 33
at the time of 108	in a study of 42	be interpreted with caution 30	for children with ASD 32
may be due to 105	findings are consistent with 42	could be related to 30	for children with autism 30
of the current study 103	in the setting of 42	in the study of 30	group of patients with 29
in the united states 101	one of the most 42	are known to be 30	of children with ASDS 28
to be associated with 96	it is not clear 41	at the level of 30	in patients treated with 28
studies have shown that 96	we believe that the 41	in the treatment group 29	of the patients in 28
on the basis of 96	has been shown that 41	it is possible to 29	in patients with OSAS 27
in the absence of 95	appears to be a 41	may be that the 29	sister with an ASD 26
it is likely that 94	reported in the literature 41	the authors declared no 29	whey peptide- based diet 26
with respect to the 94	the current study was 41	may contribute to the 29	of the upper airway 26
in this study the 90	the nature of the 41	did not find a 29	in a patient with 26
in the case of 89	a number of limitations 41	is known to be 29	
is one of the 88	the vast majority of 41	was significantly associated with 29	
in our study the 84	for the first time 41	there is a need 29	
been shown to be 84	were no significant differences 40	should be considered in 29	
may be related to 82	of our study is 40	support for the research 29	
in the context of 81	with the results of 40	been found to be 29	
the fact that the 80	that the use of 40	were found to have 29	
it should be noted 79	in our study were 40	findings from this study 29	

the present study we [77]	mothers of children with [40]	does not appear to [29]
are more likely to[76]	despite the fact that [39]	a better understanding of [29]
research is needed to [76]	patients in this study [39]	in the management of [29]
is possible that the [75]	we found that the [39]	sensitivity and specificity of [29]
we did not find [75]	are in agreement with [39]	with the control group [29]
have been shown to [75]	the best of our [39]	it is unlikely that [28]
an important role in [73]	and the use of [39]	less likely to be [28]
of this study was [72]	the extent to which [39]	more likely to report [28]
be due to the [72]	the presence of a [38]	this may be due [28]
studies are needed to[71]	best of our knowledge [38]	it is well known [28]
is the first study [71]	is in line with [38]	it is necessary to [28]
be related to the [71]	our results suggest that [38]	our study is that [28]
in addition to the [71]	has been reported that [38]	limitations to this study [28]
is important to note [70]	it has been reported [38]	our study is the [28]
publication of this article [70]	this study did not [38]	has been found to [28]
in this study were [69]	have been associated with [38]	as a consequence of [28]
should be noted that [68]	is the first to [38]	important role in the [28]
be explained by the [68]	a result of the [37]	plays an important role [28]
a significant increase in [68]	compared with the control [37]	are similar to those [28]
the use of a [68]	results are consistent with [37]	in this study had [28]
to the fact that [67]	are in line with[37]	analysis and interpretation of [28]
may be associated with [67]	in the study by [37]	brother or sister with [28]
has been reported to [67]	and the presence of [37]	the retrospective nature of [28]
at the end of [66]	to the development of [37]	the time of the [28]
it may be that [65]	with a history of [37]	in the field of [28]
is consistent with the [65]	was no significant difference [37]	was significantly higher in [28]
authorship and or publication [65]	a large number of [36]	the lack of a [27]
important to note that [64]	the majority of patients [36]	the sample size was [27]
in the presence of [63]	increase the risk of [36]	at high risk for [27]
to our knowledge this [63]	which is consistent with [36]	higher than that of [27]
it is difficult to [62]	results suggest that the [36]	the role of the [27]
more likely to be [61]	are likely to be [36]	are less likely to [27]
findings of this study [61]	further studies are needed [36]	seems to be a [27]
at the same time [61]	this study was to [36]	is considered to be [27]
the present study the [60]	we were able to[36]	included in the study [27]
in the number of [60]	could be due to [36]	would like to thank [27]
in the control group [60]	as measured by the [36]	need to be considered [27]
with the use of [60]	play an important role [36]	we were not able [27]
is likely to be [59]	the level of the [35]	this study suggest that [27]
was found to be [59]	the findings of the [35]	have shown that the [27]
there was no significant [59]	it is not possible 35	previous studies have shown [27]
more likely to have [58]	have been reported in [35]	the effectiveness of the [27]

there were no significant [58]	is associated with a [35]	in a group of [27]
between the two groups [58]	included in this study [35]	of the general population [27]
in our study we [58]	the current study is [35]	of the relationship between [27]
of this study is [58]	in agreement with the [35]	the absence of a [26]
the small sample size [57]	in terms of the [35]	due to the small [26]
we were unable to [57]	a significant difference in [35]	may play a role [26]
is in agreement with [56]	are consistent with the [34]	it is known that [26]
an increased risk of [56]	our findings suggest that [34]	can be explained by [26]
results of the present [55]	is not possible to [34]	have been described in [26]
play a role in [55]	it was not possible [34]	be associated with the [26]
for the development of [55]	further research is needed [34]	it is interesting to [26]
these findings suggest that [54]	were not able to [34]	the purpose of this [26]
it has been suggested [54]	siblings of children with [34]	the current study the [26]
in accordance with the [54]	a role in the [34]	of the study was [26]
the results of our [53]	those who did not [34]	our data suggest that [26]
were found to be [52]	significant difference in the [34]	we also found that [26]
has been suggested that [52]	a small number of [34]	be one of the [26]
the findings of this [52]	there are a number [34]	a high level of [26]
in contrast to the [52]	the majority of the [34]	there are several limitations [26]
limitations of this study [52]	are a number of [34]	there was no difference [26]
small number of patients [52]	a wide range of [34]	limitation of the study [26]
it has been shown [51]	the effect of the [33]	limitations of our study [26]
in the treatment of [51]	be more likely to [33]	in this group of [26]
by the fact that [50]	due to the fact [33]	has the potential to [26]
with the exception of [50]	is in accordance with [33]	quality of life and [26]
been reported to be [50]	in our study was [33]	the development of a [26]
in the general population [50]	as part of the [32]	of the children with [26]
for the treatment of [49]	results from this study [32]	it is reasonable to [25]
than that of the [49]	our results show that [32]	authors declared no potential [25]
this is consistent with [49]	was observed in the [32]	is well known that [25]
an increase in the [49]	was not possible to [32]	may be explained by [25]
the use of the [49]	may have contributed to [32]	studies have reported that [25]
this study is the [49]	it is noteworthy that [32]	values are expressed as [25]
with regard to the [48]	this study is that [32]	did not have a [25]
used in this study [48]	over the course of [32]	be associated with a [25]
was associated with a [47]	in the long term [32]	of our study was [25]
it is also possible [47]	limitations of the study [32]	research has shown that [25]
the small number of [47]	the size of the [32]	we have shown that [25]
the end of the [47]	in a previous study [32]	although we did not [25]
in the development of [47]	of this study are [32]	study we found that [25]
long term follow up [46]	to the use of [32]	the results from this [25]

was not associated with [46]	in the majority of [32]	to the presence of [25]
in this study was [46]	reduce the risk of [32]	the difference between the [25]
that there is a [46]	no significant differences in [32]	is similar to the [25]
the present study is [46]	study has some limitations [32]	these findings are consistent [25]
results of our study [46]	that the majority of [32]	significant difference between the [25]
as well as in [46]	with the findings of [31]	similar to those of [25]
limitation of this study [46]	have been found to [31]	when compared with the [25]
in the form of [46]	results of the current [31]	in agreement with previous [25]
study is the first [46]	we did not observe [31]	relatively small number of [25]
the first study to [46]	could be explained by [31]	was not statistically significant [25]
the presence of the [25]	a decrease in the [25]	a group of patients [25]
		as a result the [25]

The Superscript Numbers Indicate the Frequency of Each Bundle.

Appendix B

Complete List of Structural Distribution of LBs in the Corpus

Categories	Subcategories
VP-based	*(connector +) 3rd person pronoun + VP fragment* General LBs: it is possible that [310] , it is important to [163], it is likely that [94], it should be noted [79], it is difficult to [62], it has been suggested [54], it has been shown [51], it is also possible [47], it is not clear [41], it has been reported [38], it is not possible [35], it was not possible [34], it is noteworthy that [32], it is possible to [29], it is necessary to [28], it is unlikely that [28] , it is well known [28], it is interesting to [26], it is known that [26], it is reasonable to [25] *Copula be + noun phrase/adjective phrase** were more likely to [116], may be due to [105], is one of the [88], is possible that the [75], is the first study [71], is important to note [70], is consistent with the [65], is in agreement with [56], is also possible that [45], were no significant differences [40], are in agreement with [39], is in line with [38], is the first to [38], was no significant difference [37], are in line with [37], could be due to [36], are likely to be [36], is not possible to [34], were not able to [34], are a number of [34], are consistent with the [34], is in accordance with [33], be more likely to [33], was not possible to [32], are consistent with previous [31], is consistent with previous [30], are similar to those [28], was significantly higher in [28], are less likely to [27], be one of the [26], is similar to the [25], is well known that [25], was not statistically significant [25]. *Verb phrase (with non-passive verb)* play a role in [55], used in this study [48], reported in the literature [41], compared with the control [37], play an important role [36], reduce the risk of [32], does not appear to [29], may contribute to the [29], did not find a [29], plays an important role [28], included in the study [27], would like to thank [27], has the potential to [26], may play a role [26], did not have a [25]. *Verb phrase with passive verb* has been shown to [149], be due to the [72], be related to the [71],be explained by the [68], has been reported to [67], may be associated with [67], was found to be [59], were found to be [52], was associated with a [47], was not associated with [46], can be used to [44], has been reported in [44], has been associated with [43], have been associated with [38], is associated with a [35], have been reported in [35], was observed in the [32], may have contributed to [32], have been found to [31], could be explained by [31], could be related to [30], are known to be [30], may be attributed to [30], be interpreted with caution [30], may not have been [30], is known to be [29], was significantly associated with [29], were found to have [29], been found to be [29], should be considered in [29], has been found to [28], is

	considered to be [27], can be explained by [26], be associated with the [26], have been described in [26], be associated with a [25], may be explained by [25] *That-clause fragments* studies have shown that [96], should be noted that [68], it may be that [65], important to note that [64], these findings suggest that [54], has been suggested that [52], that there is a [46], these results suggest that [45], has been shown that [41], that the use of [40] , our results suggest that [38], has been reported that [38], results suggest that the [36], our findings suggest that [34], our results show that [32], this study is that [32] , that the majority of [32], may be that the [29], our study is that [28], have shown that the [27] , this study suggest that [27], our data suggest that [26], research has shown that [25], studies have reported that [25]
Clause-based	**Subject-bound LBs:** that children with ASD [39], that children with autism [39] *Wh-clause fragments* which is consistent with [36], when compared to the [30], when compared with the [25] *(verb/adjective+) to-clause fragment* to be associated with [96], been shown to be [84], have been shown to [84], may be related to [82], are more likely to [76] , be related to the [71], more likely to be [61], is likely to be [59], more likely to have [58], been reported to be [50], the first study to [46], be attributed to the [43], appears to be a [41],less likely to be [28], more likely to report [28], need to be considered [27], seems to be a [27] *Pronoun/noun phrase + be (+ . . .) ** this is the first [124], studies are needed to [71] , there was no significant [59], there were no significant [58] , this study is the [49], this is consistent with [49], this is consistent with [49], the present study is [46], study is the first [46], there was a significant [44] , findings are consistent with [42] , the current study was [41], results are consistent with [37], further studies are needed [36], this study was to [36], the current study is [35], further research is needed [34] , there are a number [34], there is a need [29], our study is the [28], the sample size was [27], there are several limitations [26], there was no difference [26], this may be due [28], these findings are consistent [25], values are expressed as [25] *Adverbial clause fragment ** as measured by the [36], although we did not [25] *(noun phrase/pronoun) +V+(complement) ** research is needed to [76], we did not find [75], we were unable to [57] , we believe that the [41], we found that the [39], this study did not [38], we were able to [36] , study has some limitations [32], we did not observe [31], study has several limitations [31], the authors declared no [29] , previous studies have shown [27], we were not able [27], we also found that [26], we also found that [26], study we found that [25], we have shown that [25]
NP-based	*(connector +) Noun phrase with of-phrase fragment* the results of this [138], the results of the [134], results of this study [125], publication of this article [70], and or publication of [65], the use of a [68], findings of this study [61], results of the present [55], the results of our [53], the findings of this [52], limitations of this study [52], small number of patients [52], the use of the [49], the end of the [47], the small number of [47], the course of the [46], limitation of this study [46], results of our study [46], parents of children with [44], the magnitude of the [43] , one of the most [42], a number of limitations [41], the vast majority of [41] , the nature of the [41] , mothers of children with [40], the best of our [39], and the use of [39], best of our knowledge [38], the presence of a [38], and the presence of [37], a result of the [37], a large number of [36], the majority of patients [36], the findings of the [35], the level of the [35], a small number of [34], a wide range of [34], the majority of the [34], siblings of children with [34] , limitations of the study [32] , the size of the [32] , the effect of the [33], the validity of the [31], and the risk of [31], results of the current [31], limitation of our study [30], group of patients with [29], analysis and interpretation of [28], the retrospective nature of [28], the effectiveness of the [27], the lack of a [27], the role of the [27], limitation of the study [26], the absence of a [26], the development of a [26], the purpose of this [26], limitations of our study [26] , quality of life and [26], the presence of the [25] , to the presence of [25], a group of patients [25] *Attributive adjectives as premodifiers** the present study the [60], the small sample size [57], long term follow up [46], conflicting interests the authors [34], mini mental state examination [37], BMI body mass index [36], whey peptide- based diet [26]. *Noun phrase with post-nominal clause fragment ** the fact that the [80], the extent to which [39], those who did not [34] *Noun phrase with prepositional phrase fragment**

PP-based	an important role in [73], a significant increase in [68] , between the two groups [58], an increased risk of [56], an increase in the [49], no significant difference in [45], a risk factor for [45], patients in this study [39], a significant difference in [35], significant difference in the [34], a role in the [34] , no significant differences in [32], results from this study [32], patients in our study [31], sensitivity and specificity of [29], a better understanding of [29], sensitivity and specificity of [29], support for the research [29], findings from this study [29], important role in the [28], a high level of [26], relatively small number of [25], the difference between the [25], significant difference between the[25], a decrease in the [25] , the results from this [25], sister with an ASD [26] *Prepositional phrase expressions* in the present study [557], in the current study [271], on the other hand [214], of the present study [154], as a result of [133], in this study we [125], at the time of [108], of the current study [103], in the united states [101] , on the basis of [96], in the absence of [95] , with respect to the [94] , in this study the [90], in the case of [89], in our study the [84] , in the context of [81], of this study was [72], in addition to the [71] , in this study were [69], to the fact that [67], at the end of [66], in the presence of [63], to our knowledge this [63], at the same time [61], in the control group [60], in the number of [60], with the use of [60], in our study we [58], of this study is [58], for the development of [55], in accordance with the [54] , in contrast to the [52], in the treatment of [51], in the general population [50], by the fact that [50], with the exception of [50], for the treatment of [49] , with regard to the [48], in patients with severe [48], in the development of [47], in the form of [46], in this study was [46], in line with the [45], to the best of [43], in a study of [42], in the setting of [42], of interest with respect [42], for the first time [41], of our study is [40] , in our study were [40], with the results of [40], despite the fact that [39] , in the study by [37] , to the development of [37], with a history of [37], in critically ill patients [36] , increase the risk of [36] , in agreement with the [35], in terms of the [35] , included in this study[35] , in our study was [33] , in the patients with [33], due to the fact [33], in a previous study [32], in the long term [32] , in the majority of [32], of this study are [32], as part of the [32], over the course of [32], to the use of [32], in relation to the [31], in this study is [31], with the findings of [31], in the study of [30], similar to that of [30], at the level of [30], in the management of [29] , in the treatment group [29], with the control group [29], in patients treated with [28] , in the field of [28], in this study had [28], of the patients in [28], as a consequence of [28],of the general population [27], of the relationship between [27], in a group of [27], at high risk for [27], in a patient with [26], in this group of [26] , of the children with [26] , of the study was [26] , due to the small [26], in agreement with previous [25], of our study was [25], similar to those of [25], as a result the [25] **Subject-bound LBs**: of children with autism [111] , of children with ASD [100], in patients with AD [73], in patients with severe [48], in children with autism [47], in patients with a [46], of the patients with [44], in patients with OSA [38] , in critically ill patients [36], in the patients with [33], in children with ASD [34], for children with ASD [32], for children with autism [30], of children with ASDS [28], in patients treated with [28], of the patients in [28], in patients with OSAS [27], of the upper airway [26], in a patient with [26] *Comparative expressions/ other expressions* as well as the [135], than that of the [49], as well as in [46], higher than that of [27]

The Superscript Numbers Show the Number of Tokens in the Corpus.

Appendix C

Complete List of Functional Distribution of LBs in the Corpus

Category	Subcategory
Referential expressions	Identification/focus this is the first [124], is one of the [88], an important role in [73], is the first study [71], play a role in [55], study is the first [46], the first study to [46], one of the most [42], is the first to [38], play an important role [36], a role in the [34], those who did not [34], important role in the [28], plays an important role [28], the role of the [27], be one of the [26] Specification of attributes *quantity specification* a significant increase in [68], in the number of [60], there was no significant [59], there were no significant [58] , the small sample size [57], an increased risk of [56], small number of patients [52], an increase in the [49], the small number of [47], a risk factor for [45], no significant difference in [45], there was a significant [44], the magnitude of the [43], a number of limitations [41], the vast majority of [41], for the first time [41], were no significant differences [40], was no significant difference [37], a large number of [36], the majority of patients [36], increase the risk of [36], a significant difference in [35],

significant difference in the [34], a small number of [34], there are a number [34], the majority of the [34] , are a number of [34], a wide range of [34], in the majority of [32] , reduce the risk of [32], no significant differences in [32], study has some limitations [32], that the majority of [32], study has several limitations [31], the validity of the [31] , was significantly higher in [28], the lack of a [27], the sample size was [27], at high risk for [27], higher than that of [27], a high level of [26], there are several limitations [26], relatively small number of [25], was not statistically significant [25], a decrease in the [25]

tangible framing attributes

publication of this article [70], findings of this study [61], limitations of this study [52], limitation of this study [46], in the form of [46], the best of our [39], best of our knowledge [38], and the risk of [31], activities of daily living [30], limitation of our study [30] , limitations of the study [32], the size of the [32], limitation of the study [26], limitations of our study [26], a group of patients [25], this group of patients [23]

intangible framing attributes

as a result of [133] , on the basis of [96], in the absence of [95], with respect to the [94], in the case of [89], in the context of [81], the fact that the [80], be related to the [71], in addition to the [71], in the presence of [63], at the same time [61], in accordance with the [54], with the exception of [50], the use of the [49], with regard to the [48], the course of the [46], the nature of the [41], the extent to which [39] , in terms of the [35], is in accordance with [33], as part of the [32], in relation to the [31], the retrospective nature of [28], of the relationship between [27] , the absence of a [26], as a result the [25]

Time/place/text-deixis bundles

Time

at the time of [108], at the end of [66], the end of the [47], long term follow up [46], over the course of [32], in the long term [32], the time of the [28]

Place

in the present study [557], in the current study [271], in the United States [101], in this study the [90], in the control group [60], between the two groups [58], in the general population [50], in the setting of [42], in the study by [37], in a previous study [32], in the study of [30], in the treatment group [29], in the field of [28], in this group of [26]

deixis bundles

in this study we [125], in our study the [84], the present study we [77], in our study we [58], this study is the [49], in our study were [40], in our study was [33], in this study is [31], in this study had [28]

Subject-bound bundles*

of children with autism [111], of children with ASD [100], in patients with AD [73], in patients with severe [48], in children with autism [47], in patients with a [46], of the patients with [44] , that children with ASD [39], that children with autism [39], in patients with OSA [38], mini mental state examination [37] , BMI body mass index [36], in critically ill patients [36], in children with ASD [34], in the patients with [33], for children with ASD [32], for children with autism [30], group of patients with [29], of children with ASDS [28], in patients treated with [28], of the patients in [28], in patients with OSAS [27], sister with an ASD [26], whey peptide- based diet [26] , of the upper airway [26], in a patient with [26]

Contrast/Comparison *

is consistent with the [65], is in agreement with [56], in contrast to the [52], this is consistent with [49], in line with the [45], findings are consistent with [42], are in agreement with [39], is in line with [38], compared with the control [37], results are consistent with [37], are in line with[37], which is consistent with [36], in agreement with the [35], are consistent with the [34], are consistent with previous [31], when compared to the [30], is consistent with previous [30], similar to that of [30], are similar to those [28], there was no difference [26], the difference between the [25], is similar to the [25], these findings are consistent [25], significant difference between the [25], similar to those of [25], when compared with the [25], in agreement with previous [25]

Epistemic stance

it is possible that [310], were more likely to [116] , may be due to [105], it is likely that [94], may be related to [82], are more likely to[76] , is possible that the [75], to the fact

Stance Expressions	that [67], may be associated with [67], it may be that [65], more likely to be [61], is likely to be [59] , more likely to have [58] , by the fact that [50], it is also possible [47], that there is a [46], is also possible that [45], has been reported in [44] , it is not clear [41], we believe that the [41] , despite the fact that [39], has been reported that [38], are likely to be [36], it is not possible [35], is not possible to [34], it was not possible [34], be more likely to [33], due to the fact [33], was not possible to [32], may have contributed to [32], it is noteworthy that [32], could be explained by [31], may be attributed to [30], may not have been [30], be interpreted with caution [30], could be related to [30], it is possible to [29], may be that the [29], the authors declared no [29] , may contribute to the [29], it is unlikely that [28], less likely to be [28], more likely to report [28], this may be due [28], it is well known [28], are less likely to [27], seems to be a [27], may play a role [26], it is known that [26] , can be explained by [26], it is reasonable to [25], authors declared no potential [25], is well known that [25], may be explained by [25], studies have reported that [25] Other stance bundles to be associated with [96], has been reported to [67], used in this study [48], was associated with a [47], was not associated with [46], has been associated with [43], reported in the literature [41], it has been reported [38], this study did not [38], have been associated with [38], have been reported in [35], is associated with a [35], did not find a [29], is known to be [29] , was significantly associated with [29], is considered to be [27], included in the study [27], have been described in [26], be associated with the [26], values are expressed as [25], did not have a [25], be associated with a [25] Attitudinal/modality stance *desire* would like to thank [27] *obligation/directive* it is important to [163], it should be noted [79,] research is needed to [76] , studies are needed to[71], is important to note [70], should be noted that [68], important to note that [64], further studies are needed [36],further research is needed [34] need to be considered [27], there is a need [29], should be considered in [29], it is necessary to [28], it is interesting to [26] *Intention/prediction bundles* of this study was [72], in this study were [69], the present study is [46], this study was to [36] , this study is that [32], our study is that [28], limitations to this study [28], the purpose of this [26] *Ability* we were unable to [57], can be used to [44], were not able to [34] , we were able to[36], we were not able [27]
Discourse organizers	Topic introduction/focus of the present study [154], of the current study [103], it is difficult to [62] , the present study the [60], in this study was [46], the current study was [41], of our study is [40], patients in this study [39], included in this study [35], the current study is [35], patients in our study [31], our study is the [28], the current study the [26], of the study was [26], of our study was [25] Topic elaboration/ clarification on the other hand [214], has been shown to [149], the results of this [138], as well as the[135] ,the results of the[134] results of this study[125], studies have shown that [96] , been shown to be [84] , we did not find [75], have been shown to [75], be due to the [72] , be explained by the [68], was found to be [59] , results of the present [55], these findings suggest that [54], it has been suggested [54], the results of our [53], were found to be [52] , has been suggested that [52], the findings of this [52], it has been shown [51],results of our study [46], as well as in [46], these results suggest that [45], has been shown that [41], appears to be a [41], with the results of [40], we found that the [39], our results suggest that [38], a result of the [37], results suggest that the [36], could be due to [36], the findings of the [35], our findings suggest that [34], the effect of the [33], results from this study [32], our results show that [32], was observed in the [32], have been found to [31] , results of the current [31], we did not observe [31], have contributed to the[30], support for the research [29], been found to be [29] , were found to have [29], findings from this study

Other bundles	[29], does not appear to [29], has been found to [28], as a consequence of [28], this study suggest that [27], have shown that the [27], previous studies have shown [27], the effectiveness of the [27], due to the small [26], our data suggest that [26], we also found that [26], research has shown that [25], we have shown that [25], although we did not [25], study we found that [25] , the results from this [25] the use of a [68] , authorship and or publication [65], to our knowledge this [63], with the use of [60], of this study is [58], for the development of [55], in the treatment of [51], been reported to be [50], for the treatment of [49], than that of the [49], in the development of [47], parents of children with [44] , siblings of children with [34], mothers of children with [40], to the best of [43], be attributed to the [43], of interest with respect [42], in a study of [42], that the use of [40], and the use of [39], the presence of a [38], and the presence of [37], to the development of [37], with a history of [37], as measured by the [36], the level of the [35], of this study are [32,] to the use of [32], with the findings of [31], are known to be [30], at the level of [30], a better understanding of [29], in the management of [29], sensitivity and specificity of [29], with the control group [29], analysis and interpretation of [28], brother or sister with [28], in a group of [27], of the general population [27], has the potential to [26], quality of life and [26], the development of a [26], of the children with [26], to the presence of [25], the presence of the [25]

The Superscript Numbers Show the Number of Tokens in the Corpus.