PASAA Journal
Volume 65, January – June 2023, 125–171

Clausal or Phrasal: Variability of Lexical Bundles in L1 and Turkish L2 English Research Articles

Fatih Güngör[a*], Hacer Hande Uysal[b]

[a] Afyon Kocatepe University, Afyonkarahisar, Türkiye

[b] Hacettepe University, Ankara, Türkiye

**Corresponding author: fgungor@aku.edu.tr*

Article information	
Abstract	This study aimed to compare how native and non-native Turkish authors use four-word lexical bundles in research articles written in English in educational sciences. The two-million-word corpus included 165 research articles in L2 English and 206 in L1 English. Our corpus query resulted in 119 bundles in L2 English and 79 bundles in L1 English with a cut-off point of 20 times per million words and the occurrence in at least 10% of the articles in each corpus. The frequency results reveal that non-native authors dominantly use lexical bundles in their research articles. The classification of bundles revealed considerable structural and functional differences, and we analyzed these differences as convergent and divergent bundles. The Turkish authors used procedure (e.g., *the study was conducted*) and inferential bundles (e.g., *it is seen that*) significantly more than other functional categories. They also heavily used verb-based structures (e.g., *it was found that*), namely clausal structures rather than phrasal structures, in their research articles. These results are expected to remark on the significant role of phrase formation and modification in EAP modules' writing and grammar lectures.
Keywords	lexical bundles, research articles, educational sciences, corpus-based study

E-ISSN: 2287-0024

APA citation:	Güngör, F., & Uysal, H. H. (2023). Clausal or phrasal: Variability of lexical bundles in L1 and Turkish L2 English research articles. *PASAA*, *65*, 125–171.

1. Introduction

Recently, several studies have addressed the register and genre-specific use of multiword units (Greaves & Warren, 2010). Given that about 21 percent of academic prose terms occur in recurrent lexical bundles, and 17 percent of those recurrent bundles show up more than once (Biber et al., 1999), studies about academic genres prevail.

Although the awareness of word combinations is essential to "demonstrating membership in a specific discourse group," achieving idiomaticity is another problematic issue (Ädel & Erman, 2012, p. 81). This difficulty led researchers to scrutinize lexical bundles in the research article genre (e.g., Bal, 2010; Cortes, 2013; Hyland, 2008a, 2008b; Öztürk, 2014; Pan et al., 2016; Pérez-Llantada, 2014; Salazar, 2010) due to the hegemony of English in academic journals and several disciplines (e.g., history, biology, medical sciences, telecommunications, and psychology) to understand the rhetorical practices of a particular discipline (e.g., Cortes, 2004, 2008; Hyland, 2008a; Pan et al., 2016). However, our study differs from these studies by focusing on L1 and L2 English research articles in the discipline of educational sciences because some previous studies (e.g., Biber, 2006; Hyland, 2008a) indicate significant usage differences among different disciplines and integrate the analyses of disciplines into the genre perspective. Also, the extraction of bundles for a particular discipline has the potential to contribute to novice academic writers in a relevant discipline (Hyland & Tse, 2007). Specifically, we extracted lexical bundles from two one million corpora of research articles in the discipline of educational sciences and subsequently categorized them structurally and functionally.

2. Literature Review

Formulaic sequences play a pivotal role for native- and non-native speakers in effectively using a language (Schmitt, 2005). Lexical bundles, called "recurrent discourse building blocks" (Biber et al., 1999, p. 991), are central for processing information (Schmitt, 2005; Wray, 2002), interacting with the audience, and raising a voice in communities (Wray, 2002). Although the knowledge of word combinations is essential for "demonstrating membership in a specific discourse community," achieving idiomaticity is difficult to the same degree (Ädel & Erman, 2012, p. 81). To be specific, the effective use of words, collocations, and lexical bundles has been associated with the competent language use within a particular register or genre (Biber & Barbieri, 2007; Cortes, 2004; Durrant, 2014; Ellis, Simpson-Vlach, & Maynard, 2008; Hyland, 2008b), and the lack of formulaic writing is frequently characterized with poor writing, L2 writers (Bestgen & Granger, 2014; De Cock et al., 1998; Granger, 1998; Peromingo, 2012) and lack of proficiency in English academic writing (Cortes, 2004, 2008; Durrant & Mathews-Aydınlı, 2011; Li & Schmitt, 2009; Neff, 2008; Römer, 2009a).

Biber et al. (1999) were the gatekeepers analyzing the significance of lexical bundles in academic texts, and their study inspired researchers to delve into the use of bundles in specific genres and registers of academic writing (e.g., Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; Durrant, 2017; Gilquin & Paquot, 2008; Leńko-Szymańska, 2014; Qin, 2014; Römer, 2009b). Although the majority of these studies attempted to identify lexical bundles in college student writing, just a few studies (e.g., Bal, 2010; Cortes, 2013; Hyland, 2008a, 2008b; Öztürk, 2014; Pan et al., 2016; Pérez-Llantada, 2014) investigated lexical bundles in research articles as a specific genre of academic writing. Of these studies, a few focused on disciplinary variation in history, biology, medical sciences, telecommunications, and psychology research articles (e.g., Cortes, 2004, 2008; Esfandiari & Barbary, 2017; Hyland, 2008a; Pan et al., 2016).

Among these studies analyzing research article genre, Hyland (2008a) investigated the structures and functions of 4-word lexical bundles in a 3.5 million-word corpus, including research articles, doctoral and master dissertations, and found that most of the lexical bundles included of-phrase. Science and engineering texts often consist of research-oriented bundles, while text-oriented bundles dominate the applied linguistics and business studies corpora. The participant-oriented bundles were mostly used in the social science texts to indicate the writers' stance. Similarly, Pan et al. (2016) compared the lexical bundles in the telecommunications discipline produced by L1 English authors and L2 English Chinese authors. and found considerable structural and functional differences, and they concluded that even the experts of an academic discipline might misuse or overuse lexical bundles. For instance, L1 professionals employed bundles consisting of noun phrases and prepositional phrases more, while L2 authors preferred to use bundles including verbs, clause fragments, and passive verb structures. Some functional differences between L1 English and L2 English were also found in the study. A noticeable result similar to the other studies in the literature was the functionally different use of lexical bundles and the misuse of some lexical bundles. For instance, L2 authors significantly used more stance-oriented and fewer research-oriented lexical bundles than L1 English authors.

Salazar (2010, 2014) analyzed lexical bundles in biomedical research articles in Philippine English and British English in her two consecutive studies. The results indicated that the British authors used a more comprehensive range of lexical bundles than the Filipino authors. The authors from the Philippines preferred passive constructions and avoided using first personal pronouns in their articles. In terms of functional usage, British authors used more participant-oriented bundles, while Filipino authors used more research-oriented bundles. Although the range of lexical bundle use was limited in the Philippine English corpus, Filipino authors overuse this limited number of lexical bundles in their research articles. Lastly, Pérez-Llantada (2014) investigated the Spanish-English Research Article Corpus (SERAC 2.0), and the results of the structural

categorization clearly showed that L2 writers' bundle usage is far from native-like. Regarding functional categorization, referential and text-organizing bundles were shared by L1 and L2 English. Besides, 61% of the bundles used in L1 Spanish and L2 English were text-organizing bundles. When Pérez-Llantada (2014) analyzed the divergent usage of bundles, she found that half of the distinctive bundles of L1 English belong to the stance category. However, the stance function is not frequent in L2 English, and Pérez-Llantada (2014) attributes this to L1 transfer.

Considering the lack of discipline-specific lexical bundle studies both in educational sciences and with Turkish L2 English, we have scrutinized the convergent and divergent structures and functions of lexical bundles in L1 and Turkish L2 English educational research articles by asking the following research questions:

- What lexical bundles do L1 English and Turkish L2 English have in common? Do the frequencies of these bundles serve distinctive functions and structures?
- What lexical bundles differentiate researchers in educational sciences from L1 English and Turkish L2 English writing? Do these lexical bundles serve distinctive structures and functions?

3. Methodology

3.1 Multilingual Corpus of Research Articles

To "adequately represent the occurrence of the features being studied" (Biber, 2006, p. 51) and adhere to the "convention" of one million-word corpus for scrutinizing lexical bundles (Cortes, 2015, p. 205), we designed a two-million-word "specialized corpus" (the MCRA-L1, L1 English research articles; the MCRA-L2, L2 English research articles by Turkish authors) for our particular research aim (Kennedy, 1998, p. 20). Even though some researchers (e.g., Pérez-Llantada, 2014) adjusted the number of articles, we chose to adjust the size of each subcorpora as one-million words, like the investigation of Pan et al. (2016), because of the gap

between the article lengths (165 L1 English articles and 206 L2 English articles) in L1 and L2 English.

We collected the research articles from peer-reviewed journals (indexed by ERIC and SSCI between 2006 and 2016) in educational sciences. The collected articles were classified into 11 fields of educational sciences (arts education, instructional technologies, elementary and middle school education, language education, music education, mathematics education, physical education, science education, pre-school education, special education, and social studies education) to ensure balance among the fields. We also added interdisciplinary education studies to the collection of research articles under a new category. To avoid misinterpretation, we removed page numbers, journal and author names, tables, figures, and quotes from interviews. The data were ready after converting pdf files to a simple text file format. When we achieved the one million-word target for each subcorpora, the corpus statistics were as in Table 1.

Table 1

The Corpus Statistics

	MCRA-L1	**MCRA-L2**
Tokens (Running words)	1,000,019	1,000,009
The number of articles	165	206
Types (Distant words)	25,445	24,743
Type/token ratio	2.61	2.58
Standardized type/token ratio	38.13	33.35
STTR std. dev.	61.16	66.30
Number of Sentences	34,821	34,978
Mean words per sentence	26.47	27.46
Standard deviation	78.35	100.43

The number of distinct words seems to be similar in two subcorpora (25,445 and 24,743), and the slight difference in favor of native authors was an expected

result. This result is reflected in the type/token ratios, which can be interpreted as the variety of word types used by L1 English authors. Although the number of sentences and mean words per sentence seem similar, the standard deviation (100.43) was higher due to the word numbers' fluctuation in the sentences in the MCRA-L2. As there are just ten words between two corpora, and this does not change the relative frequency per million words, there is no need for normalization in this study.

3.2 Extraction and Classification of Lexical Bundles

Because of their manageability in size, four-word bundles were focused on much more frequently than the others (Chen & Baker, 2010; Hyland, 2008a). They also "hold three-word bundles in their structures" (Cortes, 2004, p. 401) and are ten times more common than five-word lexical bundles (Cortes, 2004; Hyland, 2012). Along with their high frequency, four-word lexical bundles are rich in terms of different structures and functions (Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a). Therefore, this study focused on four-word lexical bundles.

Frequency and dispersion parameters are "somewhat arbitrary" for the lexical bundle research (Biber et al., 2004, p. 376). Due to the difference in the lengths of articles in both corpora (165 in the MCRA-L1 and 206 in the MCRA-L2), we included the lexical bundles appearing in at least 10% of the research articles. Thus, we avoided the idiosyncratic or similar uses of individual authors. We also put a threshold of 20 times per million words, which many researchers find conservative (see Cortes, 2004, 2008; Hyland, 2008a; Pérez- Llantada, 2014).

Although lexical bundles are incomplete in academic writing, they have "strong grammatical correlates" (Biber et al., 2004, p. 380). Taking these grammatical correlates into account, we tried to classify lexical bundles into 12 major categories, as suggested by Biber et al. (1999). Eleven lexical bundles in the MCRA-L2, which failed to fit into these categories, were classified into the new four additional categories (other noun phrases, other verbal fragments, other

adjectival phrases, and other passive fragments), and we slightly changed the category (verb phrase +) that-clause fragment as (verb phrase or noun phrase) + that-clause fragment following the study of Salazar (2014). We omitted the category of copula *be* + noun phrase/adjective phrase from the structural taxonomy due to the zero-occurrence, and the number of the structural categories increased to 15 in the context of this study.

We utilized Hyland's (2009) taxonomy and its developed version (Salazar, 2014) to examine the functions of lexical bundles in the MCRA because the taxonomy of Hyland offers "a better fit", particularly for the specific purpose of research writing (Durrant, 2017, p. 16). The results of this functional categorization verified that lexical bundles convey particular meanings. Some bundles might have more than one function (Biber et al., 2004; Byrd & Coxhead, 2010; Salazar, 2014), and identifying bundles by their most common use will be a good idea in such situations (Biber et al., 2004). This study also adopted the same methodology and classified bundles concerning their most common use. In addition to the researchers, one external applied linguist classified the randomly chosen 30% of the total bundles for the measure of agreement. The intraclass correlation coefficient (ICC) with the absolute agreement selection was calculated to assess "the consistency between judges' ratings of a set of objects" (Field, 2009, p. 788). Three raters reached 92% agreement (ICC=.916 at the 95% confidence interval), and this high agreement percentage confirmed the researcher's capability to categorize the rest of the lexical bundles.

We calculated log-likelihood (LL) values through the UCREL log-likelihood calculator (http://ucrel.lancs.ac.uk/llwizard.html) to compare the frequencies of lexical bundles and to detect statistically overused and underused lexical bundles because the frequency of occurrence should not be perceived as a sole tool to evaluate the use of the learners as native-like or not (Ellis et al., 2013). The threshold for the statistical significance was considered to be .05 (Brezina, 2016),

and Brezina and Meyerhoff (2014) calculated the following loglikelihood values for the p-values of .5, .01, .001, and .0001: 3.84, 6.63, 10.83, and 15.13.

4. Results and Discussion

The corpus query resulted in a total of 119 bundles in L2 English and 79 bundles in L1 English, which justifies the formulaicity and fixedness of the research article genre with previous studies (e.g., Pérez-Llantada, 2014). The frequency results reveal how lexical bundles are used dominantly in the research articles of non-native authors. The previous studies, for example Staples et al. (2013), also showed that non-native writings consist of many more lexical bundles than native writings do without distinguishing the functional or structural categories. We compared the most frequent 50 lexical bundles in previous studies (e.g., Bal, 2010; Öztürk, 2014; Pérez-Llantada, 2014), and a significant number (f=31, 62%) of lexical bundles in the MCRA-L2 were found in at least one of these three studies. The difference seems to be related to disciplines, and corpus size and characteristics.

4.1 Convergent Use of Lexical Bundles

Two corpora shared thirty-four lexical bundles (see Appendix 1), and these bundles will be called convergent lexical bundles. Given their high number of uses, the convergent bundles serve native and non-native scholars in structuring their research articles. Clearly, from these results, prepositional phrases are among the most frequently used postmodifiers in the academic prose of the 20th century (Parkinson & Musgrave, 2014), and their frequency considerably increased in academic writing in the last two centuries (Biber & Gray, 2011). Although the other prepositional phrase fragments were not as frequent as the *of*-phrase fragments in the MCRA-L1 and the MCRA-L2, they are also becoming more frequent in academic prose (Biber & Gray, 2011; Hyland, 2008b; Öztürk, 2014). The noun phrase with *of*-phrase fragment (e.g., *the aim of this, the findings of the*) remained as the most frequent noun structure in both corpora as in some other studies (see Biber et al., 1999; Byrd & Coxhead, 2010; Hyland, 2008b; Pan et al.,

2016; Salazar, 2014). This can be interpreted as the productive and fixed feature of the noun phrases with an *of*-phrase fragment (Biber et al., 2003; Chen & Baker, 2010; Stubbs, 2007). In addition, Scott and Tribble (2006, p. 99) argue that *of*-phrase fragments are the sign of "information-rich noun phrases."

In addition to type and token frequencies, log-likelihood tests were carried out to determine significant differences in structural classifications across the MCRA-L1 and the MCRA-L2, and overuses at the $p<0.0001$ level were shown with an asterisk. Table 2 shows that the prepositional phrase with embedded of-phrase fragment was the most common structure with 35.29% of the types in both corpora, and it was followed by the noun phrase with of-phrase fragment with 35.29% of the types. The log-likelihood statistics also indicated the overuse of prepositional-phrase fragments, noun structures, and verb structures by non-native authors at the $p<0.0001$ level in terms of their token numbers.

Table 2

Structures of the Convergent Lexical Bundles

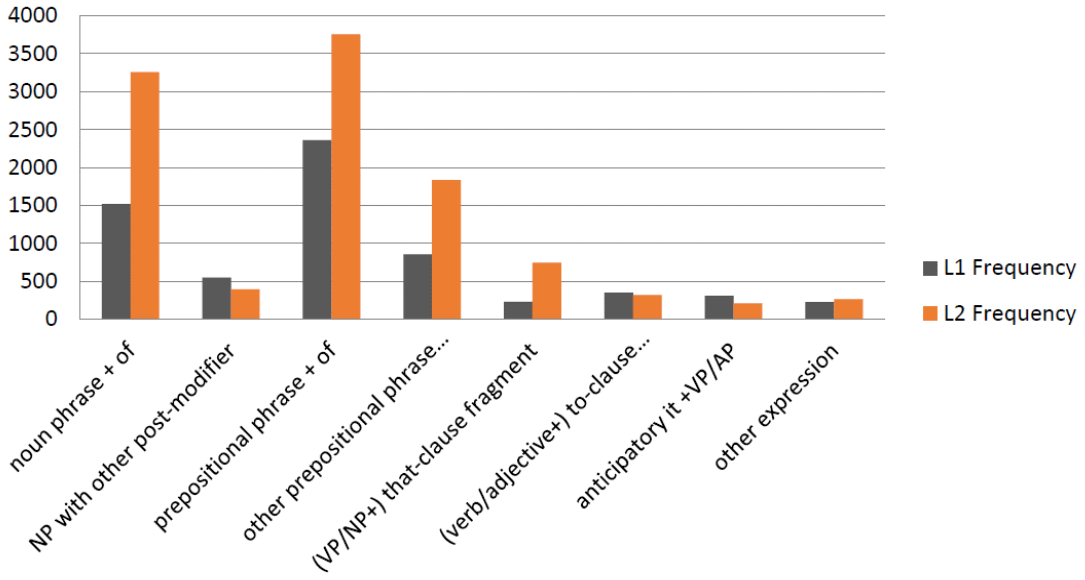
Structure	MCRA-L1		MCRA-L2				LL
	Types	%	Types	%	Types	%	
Noun Structure	**12**	**35.29**	**2068**	**32.35**	**3656***	**33.90**	**446.39**
Noun phrase with *of*-phrase fragment	10	29.41	1520	23.78	3260*	30.23	648.18
Noun phrase with other post modifier fragments	2	5.88	548*	8.57	396	3.67	24.58
Other noun phrases	0	0.00	0	0.00	0	0.00	-
Prepositional-phrase fragments	**16**	**47.06**	**3216**	**50.31**	**5592***	**51.85**	**648.95**
Prepositional phrase with embedded *of*-phrase fragment	12	35.29	2360	36.92	3756*	34.83	321.47
Other prepositional phrases (fragment)	4	11.76	856	13.39	1836*	17.03	365.09

Structure			MCRA-L1		MCRA-L2		LL
	Types	%	Types	%	Types	%	
Verb Structures	**5**	**14.71**	**884**	**13.83**	**1272***	**11.80**	**70.21**
Passive verb + prepositional phrase fragment	0	0.00	0	0.00	0	0.00	-
Other passive fragments	0	0.00	0	0.00	0	0.00	-
(Verb phrase/noun phrase +) *that*-clause fragment	2	5.88	228	3.57	744*	6.90	288.51
Anticipatory *it* + verb phrase/adjective phrase	1	2.94	308*	4.82	208	1.93	19.50
(Verb/adjective +) *to*-clause fragment	2	5.88	348	5.44	320	2.97	1.17
Pronoun/noun phrase + *be* (+…)	0	0.00	0	0.00	0	0.00	-
Other verbal fragments	0	0.00	0	0.00	0	0.00	-
Other Structures	**1**	**2.94**	**224**	**3.50**	**264**	**2.45**	**3.28**
Adverbial clause fragment	0	0.00	0	0.00	0	0.00	-
Other adjectival phrases	0	0.00	0	0.00	0	0.00	-
Other expressions	1	2.94	224	3.50	264	2.45	3.28
Total	**34**	**100.00**	**6392**	**100.00**	**10784**	**100.00**	**1135.63***

Note: *99.99th percentile; 0.01% level; $p < 0.0001$; critical LL value = 15.13

Figure 1

Structures of the Convergent Lexical Bundles



In terms of their functions, Turkish authors overused research-oriented bundles and text-oriented lexical bundles at the $p<0.0001$ level in terms of their token numbers. The text-oriented bundles ranked as the most prominent functional category, having 56.75% of the tokens in the MCRA-L2. Similar to the previous studies (see Hyland, 2008a; Pan et al., 2016; Salazar, 2014), the text-oriented bundles were found to be the most dominant category of the convergent bundles in both corpora, and the category of the research-oriented bundles preceded this category. The research-oriented bundles came out as the most common functional category, having 48.16% of the tokens in the MCRA-L1. Location bundles, the most frequent subcategory of research-oriented bundles, comprised of noun and prepositional phrases indicating the beginning and end of processes.

Table 3

Functions of the Convergent Lexical Bundles

Function			MCRA-L1		MCRA-L2		LL
	Types	%	Types	%	Types	%	
Research-oriented bundles	**16**	**47.71**	**2884**	**48.16**	**4264***	**39.54**	**268.10**
Location	5	14.71	1364	22.78	2032*	18.84	132.26
Procedure	4	11.76	388	6.48	748*	6.94	116.08
Quantification	2	5.88	536	8.95	376*	3.49	28.22
Description	4	11.76	508	8.48	676*	6.27	23.92
Grouping	1	2.94	88	1.47	432*	4.01	248.02
Text-oriented bundles	**16**	**47.06**	**2708**	**45.22**	**6120***	**56.75**	**1353.70**
Additive	4	11.76	476	7.95	1976*	18.32	985.68
Comparative	1	2.94	100	1.67	204*	1.89	36.31
Inferential	1	2.94	140	2.34	552*	5.12	262.35
Causative	2	5.88	444	7.41	1844*	17.10	920.22
Structuring	0	0.00	0	0.00	0	0.00	-
Framing	5	14.71	1100	18.37	1120	10.39	0.18
Citation	0	0.00	0	0.00	0	0.00	-
Generalization	0	0.00	0	0.00	0	0.00	-
Objective	3	8.82	448	7.48	424	3.93	0.66
Participant-oriented bundles	**2**	**5.88**	**396**	**6.61**	**400**	**3.71**	**0.02**
Stance	2	5.88	396	6.61	400	3.71	0.02
Engagement	0	0.00	0	0.00	0	0.00	-
Total	**34**	**100.00**	**5988**	**100.00**	**10784**	**100.00**	**1390.76***

Note: *99.99th percentile; 0.01% level; $p < 0.0001$; critical LL value = 15.13

The stance-oriented bundles were used in almost the same numbers, and this study corroborates with the previous studies (e.g., Chen & Baker, 2010; Hyland, 2008a; Pan et al., 2016; Salazar, 2014) on the limited use of stance bundles in the academic writing. Two bundles (*it is important to,* and *the fact that the*) were shared by native and non-native authors, and two bundles (*it is possible to*, and *it*

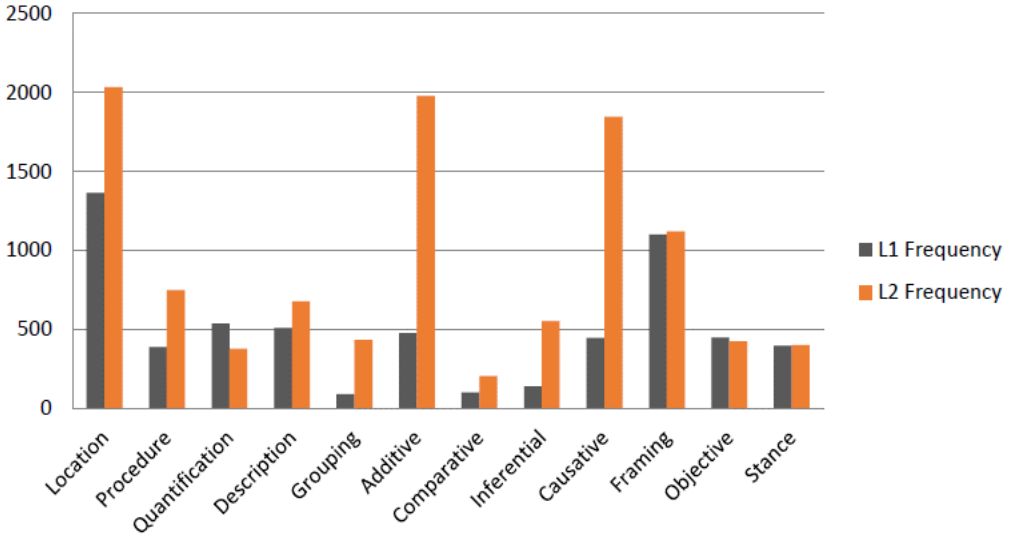
is possible that) have a slight difference between the two corpora. The first of these bundles, *it is important to*, was used to express "the writer's judgment of necessity/importance" (Hyland, 2008a, p. 19) in addition to the bundles *it is necessary to* and *an important role in* that are found in the MCRA-L2. In other words, this kind of bundle requires readers "to notice something in the text and thereby to lead them to a particular interpretation" (Hyland, 2008a, p. 19). Other stance bundles, *the fact that* and *it is possible to/that*, were used to "express degrees of certainty rather than uncertainty," as suggested by Biber et al. (2004, p. 389). Pérez-Llantada (2014, p. 91) categorized the bundle, *it is possible that* in the same category and argued that this bundle was used to make "non-face-threatening" authorial claims. The stance bundles in this study were typically constructed by the hedging anticipatory *it*-structures, copula *be* + likely to, and the hedging nouns (the fact).

(1) "Therefore, *it is important to* find answers to questions asking what can be done to increase student interest toward science classes and to turn science classrooms into more enjoyable places." (MCRA-L2, Science Education, Article 29)

(2) "This research involves several key concepts, and *it is important to* identify these concepts, define them and explain possible relationships." (MCRA-L1, Instructional Technologies, Article 7)

Figure 2

Functions of the Convergent Lexical Bundles



These structure frames were also mentioned in some previous studies (e.g., Chen & Baker, 2010; Hyland, 1998, 2008a). Clearly, the native and non-native authors used the stance bundles to highlight their arguments' significance and use their authorial authority within the current context.

4.2 Divergent Use of Lexical Bundles

The number of divergent lexical bundles (see Appendix 2) in the MCRA-L1 and the MCRA-L2 were 45 and 85, respectively. Table 4 summarizes the lexical bundles' structural classification in the MCRA-L2 with detailed type and token frequencies and percentages. Nearly half (51.11%) of the divergent lexical bundles were prepositional-phrase fragments in the MCRA-L1, and many previous studies (Byrd & Coxhead, 2010; Chen & Baker, 2010; Öztürk, 2014; Pan et al., 2016; Qin, 2014) support the higher proportionate usage of prepositional phrase fragments in native corpora. Verb phrase-based structures stand out as the most frequent structural category in divergent bundles with 55.29% of the types and 62,59% of the tokens in the MCRA-L2 as in some other studies (Güngör & Uysal, 2016, 2020; Chen & Baker, 2010; Öztürk, 2014; Pan et al., 2016). Of these verb phrase-based structures, the most frequent type, and the highest token percentage pertain to the anticipatory *it* + verb phrase/adjective phrase, and the second most frequent structure in the MCRA-L2 was the passive verb + prepositional phrase fragment.

The use of these two structures reminds a “depersonalized mode” in academic prose (Cortes, 2004, p. 408). Moreover, learners used the anticipatory *it*-phrases to prove their arguments' legitimacy much more frequently than academic professionals did (Hewings & Hewings, 2002) or to achieve a persuasive style (Güngör, 2019). In this context, the use of anticipatory it-phrases might show a naïve desire of Turkish authors to find a presence in competitive scholarly publication venues. Also, Güngör and Uysal (2020) prove that the reason for this divergent use in Turkish L2 English refers to crosslinguistic influence.

Table 4

Structures of Divergent Lexical Bundles

Structure	MCRA-L1				MCRA-L2				Token
	Types	%	Token	%	Types	%	Token	%	LL
Noun Structure	**13**	**28.89**	**1876**	**32.61**	**19**	**22.35**	**3892***	**18.27**	**719.72**
Noun phrase with *of*-phrase fragment	11	24.44	1496	26.01	11	12.94	1888*	8.86	45.51
Noun phrase with other post modifier fragments	2	4.44	380	6.61	6	7.06	1488*	6.98	702.49
Other noun phrases	0	0.00	0	0.00	2	2.35	516*	2.42	715.33
Prepositional-phrase fragments	**23**	**51.11**	**2936**	**51.04**	**13**	**15.29**	**2860**	**13.42**	**1.00**
Prepositional phrase with embedded *of*-phrase fragment	19	42.22	2528*	43.95	7	8.24	1316	6.18	388.74
Other prepositional phrases (fragment)	4	8.89	408	7.09	6	7.06	1544*	7.25	704.66
Verb Structures	**8**	**17.78**	**852**	**14.81**	**47**	**55.29**	**13336***	**62.59**	**13224.35**
Passive verb + prepositional phrase fragment	0	0.00	0	0.00	12	14.12	2112*	9.91	2927.85
Other passive fragments	0	0.00	0	0.00	6	7.06	904*	4.24	1253.21
(Verb phrase/noun phrase +) *that*-clause fragment	0	0.00	0	0.00	4	4.71	688*	3.23	953.77
Anticipatory *it* + verb phrase/adjective phrase	3	6.67	328	5.70	14	16.47	5060*	23.75	4997.65

Structure	MCRA-L1				MCRA-L2				Token
	Types	%	Token	%	Types	%	Token	%	LL
(Verb/adjective +) *to*-clause fragment	4	8.89	416	7.23	7	8.24	1488*	6.98	640.35
Pronoun/noun phrase + *be* (+...)	1	2.22	108	1.88	2	2.35	644*	3.02	423.63
Other verbal fragments	0	0.00	0	0.00	2	2.35	2440*	11.45	3382.56
Other Structures	**1**	**2.22**	**88**	**1.53**	**6**	**7.06**	**1220***	**5.73**	**1168.32**
Adverbial clause fragment	0	0.00	0	0.00	2	2.35	464*	2.18	643.24
Other adjectival phrases	0	0.00	0	0.00	1	1018	264*	1.24	356.98
Other expressions	1	2.22	88	1.53	3	3.53	492*	2.31	310.25
Total	**45**	**100**	**5752**	**100**	**85**	**100**	**21308***	**100**	**9514.96**

Note: *99.99th percentile; 0.01% level; $p < 0.0001$; critical LL value = 15.13

Figure 3

Structures of the Divergent Lexical Bundles

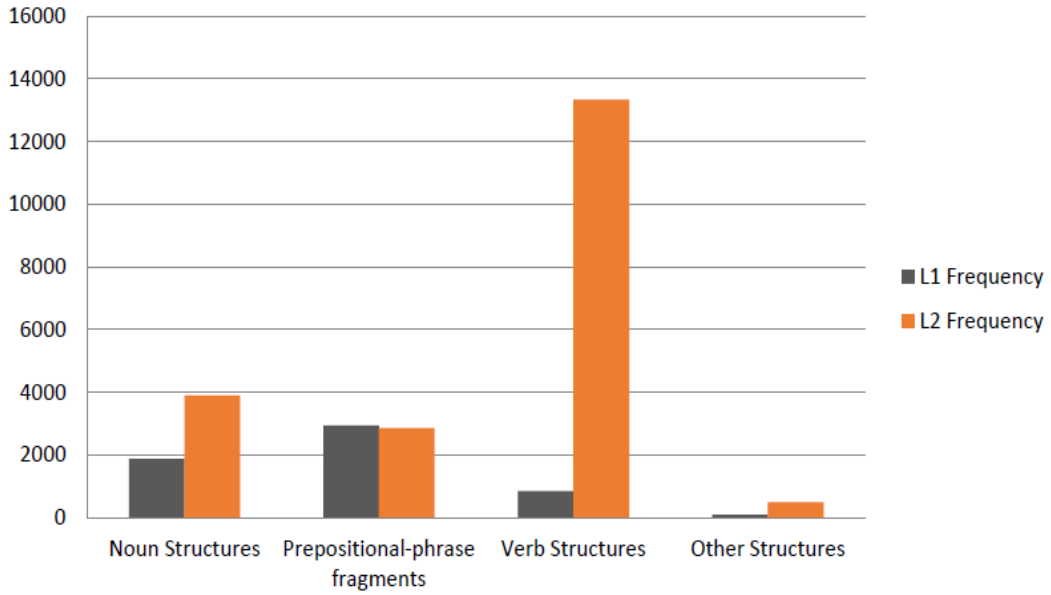


Table 5 shows that the MCRA-L2 authors used inferential, framing, comparative, structuring, objective, causative, and generalization bundle types significantly more than the MCRA-L1 authors.

Table 5

Functions of the Divergent Lexical Bundles

Function	MCRA-L1				MCRA-L2				Token
	Types	%	Tokens	%	Types	%	Tokens	%	LL
Research-oriented bundles	**28**	**62.22**	**3628**	**63.07**	**29**	**34.12**	**4324***	**21.75**	**61.00**
Location	5	11.11	688*	11.96	0	0.00	0	0.00	953.77
Procedure	6	13.33	868	15.09	19	22.35	2812*	14.15	1081.00
Quantification	6	13.33	676*	11.75	0	0.00	0	0.00	937.13
Description	8	17.78	1104	19.19	6	7.06	992	4.99	5.99
Grouping	3	6.67	292	5.08	4	4.71	520*	2.62	64.89
Text-oriented bundles	**13**	**28.89**	**1700**	**29.55**	**49**	**57.65**	**13736***	**69.11**	**10692.69**
Additive	0	0.00	0.00	0.00	0	0.00	0	0.00	0.00
Comparative	0	0.00	0	0.00	8	9.41	2000*	10.06	2772.59
Inferential	2	4.44	248	4.31	16	18.82	5612*	28.24	6069.74
Causative	0	0.00	0	0.00	2	2.35	520*	2.62	720.87
Structuring	0	0.00	0	0.00	7	8.24	2712*	13.64	3759.63
Framing	9	20.00	1264	21.97	8	9.41	1600*	8.05	39.51
Citation	0	0.00	0.00	0.00	0	0.00	0	0.00	0.00
Generalization	0	0.00	0	0.00	3	3.53	424*	2.13	587.79
Objective	2	4.44	188	3.27	5	5.88	868*	4.37	474.68
Participant-oriented bundles	**4**	**8.89**	**424**	**7.37**	**7**	**8.24**	**1816***	**9.14**	**931.67**
Stance	4	8.89	424	7.37	4	4.71	748*	3.76	90.75
Engagement	0	0.00	0	0.00	3	3.53	1068*	5.37	1480.56
Total	**45**	**100.00**	**5752**	**100.00**	**85**	**100.00**	**19876***	**100.00**	**8235.52**

Note: *99.99th percentile; 0.01% level; $p < 0.0001$; critical LL value = 15.13

Table 5 also shows the existence of divergent inferential bundles (*f*=16) in the MCRA-L2, and clausal structures took the lead with 87.5%. The high number of inferential bundles might be related to its significant role in presenting the results of studies. Eight bundles were passive, and six were active structures. Although the reporting verbs used in the current subcorpora corroborate with the study of Hyland (2000, 2002) in which he analyzed the reporting verbs in the research articles across disciplines, the Turkish authors seem to deviate from the norms of native academic writing by using clausal and passive structures

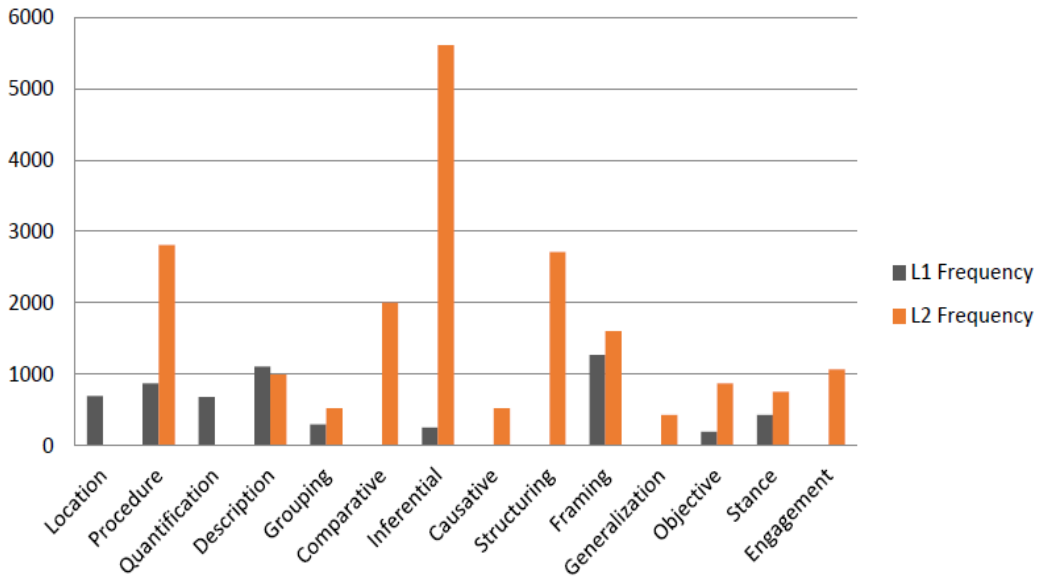
redundantly and excessively. This deviation might be linked to the predominance of phrasal structures in the academic writing of L1 English (see Adel & Erman, 2012; Biber & Conrad, 2019; Biber et al., 2004; Biber et al., 1999; Biber & Gray, 2010, 2011; Byrd & Coxhead, 2010; Chen & Baker, 2010) and the transition from the clausal style of less proficient writers to the phrasal style of the advanced writers (Biber et al., 2011). The following concordance lines show that the anticipatory *it*-structures were used redundantly in some cases (Examples 3 and 4), and these redundant clausal structures can be expressed with phrasal structures shortly.

(3) "It was determined that there is a significant and positive relation at the intermediate level between organizational loyalty and organizational power…" (MCRA-L2, Interdisciplinary Education Studies, Article 127)

(4) "It is seen that there is a significant difference between pre-test and posttest results of the experimental group in the table…" (MCRA-L2, Language Teaching, Article 19)

Figure 4

Functions of the Divergent Lexical Bundles



The divergent text-oriented bundles comprised a high proportion of the framing and comparative bundles in the MCRA-L2. Framing bundles were mainly in the form of *be* + prepositional phrases with embedded *of*-fragment (6 out of 9 types in the MCRA-L1 and 4 out of 8 types in the MCRA-L2), as also suggested by Hyland (2008a) and Salazar (2014), and they primarily functioned to frame the arguments of the authors by indicating specific cases (*in the case of*), indicating a method or style of doing something (*as a means of, as a way of, and as a way to*), specifying conditions and limitations (*in the context of, on the basis of, within the context of, within the framework of, within the scope of, is based on the, with regard to the,* and *in relation to the*), and highlighting aspects of an argument (*in terms of the*). The statistically significant overuse of the comparative lexical bundles in this context was in line with Salazar's (2014) study. Comparative bundles were mainly used to compare the statistical differences between two constructs or elements in the research article corpora of educational sciences (see Example 5). Durrant (2013) also confirms this by retrieving 18 four-word lexical bundles describing statistical significances in his 2.5 million-word academic corpus in Turkish.

(5) "There is *a significant difference between* 4th and 5th grades students in terms of achievement level of the specific objectives in English courses." (MCRA-L2, Language Teaching, Article 7)

The structuring lexical bundles were the fourth category of text-oriented bundles used significantly more frequently by L2 English authors. While the authors in the MCRA-L1 did not use any structuring lexical bundles in their research articles, their non-native counterparts used seven structuring four-word lexical bundles. The use of structuring lexical bundles might indicate a robust reader orientation, and these features are known to enhance understanding (Pan et al., 2016). The structuring lexical bundles took the form of passives in the MCRA-L2, and they were mostly used with a preposition to signal a table or a figure or data in the texts (see Example 4). This might be attributed to the finding that

Turkish authors prefer to present their results with many tables instead of interpreting results in words.

(6) "*As can be seen* from the data in Table 2 based on the expressions of participants, squares were mostly constructed based on rectangle (21 people)." (MCRA-L2, Mathematics Education, Article 31)

The objective lexical bundles included seven percent of the bundle types in both corpora, but non-native authors used objective bundles significantly more frequently than their native counterparts. Most of the objective bundles (80%) were in the form of (verb/adjective) + *to*-clauses (see Example 7). Although native authors did not use any lexical generalization bundles in the MCRA-L1, their Turkish counterparts used three types (*it is thought that, it is believed that, and can be defined as*) for this category. In addition to the generalization function of the bundles, *it is believed that* and *it is thought that*, the Turkish authors also expressed their expectations on how their studies would fill the gap in their fields (see Example 8) or they claimed the centrality of their studies (see Example 9) by generalizing their arguments. The first bundle (*it is thought that*) was used one time per million words in COCA and five times per million words in BNC, and the second bundle (*it is believed that*) was used two times per million words in COCA and one time per million words in BNC. The third bundle (*can be defined as*) was used two times per million words in both COCA and BNC. These three structures (*it is thought that, it is believed that,* and *can be defined as*) cannot be defined as ungrammatical or wrong due to the frequencies, but it can be said that the Turkish authors overused these bundles significantly for some potential reasons such as implicit learning and crosslinguistic influence. For the causative function, the Turkish authors used two different bundles (*the effect of the* and *the findings of the*). These bundles, used almost the same as the native norms, were also retrieved by similar studies (see Bal, 2010; Hyland, 2008b; Öztürk, 2014; Pérez-Llantada, 2014).

(7) "*This study aims to* determine pre-service music teachers' opinions about the significance of the choir lesson." (MCRA-L2, Music Education, Article 14)

(8) "*It is believed that* new studies related to the attitudes of other segments of society such as children, the youth and retired individuals towards the mentally retarded will provide more detailed and clear ideas in the field." (MCRA-L2, Interdisciplinary Education Studies, Article 17)

(9) "*It is thought that* studying the effects of these reforms in chemistry education on educating people, as chemical literate is very important." (MCRA-L2, Science Education, Article 24)

Research-oriented divergent bundles were the most commonly used functional category in the MCRA-L1 and the second most widely used category in the MCRA-L2 (see Table 5). While the MCRA-L2 authors used procedure and grouping bundles significantly more than the MCRA-L1 authors, their native counterparts more frequently used location and quantification bundles. Description bundles were used similarly in both corpora. These bundles were mainly used for the description of the research objects and contexts in the MCRA-L1 and the MCRAL2, and they were typically constructed by the noun phrase + *of*-structures, as also suggested by Hyland (2008a). Some of the description bundles were used to identify the aim of the studies. These kinds of bundles (e.g., *the aim of this, the main purpose of,* and *the purpose of this*) were significantly overused in the MCRA-L2; however, there were some similar description bundles in the MCRA-L1 (e.g., *the focus of the,* and *the purpose of this*). The three of these bundles signified the degree of impact, importance, and quality in the MCRA-L1.

The procedure bundles show how research studies were carried out. In other words, they are at the heart of research-oriented bundles with their function to describe actions, processes, methods, and activities. The dominance of procedure bundles in research-oriented bundles corroborated the studies of

Hyland (2008a) and Salazar (2014). In the MCRA-L1, four out of six divergent procedure bundles were the prepositional phrase fragments with embedded of fragments, and one procedure bundle was the noun phrase with embedded of fragment. Seventeen divergent procedure bundles in the MCRA-L2 were clausal structures, and 15 of these structures were passive. Non-native authors in the educational sciences use a higher number of procedure bundles. This might show that they are well aware of the importance of unbiased reporting of the research processes (Salazar, 2014). The passive structure of most bundles can also be interpreted as their effort to sound objective. However, the extensive use of procedure bundles by novice writers such as postgraduate students is not an expected result, and this might be attributed to their attempts to show their competence in research methods (Hyland, 2008a). Considering these two perspectives together, non-native authors might prefer using many lexical bundles to compensate for their disadvantage while writing in a foreign or second language. Remarkably, the methodology part seems ideal for using lexical bundles to get credit and spend these in the further discipline-specific parts.

Of the research-oriented bundles, the quantification subcategory was a subcategory in which the native authors used bundle types and tokens significantly more frequently than their non-native Turkish counterparts did. Pan et al. (2016) found similar results and suggested that novice writers might not realize the importance of the quantification bundles in their academic writing. Cortes (2004) also links this absence of quantification expressions to a sign of novice or student writing. The Turkish authors used the bundle *the extent to which* significantly less than the native authors and abstained from using the bundle *the degree to which*. These expressions might be challenging to acquire due to their different word order from Turkish. Location bundles were one of the two categories overused significantly in the MCRA-L1. The location bundles comprised mostly temporal bundles and prepositional phrases in both corpora. Of 15 bundles in both corpora, 11 bundles were phrase frames of the words like beginning, end, and start.

One group of research-oriented bundles, grouping, was significantly overused by the non-native authors in terms of the bundle tokens. However, the grouping bundle used by Turkish authors had a uniform structure, including the word one or a in their structures, and such bundles were used to express abstract constructs. In other terms, they define one of the most critical constructs but rarely mention the others.

4.2.1 Participant-Oriented Lexical Bundles

The participant-oriented bundles were the least commonly used functional category in both corpora. A total of four types and 424 tokens served for stance function in the MCRA-L1, having 8.89% of all bundle types and 7.37% of all bundle tokens, and seven bundles functioned as participant-oriented in the MCRA-L2, having 8.24% of all bundle types and 9.14% of all bundle tokens. The participant-oriented lexical bundles in the MCRA-L2 had two functions: stance (4 types, 4.71%; 748 tokens, 3.76%) and engagement (3 types, 3.53%; 1068, 5.37%). The limited number of stance bundles might be related to two different reasons. First, authors might express stance and engagement in other ways than four-word lexical bundles (Biber, 2006; Hyland, 2005). Secondly, novice authors might avoid using stance bundles to argue their claims (Pérez-Llantada, 2014).

Briefly, stance refers to “the ways writers present themselves and convey their judgments, opinions, and commitments” with the saying of Hyland (2005, p. 176). Although non-native authors were expected to use the bundles, *it is difficult to*, and *it is easy to* in order to show the personal evaluation (Pan et al., 2016), native authors tend to use these structures in the MCRA-L1 within this context.

We came across engagement bundles just in the MCRA-L2. These bundles were structured through the anticipatory *it*-phrases, and each included the auxiliary can to engage the readers logically or mentally. The authors could have expressed their views directly without engaging readers; however, they wanted to empower their arguments with readers’ engagement. It is obvious from the limited

use of these structures in the native corpora that native authors do not prefer the anticipatory *it*-structures and passive structures together in their academic writing.

5. Conclusion and Implications

Although non-native speakers of English were expected to produce fewer (Erman, 2009; Howarth, 1998) and less various (Granger, 1998) lexical bundles, the Turkish authors used a wide range of lexical bundles in English in high frequencies, as validated by some previous studies (Güngör & Uysal, 2016; Öztürk, 2014; Pan et al., 2016; Pérez-Llantada, 2014). On the one hand, L2 authors use a considerable number of lexical bundles in their academic prose (Biber et al., 1999; Greaves & Warren, 2010) and rely on formulaicity and fixedness, especially in the research article genre (Pérez-Llantada, 2014). On the other hand, only 34 (29%) of the bundles used by the Turkish authors were the same as the native norms in the MCRA-L1, and this means that 85 (71%) of the lexical bundles in the MCRA-L2 did not occur in the native corpus. A previous study (Güngör & Uysal, 2020) attributed this divergent use to the crosslinguistic influence of their L1.

A closer look into the functions of the lexical bundles revealed that nonnative academics tend to use particular functions more significantly than their native counterparts, as confirmed by Pan et al. (2016). The Turkish authors used procedure and inferential bundles substantially more than other functional categories. However, the Turkish authors' redundant and excessive use of clausal and passive structures seems to deviate from the norms of native academic writing. In the divergent procedure and inferential bundles together, 41.17% of the types and 42.39% of the tokens were not shared with native writing. This result underlines the need to instruct postgraduate students and researchers, especially in how native speakers present their methodology and discuss their results.

Obviously, the results of this study contribute to the consensus that academic writing is phrasal rather than clausal (see Adel & Erman, 2012; Biber & Conrad, 2019; Biber et al., 2004; Biber & Gray, 2010, 2011; Biber et al., 2016; Biber

et al., 1999; Byrd & Coxhead, 2010; Chen & Baker, 2010) and its reliance on the phrasal features are growing day by day (Parkinson & Musgrave, 2014). On the contrary, the statistics of divergent lexical bundles revealed that the Turkish authors heavily used (55.29%) verb-based structures, namely clausal structures rather than phrasal structures, in their research articles. In other words, in their writing, less proficient authors tend to show a more clausal style than proficient authors, as stated in the study of Parkinson & Musgrave (2014). When writers become more advanced, the relationship between syntactic complexity and L2 proficiency becomes non-linear (Ortega, 2003). Therefore, we can suggest universities establish academic writing centers and give training about the conventions of research article genre.

Since acquiring formulaic idiomaticity is critical in academic writing (Meunier & Granger, 2008; Schmitt, 2004), the primary focus of language instruction should be on the ready-made units or collocations (Nattinger, 1980) and their significance in the language learning process (Ebeling & Hasselgård, 2015; Hasselgren, 1994; Hoey, 2000; Leech, 2011; Lewis, 1993; Meunier, 2012; Wible & Tsao, 2008). Of the formulaic units, lexical bundles should be emphasized in academic writing classes because knowing common lexical bundles can be a crucial element of writing in a native-like manner (Ellis et al., 2008). Furthermore, identifying the bundles for a particular discipline was essential for novice academic writers (Hyland & Tse, 2007) so that they will be aware of the common lexical and rhetorical practices in their communities (Hyland, 2008a).

The phrasal style of L1 writing seems purposeful because it is more economical (Biber & Gray, 2010) and allows writers to write a concise paper (Halliday, 1993) and construct their writing in a nominal way (Parkinson & Musgrave, 2014). Furthermore, the journals limit the research articles' running words, which motivates researchers to write concisely. Writing a concise paper requires researchers to use phrasal expressions rather than clausal expressions (Halliday, 1993). Thus, writers concentrate on the clausal meaning by using

phrasal expressions, and they become advantageous by constructing their texts nominally (Parkinson & Musgrave, 2014). Although this writing style causes implicitness, native speakers tolerate this implicitness on the phrase level (Yli-Jokipii & Jorgensen, 2004). Considering these, we should encourage second language writers to use a phrasal style in their academic writing (Parkinson & Musgrave, 2014) and design intensive reading courses to help less proficient writers decode texts, including complex phrasal structures (Biber & Gray, 2010).

From this perspective, Biber et al. (2016, p. 9) built on the article of Scott and Tribble (2006) and emphasized the role of noun phrases for "informational communicative purposes" in academic writing. Also, the increasing complexity of noun phrases can be observed in the later phases of academic writing development (Biber et al., 2011) because fitting all information focus into noun phrases is becoming vital for EAP writers (Parkinson & Musgrave, 2014). Halliday (1989) considers this issue from the complexity of written language and the lexical density perspective and argues for the inclusion of the richness of nouns and noun phrases with postmodifiers in academic texts. However, the significance of noun phrases is mostly underestimated in traditional grammar classes, with the assumption that modifying phrases is more effortless than producing clauses (Biber & Gray, 2010). Therefore, the current study results should highlight the significant role of phrase formation and modification in the writing and grammar lectures of EAP modules. Reading courses can also supplement these with practices in analyzing the phrases inside text structures (Biber & Gray, 2010).

For the need for novel learning materials, especially for lexical bundles (Wible, 2008), the design of new materials based on the research findings might help to establish a relationship between EAP and writing pedagogy (Gilquin et al., 2007). The divergent uses might be beneficial for offering solutions to students' divergent or problematic writing (Tribble, 2002), and this, in turn, is expected to help less proficient writers and language learners use lexical bundles appropriately (Cortes, 2015) or in a native-like manner (Ebeling & Hasselgård, 2015). Writing

instructors should mainly focus on formulaic sequences and collocations used in the academic prose by the authors of a specific genre (e.g., dissertations, research articles, and so on) or discipline (e.g., educational sciences, engineering, or, more specifically, English language teaching). Thus, learners become able to learn formulaic sequences first, as suggested by some researchers (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004; Weinert, 1995), and, at the same time, they master the rhetorical practices specific to their communities (Hyland, 2008a). Lastly, this kind of corpus-informed study might lead to projects such as Academic Phrasebank (https://www.phrasebank.manchester.ac.uk/). Another idea might be guiding and encouraging postgraduate students to work on L1 and L2 English corpora to notice similarities and differences.

6. About the Author

Fatih Güngör is an associate professor at the Foreign Languages Education Department, Faculty of Education, Afyon Kocatepe University, Türkiye. He received his master's degree from Pamukkale University, Turkey, and a Ph.D. from Gazi University in English Language Teaching. He mainly focuses on corpus linguistics, academic writing, technology in language learning and teaching, and teacher education.

Hacer Hande Uysal is currently a full professor at the Foreign Languages Education Department, Faculty of Education, Hacettepe University, Turkey. She received her master's degree in English education and her Ph.D. in foreign language/ESL education from the University of Iowa. Her research interests are second language writing, language planning, and teacher education.

7. Acknowledgement

This study is a part of the first author's doctoral dissertation, and we thank The Scientific and Technological Research Council of Turkey for sponsoring the research stay of the first author at the University of Exeter. Huge thanks also go to the editor-in-chief, Assoc. Prof. Dr. Punchalee Wasanasomsithi, the editor, Asst.

Prof. Boonyakorn Siengsanoh, and the two reviewers who helped us strengthen the content and methodology of this study.

8. References

Ädel, A., & Erman, A. (2012). Recurrent word combinations in academic writing by native and non-native 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

Bal, B. (2010). *Analysis of four-word lexical bundles in published research articles written by Turkish scholars* [Unpublished master's thesis]. Georgia State University.

Bestgen, Y. & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing, 26*, 28–41. https://doi.org/10.1016/j.jslw.2014.09.004

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers (Vol. 23)*. John Benjamins. https://doi.org/10.1075/scl.23

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. (2019). *Register, genre, and style*. Cambridge University Press. https://doi.org/10.1017/9781108686136

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. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics, 15*(2), 223–250. https://doi.org/10.1017/S1360674311000025

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37*(5), 639–668. https://doi.org/10.1093/applin/amu059

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Brezina, V. (2016). *Statistics in corpus linguistics*. Lecture presented at the Lancaster Summer Schools in Corpus Linguistics for Language Studies, July 12-15, Lancaster.

Brezina, V., Meyerhoff, M. (2014). Significant or random: A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics. 19*(1), 1–28. https://doi.org/10.1075/ijcl.19.1.01bre

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*, 31–64.

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

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

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

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.

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes & E. Csomay

(Eds.), *Corpus-based research in Applied Linguistics: Studies in honor of Doug Biber* (pp. 197–216). John Benjamins.

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

Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory, 9*(1), 1–38.

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics, 35*(3), 328–356.

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

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes, 30*(1), 58–72.

Ebeling, S. O., & Hasselgård, H. (2015). Phraseology in learner corpus research. To appear in S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 207–230). Cambridge University Press.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, *63*, 25–51. https://doi.org/10.1111/j.1467-9922.2012.00736.x

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.

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In R. Corrigan, E. Moravcsik, H. Quali & K. Wheatley (Eds.), *Formulaic language, Vol 2: Acquisition, loss, psychological reality, and functional explanations* (pp. 323–346). John Benjamins.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology

research articles. *Journal of English for Academic Purposes, 29*, 21–42. https://doi.org/10.1016/j.jeap.2017.09.002

Field, A. (2009). *Discovering statistics using SPSS statistics* (3rd ed.). Sage Publications Ltd.

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes, 6*(4), 319–335. https://doi.org/10.1016/j.jeap.2007.09.007

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41–61. https://doi.org/10.1075/etc.1.1.05gil

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–160). Clarendon Press.

Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 212–226). Routledge.

Güngör, F., & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching, 9*(6), 176–188. http://doi.org/10.5539/elt.v9n6p176

Güngör, F. (2019). It was found that…: Introductory it patterns by native and non-native authors. *Eurasian Journal of Applied Linguistics, 5*(3), 473–493. http://doi.org/10.32601/ejal.651341

Güngör, F., & Uysal, H. H. (2020). Lexical bundle use and crosslinguistic influence in academic texts. *Lingua, 242,* 102859. https://doi.org/10.1016/j.lingua.2020.102859

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford University Press.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education, 5*(2), 93–116. https://doi.org/10.1016/0898-5898(93)90026-7

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International*

Journal of Applied Linguistics, 4(2), 237–258. https://doi.org/10.1111/j.1473-4192.1994.tb00065.x

Hewings, M., & Hewings, A. (2002). "It is interesting to note that…": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes, 21*(4), 367–383.

Hoey, M. (2000). A world beyond collocation: New perspectives on vocabulary teaching. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 224–243). Language Teaching Publications.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24–44. https://doi.org/10.1093/applin/19.1.24

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics, 30*(4), 437–455.

Hyland, K. (2000). *Disciplinary discourses: Social interaction in academic writing.* Longman.

Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115–130). Routledge.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192.

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

Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum.

Hyland, K (2012). English for academic purposes and discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge handbook of discourse analysis* (pp. 412–423). Routledge.

Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly, 41*(2), 235–253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Leech, G. (2011). Frequency, corpora and language learning. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 7–32). John Benjamins.

Leńko-Szymańska, A. (2014). The acquisition of formulaic language by EFL learners: A cross-sectional and cross-linguistic perspective. *International Journal of Corpus Linguistics, 19*(2), 225–251. https://doi.org/10.1075/ijcl.19.2.04len

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.

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.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics, 32*, 111–129. https://doi.org/10.1017/S0267190512000128

Meunier, F., & Granger, S. (2008). *Phraseology in foreign language learning and teaching*. John Benjamins. https://doi.org/10.1075/z.138

Nattinger, J. R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly, 14*(3), 337–344.

Nattinger, J. R., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Neff, J. (2008). Contrasting English-Spanish interpersonal discourse phrases: A corpus study. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching (*pp. 85–99). John Benjamins. https://doi.org/10.1075/z.138.08aer

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Öztürk, Y. (2014). *Lexical bundle use of Turkish and native English writers: A corpus-based study* [Unpublished master's thesis]. Anadolu University.

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*(2), 60–71.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes, 14*, 48–59. https://doi.org/10.1016/j.jeap.2013.12.001

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

Peromingo, J. P. R. (2012). Corpus analysis and phraseology: Transfer of multi-word units. *Linguistics and the Human Sciences, 6*(1-3), 321–343. https://doi.org/10.1558/lhs.v6i1-3.321

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42,* 220–231. https://doi.org/10.1016/j.system.2013.12.003

Römer, U. (2009a). The inseparability of lexis and grammar. *Annual Review of Cognitive Linguistics, 7*(1), 140–162.

Römer, U. (2009b). English in academia: Does nativeness matter. Anglistik: *International Journal of English Studies, 20*(2), 89–100.

Salazar, D. (2010). Lexical bundles in Philippine and British scientific English. *Philippine Journal of Linguistics, 41*, 94–109.

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

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing and use*. John Benjamins.

Schmitt, N. (2005). Processing advantages of formulaic sequences vs nonformulaic sequences in text passages. In C. Cosme, C. Gouverneur, F. Meunier & M. Paquot (Eds.), *Proceedings of the Phraseology 2005 conference* (pp. 381–382). Universite Catholique de Louvain: Louvain-la-Neuve, 13–15 October 2005.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). John Benjamins.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins. https://doi.org/10.1075/scl.22

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214–225. https://doi.org/10.1016/j.jeap.2013.05.002

Stubbs, M. (2007). Quantitative data on multi-word sequences in English: The case of the word world. In M. Hoey, M. Mahlberg, M. Stubbs & W. Teubert (Eds.), *Text, discourse and corpora: Theory and analysis* (pp. 163–189). Continuum.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131–149). Routledge.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics, 16*(2), 180–205. https://doi.org/10.1093/applin/16.2.180

Wible, D. (2008). Multiword expressions and the digital turn. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 163–181). John Benjamins.

Wible, D., & Tsao, N. L. (2011). Towards a new generation of corpus-derived lexical resources for language learning. In F. Meunier, S. De Cock, G.

Gilquin, & M. Paquot (Eds.), *A Taste for Corpora: In honour of Sylviane Granger* (pp. 237–256). John Benjamins.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Yli-Jokipii, H., & Jorgensen, P. E. F. (2004). Academic journalese for the Internet: A study of native English-speaking editors' changes to texts written by Danish and Finnish professionals. *Journal of English for Academic Purposes, 3*(4), 341–359. https://doi.org/10.1016/j.jeap.2004.07.006

Appendix 1

Convergent Lexical Bundles

Lexical Bundles	f in L1	%	f in L2	%	Structure	Function
as a result of	76	27.88	236	54.37	Prepositional phrase with embedded of-phrase fragment	Causative (TO)
as well as the	56	27.88	66	20.39	Other expressions	Additive (TO)
at the beginning of	42	16.97	72	17.96	Prepositional phrase with embedded of-phrase fragment	Location (RO)
at the end of	116	35.15	160	38.83	Prepositional phrase with embedded of-phrase fragment	Location (RO)
at the same time	62	27.88	67	23.79	Other prepositional phrases (fragment)	Additive (TO)
for the development of	27	11.52	27	10.68	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
for the purposes of	26	13.94	26	10.68	Prepositional phrase with embedded of-phrase fragment	Objective (TO)
in addition to the	39	18.18	30	12.14	Other prepositional phrases (fragment)	Additive (TO)
in relation to the	51	18.18	31	10.19	Other prepositional phrases (fragment)	Framing (TO)
in terms of the	63	25.45	126	28.64	Prepositional phrase with embedded of-phrase fragment	Framing (TO)

Lexical Bundles	f in L1	%	f in L2	%	Structure	Function
in the context of	87	26.67	34	11.17	Prepositional phrase with embedded of-phrase fragment	Framing (TO)
in the field of	29	11.52	72	20.87	Prepositional phrase with embedded of-phrase fragment	Location (RO)
in the form of	45	22.42	41	14.08	Prepositional phrase with embedded of-phrase fragment	Framing (TO)
in the process of	23	10.91	69	17.96	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
it is important to	77	31.52	52	18.45	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
on the basis of	29	15.15	48	16.5	Prepositional phrase with embedded of-phrase fragment	Framing (TO)
on the other hand	62	24.24	331	62.14	Other prepositional phrases (fragment)	Additive (TO)
one of the most	22	12.12	108	32.04	Noun phrase with of-phrase fragment	Grouping (RO)
that there is a	35	20	138	35.92	(Verb phrase/noun phrase +) that-clause fragment	Inferential (TO)
the analysis of the	20	10.3	63	22.82	Noun phrase with of-phrase fragment	Procedure (RO)
the beginning of the	40	16.36	64	16.99	Noun phrase with of-phrase fragment	Location (RO)
the content of the	25	12.12	31	11.65	Noun phrase with of-phrase fragment	Description (RO)
the end of the	114	30.91	140	37.38	Noun phrase with of-phrase fragment	Location (RO)
the extent to which	112	30.91	48	15.53	Noun phrase with other post-modifier fragments	Quantification (RO)
the fact that the	22	10.91	48	15.53	(Verb phrase/noun phrase +) that-clause fragment	Stance (PO)
the importance of the	31	14.55	24	10.68	Noun phrase with of-phrase fragment	Description (RO)

Lexical Bundles	f in L1	%	f in L2	%	Structure	Function
the majority of the	22	11.52	46	14.56	Noun phrase with of-phrase fragment	Quantification (RO)
the purpose of this	45	18.79	83	28.64	Noun phrase with of-phrase fragment	Description (RO)
the quality of the	26	10.3	31	12.14	Noun phrase with of-phrase fragment	Description (RO)
the relationship between the	25	11.52	51	16.99	Noun phrase with other post-modifier fragments	Comparative (TO)
the results of the	35	14.55	225	47.09	Noun phrase with of-phrase fragment	Causative (TO)
to be able to	55	18.79	40	15.05	(Verb/adjective +) to-clause fragment	Objective (TO)
to participate in the	31	14.55	40	14.56	(Verb/adjective +) to-clause fragment	Objective (TO)
to the development of	27	13.33	28	11.17	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)

Appendix 2

Divergent Lexical Bundles in the MCRA-L1

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
the ways in which	66		35	21.21	Noun phrase with other post-modifier fragments	Description (RO)
as part of the	60		44	26.67	Prepositional phrase with embedded of-phrase fragment	Grouping (RO)
the role of the	56		36	21.82	Noun phrase with of-phrase fragment	Description (RO)
in a variety of	50		40	24.24	Prepositional phrase with embedded of-phrase fragment	Quantification (RO)
in the case of	47		26	15.76	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
the nature of the	46		32	19.39	Noun phrase with of-phrase fragment	Description (RO)
a wide range of	45		32	19.39	Noun phrase with of-phrase fragment	Quantification (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
at the time of	44		37	22.42	Prepositional phrase with embedded of-phrase fragment	Location (RO)
in the development of	43		30	18.18	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
within the context of	38		24	14.55	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
in the area of	38		24	14.55	Prepositional phrase with embedded of-phrase fragment	Location (RO)
the development of the	37		26	15.76	Noun phrase with of-phrase fragment	Procedure (RO)
the use of the	36		30	18.18	Noun phrase with of-phrase fragment	Description (RO)
over the course of	36		21	12.73	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
for each of the	35		27	16.36	Prepositional phrase with embedded of-phrase fragment	Grouping (RO)
the impact of the	33		17	10.3	Noun phrase with of-phrase fragment	Description (RO)
it is possible that	31		24	14.55	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
on the development of	31		23	13.94	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
with regard to the	31		23	13.94	Other prepositional phrases (fragment)	Framing (RO)
it is clear that	31		26	15.76	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
were more likely to	30		20	12.12	(Verb/adjective +) to-clause fragment	Stance (PO)
an understanding of the	29		24	14.55	Noun phrase with of-phrase fragment	Description (RO)
the degree to which	29		18	10.91	Noun phrase with other post-modifier fragments	Quantification (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
can be used to	28		21	12.73	(Verb/adjective +) to-clause fragment	Procedure (RO)
by the end of	28		18	10.91	Prepositional phrase with embedded of-phrase fragment	Location (RO)
there has been a	27		21	12.73	Pronoun/noun phrase + be (+…)	Inferential (TO)
the start of the	27		18	10.91	Noun phrase with of-phrase fragment	Location (RO)
to meet the needs	26		20	12.12	(Verb/adjective +) to-clause fragment	Objective (TO)
in a way that	25		20	12.12	Other prepositional phrases (fragment)	Framing (RO)
as a means of	25		19	11.52	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
on the part of	25		19	11.52	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
through the use of	24		17	10.3	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
a small number of	24		17	10.3	Noun phrase with of-phrase fragment	Quantification (RO)
on the one hand	23		17	10.3	Other prepositional phrases (fragment)	Inferential (TO)
in a number of	23		21	12.73	Prepositional phrase with embedded of-phrase fragment	Quantification (RO)
as a way to	23		17	10.3	Other prepositional phrases (fragment)	Framing (RO)
as a way of	22		17	10.3	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
at the start of	22		17	10.3	Prepositional phrase with embedded of-phrase fragment	Location (RO)
what it means to	22		18	10.91	Other expressions	Description (RO)
the focus of the	21		17	10.3	Noun phrase with of-phrase fragment	Description (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
in a range of	21		18	10.91	Prepositional phrase with embedded of-phrase fragment	Quantification (RO)
it is difficult to	20		20	12.12	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
the rest of the	20		18	10.91	Noun phrase with of-phrase fragment	Grouping (RO)
on the use of	20		17	10.3	Prepositional phrase with embedded of-phrase fragment	Procedure (RO)
in order to provide	20		19	11.52	(Verb/adjective +) to-clause fragment	Objective (TO)

Appendix 3

Divergent Lexical Bundles in the MCRA-L2

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
that there is no	39		30	14.56	(Verb phrase/noun phrase +) that-clause fragment	Inferential (RO)
that there was no	60		27	13.11	(Verb phrase/noun phrase +) that-clause fragment	Inferential (RO)
that they did not	41		23	11.17	(Verb phrase/noun phrase +) that-clause fragment	Inferential (RO)
the results showed that	32		23	11.17	(Verb phrase/noun phrase +) that-clause fragment	Inferential (RO)
in order to be	26		23	11.17	(Verb/adjective +) to-clause fragment	Objective (RO)
in order to determine	101	0.01	62	30.1	(Verb/adjective +) to-clause fragment	Objective (RO)
in order to test	23		21	10.19	(Verb/adjective +) to-clause fragment	Objective (RO)
to be used in	24		22	10.68	(Verb/adjective +) to-clause fragment	Objective (RO)
used in order to	32		23	11.17	(Verb/adjective +) to-clause fragment	Procedure (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
was found to be	132	0.01	59	28.64	(Verb/adjective +) to-clause fragment	Inferential (RO)
was used to determine	34		24	11.65	(Verb/adjective +) to-clause fragment	Procedure (RO)
as can be seen	74		36	17.48	Adverbial clause fragment	Structuring (RO)
as seen in table	42		24	11.65	Adverbial clause fragment	Structuring (RO)
it can be concluded	33		28	13.59	Anticipatory it + verb phrase/adjective phrase	Engagement (PO)
it can be said	158	0.02	64	31.07	Anticipatory it + verb phrase/adjective phrase	Engagement (PO)
it can be seen	76		40	19.42	Anticipatory it + verb phrase/adjective phrase	Engagement (PO)
it is believed that	34		24	11.65	Anticipatory it + verb phrase/adjective phrase	Generalization (RO)
it is necessary to	50		38	18.45	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
it is observed that	59		29	14.08	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it is possible to	63		38	18.45	Anticipatory it + verb phrase/adjective phrase	Stance (PO)
it is seen that	165	0.02	54	26.21	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it is thought that	42		29	14.08	Anticipatory it + verb phrase/adjective phrase	Generalization (RO)
it was concluded that	46		32	15.53	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it was determined that	118	0.01	45	21.84	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it was found that	193	0.02	65	31.55	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it was observed that	112	0.01	47	22.82	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
it was seen that	116	0.01	43	20.87	Anticipatory it + verb phrase/adjective phrase	Inferential (RO)
a high level of	37		24	11.65	Noun phrase with of-phrase fragment	Description (RO)
analysis of the data	36		26	12.62	Noun phrase with of-phrase fragment	Procedure (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
one of the important	25		21	10.19	Noun phrase with of-phrase fragment	Grouping (RO)
sample of the study	33		22	10.68	Noun phrase with of-phrase fragment	Procedure (RO)
the aim of this	65		53	25.73	Noun phrase with of-phrase fragment	Description (RO)
the characteristics of the	25		23	11.17	Noun phrase with of-phrase fragment	Description (RO)
the effect of the	40		24	11.65	Noun phrase with of-phrase fragment	Causative (RO)
the findings of the	90		51	24.76	Noun phrase with of-phrase fragment	Causative (RO)
the main purpose of	26		22	10.68	Noun phrase with of-phrase fragment	Description (RO)
the participants of the	25		21	10.19	Noun phrase with of-phrase fragment	Description (RO)
the reliability of the	70		49	23.79	Noun phrase with of-phrase fragment	Description (RO)
a significant difference between	118	0.01	46	22.33	Noun phrase with other post-modifier fragments	Comparative (TO)
a significant difference ir	61		26	12.62	Noun phrase with other post-modifier fragments	Comparative (TO)
an important role in	40		33	16.02	Noun phrase with other post-modifier fragments	Stance (PO)
no significant difference between	63		30	14.56	Noun phrase with other post-modifier fragments	Comparative (TO)
the data obtained from	43		36	17.48	Noun phrase with other post-modifier fragments	Framing (RO)
the difference between the	47		28	13.59	Noun phrase with other post-modifier fragments	Comparative (TO)
higher than those of	66		22	10.68	Other adjectival phrases	Comparative (TO)
before and after the	51		21	10.19	Other expressions	Comparative (TO)
which is one of	24		22	10.68	Other expressions	Grouping (RO)
who participated in the	48		23	11.17	Other expressions	Procedure (RO)
according to the findings	35		25	12.14	Other noun phrases	Inferential (RO)
according to the results	94		49	23.79	Other noun phrases	Inferential (RO)
significant difference was found	42		21	10.19	Other passive fragments	Procedure (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
students were asked to	38		23	11.17	Other passive fragments	Procedure (RO)
study was carried out	40		32	15.53	Other passive fragments	Procedure (RO)
the data were collected	33		27	13.11	Other passive fragments	Procedure (RO)
the study was conducted	36		33	16.02	Other passive fragments	Procedure (RO)
they were asked to	37		26	12.62	Other passive fragments	Procedure (RO)
in accordance with the	77		46	22.33	Other prepositional phrases (fragment)	Framing (RO)
in line with the	59		37	17.96	Other prepositional phrases (fragment)	Comparative (TO)
in the current study	66		34	16.5	Other prepositional phrases (fragment)	Structuring (RO)
in the present study	106	0.01	50	24.27	Other prepositional phrases (fragment)	Structuring (RO)
to the fact that	34		29	14.08	Other prepositional phrases (fragment)	Stance (PO)
with respect to the	44		26	12.62	Other prepositional phrases (fragment)	Framing (RO)
participated in the study	74		42	20.39	Other verbal fragments	Procedure (RO)
this study aims to	43		31	15.05	Other verbal fragments	Objective (RO)
are given in table	70		26	12.62	Passive verb + prepositional phrase fragment	Structuring (RO)
are presented in table	129	0.01	53	25.73	Passive verb + prepositional phrase fragment	Structuring (RO)
are shown in table	56		29	14.08	Passive verb + prepositional phrase fragment	Structuring (RO)
can be defined as	30		25	12.14	Passive verb + prepositional phrase fragment	Generalization (RO)
included in the study	29		25	12.14	Passive verb + prepositional phrase fragment	Procedure (RO)
is based on the	23		21	10.19	Passive verb + prepositional phrase fragment	Framing (RO)

Lexical Bundles	f in L1	%	Texts	%	Structure	Function
used in this study	40		35	16.99	Passive verb + prepositional phrase fragment	Procedure (RO)
was carried out with	30		25	12.14	Passive verb + prepositional phrase fragment	Procedure (RO)
was used as a	26		23	11.17	Passive verb + prepositional phrase fragment	Procedure (RO)
was used in the	35		27	13.11	Passive verb + prepositional phrase fragment	Procedure (RO)
were included in the	34		26	12.62	Passive verb + prepositional phrase fragment	Procedure (RO)
were informed about the	26		24	11.65	Passive verb + prepositional phrase fragment	Procedure (RO)
as a part of	27		22	10.68	Prepositional phrase with embedded of-phrase fragment	Grouping (RO)
as one of the	54		36	17.48	Prepositional phrase with embedded of-phrase fragment	Grouping (RO)
in the light of	44		38	18.45	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
with the help of	44		29	14.08	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
with the results of	35		30	14.56	Prepositional phrase with embedded of-phrase fragment	Comparative (TO)
within the framework of	33		24	11.65	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
within the scope of	92		35	16.99	Prepositional phrase with embedded of-phrase fragment	Framing (RO)
there was a significant	89		34	16.5	Pronoun/noun phrase + be (+...)	Inferential (RO)
there was no significant	72		23	11.17	Pronoun/noun phrase + be (+...)	Inferential (RO)