

## LEXICAL BUNDLES IN L1 AND L2 ENGLISH DOCTORAL DISSERTATIONS

Ilyas Yakut<sup>1</sup>, Fatma Yuvayapan<sup>2</sup>, Erdogan Bada<sup>3</sup>

<sup>1</sup>Zonguldak Bulent Ecevit University, Turkey

<sup>2</sup>Kahramanmaraş Istiklal University, Turkey

<sup>3</sup>Çukurova University & Hakkari University, Turkey

**Abstract.** *Based on contrastive interlanguage analysis, this study explores the usage of lexical bundles occurring in doctoral dissertations produced in the English language related studies in the USA by L1 American English speakers and in Turkey by Turkish speakers of English in the last ten years between 2010-2019. In our analysis of the data, we identified a significant number of types of 4-word bundles from both corpora from a structural and functional perspective. The findings regarding the types of lexical bundles and their structural and functional dispersions revealed significant differences between the two corpora. While L1 English writers refrained from heavily utilizing formulaic sequences, the opposite could be observed in the works of L2 English authors. This study has significant implications for academic writers producing work in the English language since corpus-informed lists and concordances might be of great help to L2 speakers of English in identifying appropriate lexical bundles that are specific to their disciplines.*

**Key words:** *lexical bundles, doctoral dissertations, functional classification, structural classification*

### 1. INTRODUCTION

Constructed from culturally and disciplinary conventions, academic writing is different from other writing registers. It is not merely a presentation of propositional content but possesses many pragmatic functions like creating writer-reader engagement, building authorial stance, and most prominently, it is a key for academics to get a credible place in an academic community. Hyland (2002) states that academic writing is rested on practices and structures of social communities and conveys a representation of the writer reflected by the discursive choices to convince readers about the truth of propositional content and arguments.

Murray & Moore (2006) explain that what and how we write shapes our career development as an academic writer. It is this realization which puts academic writing at the core of academic performance and success. It is a socially-constructed means of sharing ideas under the framework of expected conventions in a shared academic context. Similarly, Burke (2010) contends that academic writing is a social process in which writers,

---

Submitted February 8<sup>th</sup>, 2021, accepted for publication March 18<sup>th</sup>, 2021

*Corresponding author:* Ilyas Yakut. Zonguldak Bulent Ecevit University, Department of English Language and Literature, Turkey | E-mail: [ilyasyakut@beun.edu.tr](mailto:ilyasyakut@beun.edu.tr)

by making institutionally motivated rhetorical choices in a particular discipline, convey meaning, marshal arguments and reach agreements with their readers through drawing on strategies at an interpersonal level.

It is undeniable that English, as the lingua-franca of a global academic world, is the key to gaining credibility in academia (Genç & Bada, 2010) since a number of articles are published, presentations in international conferences are performed, and in recent years projects with the contribution of many stakeholders from different countries are conducted in this global language. In all of these academic activities, academic writers employ disciplinary and culturally-situated linguistic norms to meet the expectations of a worldwide audience. Among these linguistic norms, lexical bundles (LBs), i.e., formulaic sequences of words, are beneficial in creating coherence in an academic discourse and mitigating authorial stance.

Corpus linguistics is one of the fields that investigates LBs from a variety of perspectives such as forms, structures and functions (Cortes, 2004; Biber, Conrad & Cortes, 2004; Hyland, 2008a) historical changes (Hyland & Jiang, 2018); and cultural and disciplinary variations (Lu & Deng, 2019; Muşlu, 2018; Karabacak & Qin, 2013; Dontcheva-Navrotilova, 2012). The results of these studies that are based on the analysis of a large amount of language data have improved our understanding of formulaic sequences.

Most studies in the literature have also focused on different academic genres: research articles (Güngör & Uysal, 2016; Hyland, 2008a); argumentative essays (Karabacak & Qin, 2013); BA theses (Dontcheva-Navratilova, 2012); acknowledgment in research articles (Demirel & Hesamoddin, 2013); research articles, MA and PhD theses (Hyland, 2008a); literature reviews (Wright, 2019). However, no previous study has dealt with the analysis of LBs regarding their structures and functions in a corpus of PhD dissertations written by L1 and L2 English academic writers.

In this study, we intend to examine the structures and functions of LBs in a corpus of doctoral dissertations written by L1 and L2 English academic authors. As Hyland (2008a) states, all research genres reflect similar use of academic register, but students' genres have some challenges and constraints. MA and PhD theses are high stakes genres that students have to deal with in their university lives. The credibility of these genres improves students' chances of succeeding in their academic community. Yet, it is a difficult process that they face while they endeavor to comply with the appropriate linguistic norms of their disciplines. To this end, the following research questions constituted the essence of this study:

1. What are the frequently emerging LBs in PhD dissertations written by L1 English writers and Turkish-speaking academic writers of English regarding their structures and functions?
2. Do L1 English writers and Turkish-speaking academic writers of English significantly differ in their use of LBs in their PhD dissertations regarding their structures and functions?
3. What are the common LBs utilized by L1 English writers and Turkish-speaking academic writers of English in PhD dissertations?

## 2. RESEARCH IN LEXICAL BUNDLES

One of the most striking features of academic written texts is the employment of LBs. Biber et al. (1999) define LBs as “recurrent expressions, regardless of their idiomaticity, and regardless of their structural status” (p. 990). They can be formed from two or more

words and must frequently recur to represent a typical LB. They are also widely used by different speakers in different situations.

Hyland (2008a) identifies them as structural units of words to establish coherence in a text, and their identification relies mainly on the frequency of counts. The minimal cut-off set for recurrent sets of words to be conceived as LBs is at least ten times per million words. The use of 4-word bundles is more common in research since they present a wider range of structures. It seems, therefore, that the criteria for the cut-off sets depend on the scope of the study. According to Biber (2006), general characteristics of LBs could be explained as;

- A frequency-driven approach is essential to identify LBs. The occurring sets of words that meet the frequency criteria qualify as LBs, and they are extracted from a corpus using computer software.
- They are incomplete grammatical structures occurring frequently. They connect two structural units; usually, the last of a bundle is followed by the first element of another structure. Although having unidiomatic meaning, they reflect particular pragmatic and discoursal functions.
- The criteria applied for the frequency cut-off used to label them is not definite. It can be employed differently, depending on the number of words to be analyzed.
- To be identified as an LB, a multi-word sequence must occur in at least five different texts.

Hyland (2008b) states that bundles are beneficial in organizing meaning in texts and facilitating distinctiveness in a register and participation in a community. To maintain a new language or a register, novices have to be aware of the expert's preferences of LBs. In other words, disciplines rely upon some fixed phrases, and it is necessary for L2 English academic writers to learn the way they are used to gain communicative and linguistic competence in their academic disciplines. Similarly, Hyland & Jiang (2018) emphasize that LBs contribute to efficient communication in terms of pragmatic issues, and do also facilitate more specific purposes in academic texts. The usage of familiar recurrent structures helps readers to understand the texts in a short time by guiding them to particular propositions. They are also disciplinary-based linguistic features that create engagement of readers and construal of author stance.

Studies on characteristics of LBs have led to the emergence of some categories regarding the classification of LBs grounded in their structural units and functions. For instance, Cortes (2004) suggested a structural categorization of LBs: noun phrase with 'of' phrase fragment, noun-phrase with post-nominal clause fragment, prepositional phrase with embedded 'of' phrase, other prepositional fragments, verb (be)+ complement (noun phrase), other expressions. According to their functions performed in context, LBs were grouped into two categories: referential bundles and text organizers. In the same year, Biber et al. (2004) developed a more detailed categorization of LBs, which distinguishes among three main structural categories: LBs that incorporate verb phrase fragments, LBs that incorporate dependent clause fragments, LBs that incorporate noun phrase and prepositional phrase fragments. Functional aspects were divided into three categories: stance expressions, discourse organizers, and referential expressions. As revealed by Biber et al. (2004: 384), stance bundles display attitudes or assessments of propositional certainty, and discourse organizers state the relationship between prior and upcoming discourse. On the other hand, referential bundles suggest direct referencing to physical or abstract entities, or to textual contexts.

A modification of a framework was proposed by Hyland (2008a) with more or less the same content yet with differing terminology. Upon completing the analysis of research articles and MA and PhD theses, Hyland proposed three broad functional categories of LBs: research, text, and participant-oriented. Research-oriented bundles "help writers to structure their activities and experiences of the real world." Text oriented "clusters are concerned with the organization of the text and the meaning of its elements as a message or argument," and participant-oriented bundles "focused on the writer or reader of the text" (p. 49).

With the advances in corpus linguistics, researchers have focused on the continuous sequence of frequently recurring words to have a better understanding of these formulaic units. The frequency-driven approach has inspired new and various studies (Salazar, 2014). Similarities and differences between written and spoken genres have been examined in some studies; a pioneering one being made by Biber et al. (1999) introduced the concept of LBs for the first time in literature. The researchers analyzed large corpora, including both written and spoken academic registers in terms of frequency counts. Following a frequency-driven approach, Biber et al. (2004) later examined LBs in two registers: textbooks and classroom teaching. They indicated that LBs reflected important functions on the construction of discourse; they guide readers/listeners to interpret a text in terms of stance, discourse organization, and referential status, and signal well-defined structural forms. In a corpus consisting of written and spoken activities related to university life, Biber & Barbieri (2007) found that the usage of bundles is closely associated with writers' or speakers' communicative purposes. The reasons for the less frequent employment of bundles in written registers are that these particular registers have limited communicative purposes compared to spoken ones. While LBs tend to appear at high frequencies in sub-corpora of service-encounters and class management in the corpus of spoken registers, in written registers, however, they are utilized most frequently in written course management. Regarding their functions, each register was rested on different functions. Stance bundles were dominant in spoken university registers, while referential functions were more common in written registers.

Some researchers have attempted to investigate LBs in a pedagogical framework. Byrd & Coxhead (2010) examined the most frequently occurring LBs in the corpus of Academic Word List developed by Coxhead in 2000. This corpus includes various genres, including articles, book chapters, course workbooks, laboratory manuals, and course notes. Cortes (2004) analyzed the functions of LBs in two academic prose written by published authors and students in two disciplines. After identifying the most frequent LBs used by published authors, they examined them in the writing of students and found that these bundles did not exist in students' writings. Neely & Cortes (2009) concentrated on LBs in academic lectures and suggested that EFL teachers should utilize corpus-based research tools to aid students in discovering their functions.

A considerable amount of literature has been published on the employment of LBs in L2 academic writing. Hyland (2008b) analyzed the forms, structures, and functions of 4-word bundles in corpora consisting of research articles, MA theses, and PhD dissertations written by Cantonese writers of English across four disciplines and observed disciplinary variations in the employment of LBs in the corpora. Writers in different fields marshal their arguments, gain credibility, and communicate with their readers through the use of different bundles. In another study, Hyland (2008a) compared the use of LBs in the same corpus, revealing considerable variations in terms of structures and functions of bundles in three genres: published articles, dissertations, and student writings. The preferences of writers were shaped to improve their arguments, persuade their readers and build their stance.

In a recent study, Shin et al. (2018) specifically focused on definite articles in LBs in academic writing produced by L2. The analysis of this learner corpus showed a number of the- errors in bundles. Presenting a problematic area in the formation of LBs by L2 English writers, they recommended the implementation of article-related instruction in the teaching of LBs. Shin (2019) compared L1 and L2 English novice academic writing in terms of utilization of frequency of LBs, and observed the existence of many common functions, especially stance expression and referential LBs.

Some studies were also conducted by Turkish academic writers. Nesi & Basturkmen (2006) investigated LBs in a corpus of 160 university lectures. The corpus consisted of lectures from the Corpus of British Academic Spoken English (BASE) and Michigan Corpus of Academic Spoken Corpus (MICASE). The researchers stated that LBs played a pivotal role in discourse, suggesting that language learners should be aware while using them. It is likely that L2 English writers may not have a clear understanding of the functions of bundles. A study conducted by Muşlu (2018) showed that while there was a heavy reliance on LBs by Japanese and Turkish writers, L1 English writers did not have a tendency towards employing such bundles frequently in their argumentative essays. Karabacak & Qin (2013) compared the argumentative essays written by Turkish, Chinese, and American students with an academic writing corpus taken from the New York Times magazine. Five-item bundles were found to be employed less by American students compared to Turkish students, and the Chinese used them the least. LBs used by L1 English speakers were different from those employed by Turkish and Chinese writers of English. Güngör & Uysal (2016) examined the utilization of LBs in research articles and found that Turkish researchers used more LBs than L1 English writers.

Recent work has concentrated on particular sections of academic genres. Looking at forms and functions of LBs in doctoral dissertation abstracts written by L1 English and Chinese writers of English, Lu & Deng (2019) found that Chinese writers employed LBs more frequently than L1 English writers regarding structural and functional categories. Wright (2019) analyzed LBs in published literature reviews in three disciplines: psychology, education, and medicine. The results reflected a frequent utilization of referential bundles followed by stance and discourse bundles.

Few studies have been carried out to discover potential changes in the usage of LBs. Most prominently, we see Hyland & Jiang's (2018) study highlighting that LBs signal the argument patterns and the purposes of writers and readers. In this study, the researchers tracked changes in terms of frequency, forms, and functions of LBs in research writing in the last 50 years. They analyzed a corpus of articles in top journals in four disciplines: applied linguistics, sociology, electrical engineering, and biology, and they discovered that noun/preposition related phrases were the most common types used, yet there was a decline in this structural category. According to researchers, regarding functions, there was a shift from research-oriented to participant-oriented forms in the hard sciences. In contrast, the soft knowledge fields have moved in the opposite direction.

As can be seen from this overview, LBs, which are frequently recurring expressions functioning as structural units, have proven to be highly influential in organizing texts, persuading readers, and mitigating stance in academic writing. Lack of mastery in proper utilization of such expressions may have a significant impact on L2 English writers' credibility in a global, academic world. Since a wide range of functions are realized through LBs in academic texts, ineffective and/or erroneous utilization may lead to failure in

responding to readers' expectations. Appropriate utilization of these formulaic structures, thus, may be an area to which L2 English academic writers should give due attention.

### 3. METHODS

The study adopts a corpus-based contrastive interlanguage analysis (CIA) (Granger, 1996) approach to figure out the use, structures and functions of 4-word LBs occurring in the corpus of L1 English authors (CNAE) and the corpus of Turkish speaking academic authors of English (CTAE). To compare the use of LBs across the two corpora, we performed both descriptive and inferential statistics.

#### 3.1. Data collection

The data for this research consists of two corpora of PhD dissertations written by L1 and L2 English authors. PhD dissertations in the English language related fields written in English by L1 American English speakers studying at various universities in the USA constituted CNAE while PhD dissertations in the same fields produced in English by Turkish PhD students studying at the Turkish universities formed CTAE. Informed consent of the authors was obtained in the process of corpus compilation, and the corpora consisted of dissertations constructed through a convenience sampling technique. The data was compiled from PhD dissertations written between 2010 and 2019 in the English language related fields. In the process of corpus compilation, sections such as introduction, results and discussions, and conclusion were included into the corpora; all quotations and paraphrases were excluded to avoid potential interference of other authors cited in dissertations. CNAE consisted of 63 dissertations, totaling 1.245.681 words while CTAE comprised 64 dissertations, totaling 1.413.876 words.

#### 3.2. Identification of lexical bundles

In this study, we decided to explore the use, structures, and functions of 4-word formulaic sequences that emerged in PhD dissertations written by L1 English and Turkish L2 English speakers. As stated by Hyland (2008b), 4-word bundles "[...] are far more common than 5-word strings and offer a clearer range of structures and functions than 3-word bundles" (p. 8). In addition, the practice of previous research demonstrated that 4-word bundles were the most frequently explored units in the field due to their frequency, prevalence, and variety, which enable scholars to make further analysis (Chen & Baker, 2010). Similarly, Hyland (2012) revealed that 3-word bundles were very common and could not be handled in large corpora while 5+ bundles were relatively rare, and they generally encapsulated shorter bundles. Since 4-word bundles "offer a wider variety of structures and functions to analyze" (Hyland, 2012: 151) and "are more phrasal in nature and correspondingly less common" (Biber et al., 1999: 992), the 4-word bundles occurring in CNAE and CTAE were analyzed in the current study.

LBs are identified through two criteria: frequency and range. The methodology of identifying the formulaic sequences is not under debate in the field (Hyland & Jiang, 2018), whereas frequency and dispersion thresholds are arbitrary. For this reason, the frequency and range cut-off set by the scholars differ. Frequency thresholds in the field have settled on 10 (Biber et al., 1999; Biber, 2006), 20 (Cortes, 2004; Hyland, 2008a; 2008b;

Lu & Deng, 2019) or 40 (Biber & Barbieri, 2007; Pan, et al., 2016) times of occurrences per million words. In addition, there have been studies taking the raw frequency counts as the baseline of bundle identification (Chen & Baker, 2010). Even though the threshold frequency shows differences across studies, the purpose of setting a frequency cut-off “is to identify bundles that recur often enough to be regarded as typical of the target” (Pan et al., 2016: 63).

Range thresholds, the second identification criterion, have to be set to make sure that a specific multi-word sequence is used in a variety of texts which “guard[s] against idiosyncratic uses by individual speakers or authors” (Biber & Barbieri, 2007: 268). Similar to the frequency thresholds, the cut-off breadth of use is also somewhat arbitrary. While some of the studies set raw frequency such as three to five texts as the threshold range (Biber et al., 1999; Biber et al., 2004; Biber & Barbieri, 2007, Pan et al., 2016), Hyland (2008a, b) highlighted the importance of setting the threshold in proportion to the number of texts construing the corpus. For this reason, they suggested including word strings emerging at least 10% of the texts in the corpora.

Regarding the frequency and range thresholds, we decided to take a conservative approach by following the suggestions made by Hyland (2008a, 2008b, 2012), and we set the frequency cut-off of 20 times per million words and included only the 4-word LBs used in at least 10% of texts. Using Ant Conc version 3.5.8 (Anthony, 2019), the bundle lists for the texts in the two sub-corpora were generated by applying the aforementioned criteria. Through concordance, the overlapping LBs constituting “complete overlap” (Chen & Baker, 2010: 33), where two or more 4-word bundles are part of 5-word formulaic sequence, were manually removed as they might inflate the quantitative results (Chen & Baker, 2010; Bychkovska & Lee, 2017).

### 3.3. Annotation of the data

The refined data was annotated structurally regarding grammatical features, and functionally in terms of discourse functions in the text. Structural tagging was performed by one researcher through a vertical reading of the data. Each bundle was tagged regarding its part of speech based on the structural categories generated by Biber et al. (1999) and updated by Hyland & Jiang (2018). While the former taxonomy consisted of 12 categories of academic bundles, the latter identified three distinct main categories with various sub-categories. In our study, we adapted the abovementioned classifications by making some amendments. Similar to Hyland and Jiang’s structural categories, we divided the structures into verb phrase-related bundles (VP), clause-related bundles, noun phrase-related bundles (NP), and prepositional phrase-related bundles (PP). Even though we benefited from Hyland and Jiang’s structural categories, it should be noted that the scholars merged noun and prepositional phrase-related bundles into the category of noun/preposition-related bundles, and this was criticized by Lu & Deng (2019) pointing out that combining bundles with different phrases (e.g., noun phrase and prepositional phrase based bundles) might “obscure important distribution patterns” (p. 25). Therefore, we added the category of others, which includes structures different from the abovementioned categories.

As for functional annotation, we adopted the functional taxonomy suggested by Hyland (2008a, b) since their corpora also consisted of texts compiled from academic writings. According to Hyland’s (2008b) framework, the functions of LBs are organized around three main functional categories, which are research-oriented, text-oriented, and

participant-oriented, and each category is divided into various sub-categories grouping more specific roles of word strings.

Although structural tagging was not ambiguous, and it could be performed through part of speech tagging, functional annotation of the data was relatively subjective because a specific LB could have multiple functions. In this study, the primary function of a bundle at hand was determined through an integrated reading of the data (both vertical and horizontal reading) to figure out how the bundle was employed in the text.

For functional annotation, two researchers tagged 10% of the data independently for an inter-rater reliability test before annotating the whole data. According to Loewen & Plonsky (2016), an inter-rater reliability test is "[a] method of dual coding of data to ensure that the coding categories or scores are being used in a consistent manner" (p. 93). The reliability test revealed a high correlation between the raters since the p-value was above 0.80 ( $p=0.84$ ). For the rest of the functional annotation, the three researchers were in constant interaction.

### 3.4. Statistical measures

The principal aim of this study is to reveal overall usages of 4-word LBs by L1 and L2 English writers in doctoral dissertations. Hence, the study is a corpus-based study, and it adopts CIA to reach its abovementioned aim. To reach this goal, we performed descriptive statistics to set out overall usage rates of LBs along with distributions of structural and functional categories of the 4-word LBs in both corpora. In addition, the log-likelihood (LL) ratio of raw frequencies of the 4-word bundles was computed through Rayson's spreadsheet calculator at <http://ucrel.lancs.ac.uk/llwizard.html> to compare the overall distribution of bundles, their structures and functions across the two corpora.

## 4. FINDINGS AND DISCUSSION

This study mainly concentrated on the comparison of LBs occurring in doctoral dissertations written by L1 and L2 English writers. In line with our aim, the data is presented in tabular form, and analysis and commentary were made related to each cited table. In the tables, L1 stands for L1 English writers' dissertations included in the Corpus of Native Academic Authors of English; L2, for L2 English writers' dissertations included in the Corpus of Turkish Academic Authors of English; N, for the raw frequency of LBs; N/10000, for the number of LB occurrences per 10000 words; %, for percentages of LB usages; and LL, for log-likelihood in the corpora.

Table 1 compares the overall distribution of LBs in two corpora. L1 consisted of 1.245.681 words, while L2 included 1.413.876 words. We identified 103 and 275 types of LBs in L1 and L2, respectively, which demonstrates that there might be a relationship between author's L1 background and academic writing conventions employed in the disciplines related to the English language. This fact was also supported by the normalized frequency counts of 37.24 per 10.000 words in L1 and 119.22 per 10.000 words in L2. Type/token ratio of 0.008 and 0.020 in L1 and L2 allowed us to state that LBs were less varied in the product of L1 compared to L2. The results were consistent with Öztürk (2014), who found that LBs were employed more frequently by Turkish researchers against L1. However, Muşlu (2018) found that L1 English corpus was more varied than L2 English corpus in terms of LB usage.



Table 1 The overall distribution of LBs in the two corpora

	L1	L2
Corpus size in words	1245681	1413876
Lexical Bundle type	103	275
N /10000*	0.83	1.95
N	4639	16856
N/10000**	37.24	119.22
T/T ratio***	0.008	0.020

\* lexical bundle types per 10.000 words

\*\* frequency of lexical bundles per 10.000 words

\*\*\* type/token ratio: percentage of number of lexical bundles (types) in total words in each corpus

We conducted LL statistics to calculate whether the difference of frequency counts in the two corpora was statistically significant. As shown in Table 2, while L1 did not heavily rely on formulaic sequences in their academic writing, L2 resorted to such forms quite frequently. For this reason, there was a statistically significant LB usage difference between L1 and L2 in terms of types (-60.95) and frequency counts (-5915.22) against. Hence, it could be stated that there was a substantial reliance on the use of LBs in doctoral dissertations written by L2. Hyland (2012) states that postgraduate genres are more phrasal than published ones. It may be due to the fact that Turkish postgraduate students tend to distance themselves from their dissertations and assert credibility for readers by using a large number of LBs.

Table 2 LL ratio of LBs in the two corpora regarding types and raw frequencies

	L1		L2		LL Ratio (p<0.05)	ELL
	(O1)	1%	(O2)	2%		
LB type	103	0.01	275	0.02	- 60.95	0.00000
LB N	4639	0.37	16856	1.19	- 5915.22	0.00024

Table 3 compares the distribution of LBs in two corpora and shows LL ratio values regarding their structural categories. A quick glance at the table shows that type and frequency counts of all major and sub-categories of LBs were significantly lower among L1 compared to L2. Additionally, both L1 and L2 had different tendencies in the utilization of LBs regarding main categories and sub-categories, both groups employing a great number of NP and PP in their doctoral dissertations.

While PP utilization was at the top in L1, accounting for half of all LBs with a frequency of 22.09 per 10.000 words and 55 types, it was the second most common structural category of LBs in L2 with 83 types and 39.90 times of occurrences per 10.000 words. PP with embedded of-phrase was observed as the most frequented sub-category of this structure, which was followed by other PP. Biber et al. (1999) state that of-phrases are common features of written and spoken registers. Although it was the most preferred structure of LBs in L1, the LL ratio of -682.81 indicated a significant overuse of this structure by L2. It can be inferred from (S1) and (S2) that this type of structure is a sign of logical relations between two claims or propositional elements.

Table 3 The structural distribution of LBs in the two corpora

Major structure	Sub-Structure	Type				Frequency				LL Ratio (p<0.05)
		L1		L2		L1		L2		
		N	%	N	%	N	N/10000	N	N/10000	
Noun-phrase	with embedded of-phrase	13	12.62	67	24.36	489	3.93	3964	28.04	-2668.30
	other noun phrase	8	7.77	26	9.45	405	3.25	1884	13.33	-858.48
	Sub-total	21	20.39	93	33.82	894	7.18	5848	41.36	-3469.66
Prepositional-phrase	with embedded of-phrase	21	20.39	41	14.91	1162	9.33	2756	19.49	-481.51
	other prepositional phrase	31	30.10	41	14.91	1423	11.42	2751	19.46	-278.48
	comparative expressions	3	2.91	1	0.36	167	1.34	134	0.95	+9.01
	Sub-total	55	53.40	83	30.18	2752	22.09	5641	39.90	-682.81
Verb-phrase	Copula be + noun/adjective phrase	5	4.85	3	1.09	183	1.47	188	1.33	+0.92
	imperative + verb phrase	2	1.94	-	-	56	0.45	-	-	-
	Verb + to clause fragment	1	0.97	14	5.09	63	0.51	617	4.36	-455.52
	with passive verb	2	1.94	21	7.64	71	0.57	953	6.74	-796.03
	Sub-total	10	9.71	38	13.82	373	2.99	1758	12.43	-810.69
Clause-related	abstract subject	1	0.97	6	2.18	27	0.22	205	1.45	-133.13
	Anticipatory it	5	4.85	20	7.27	231	1.85	1289	9.12	-683.86
	as + fragments	4	3.88	9	3.27	127	1.02	758	5.36	-422.55
	human subject	1	0.97	5	1.82	38	0.31	314	2.22	-213.51
	if + fragments	-	-	2	0.73	-	-	58	0.41	-
	that + fragments	2	1.94	7	2.55	96	0.77	425	3.01	-184.82
	there+fragments	-	-	4	1.45	-	-	212	1.50	-
	wh + fragments	1	0.97	4	1.45	25	0.20	187	1.32	-120.41
Sub-total	14	13.59	57	20.73	544	4.37	3448	24.39	-2003.54	
Others	Sub-total	3	2.91	4	1.45	76	0.61	161	1.14	-21.36

(S1) In the case of American literature anthologies, these rhetorical moves include explicit mention of additions to each edition without mentioning deletions; in composition textbooks, these rhetorical moves emerge in the promotional overview of what the editions include with no mention of what the editors have opted to exclude. (L1-10)

(S2) This outcome is again in line with the related literature indicating EFL learners' positive attitudes towards corpus use in language classes. (L2-10)

Being the second most frequented category in L1 and the most frequently employed one in L2 in terms of the types and frequency counts, NP usage was represented by 21 and 93 types with a frequency of 7.18 and 41.36 per 10.000 words in L1 and L2, respectively. Clearly, L1 adhered to less varied types of NPs compared to L2 in their doctoral dissertations.

The LL ratio of -3469.66 proved the statistically significant fewer usage rates of NP category by L1 against L2. Regarding the sub-categories, LBs of NP with embedded of-phrase were seen more frequently than other NP types in both corpora. Hyland (2008b) found that NP with embedded of phrase was the most dominant structure of LBs. This finding does not collaborate with some studies in the literature. In Byrd & Coxhead's (2010) study, for instance, PP represented 38% of the LBs in the corpus, whereas NP had a percentage of 24 in this analysis. As can be seen from (S3) and (S4), the author attempted to build a bridge between two contents in (S3), and to direct readers to a particular section in the dissertation in (S4).

(S3) Unable to explain the meaning of the lyrics, he tells the Kakyaks, "it is one of the many coon songs which are so immensely popular in my country. (L1-1)

(S4) In this part of the study, the relationship between the training and students' awareness and use of SRL strategies will be discussed to address the third research question. (L2-7)

Biber & Gray (2010) emphasize that NPs are grammatical devices of structural complexity, especially common in academic writing. Liu & Li (2016) further explain that achieving structural complexity is an indication of linguistic maturity in academic literacy. An understanding of the NP centered structural units contributes to the development of this academic literacy. An overuse of NP structures in L2 allowed us to suggest that L2 did not seem to have a complete knowledge of the appropriate usage of NP. They produced a style rested on massive use of NPs to give detailed information about the propositional content in terms of quantity, quality, size, and place.

These findings were parallel to Güngör (2016) and Güngör & Uysal (2016) who found that PP was the most frequently applied structural category of LBs among L1 while NP was mostly used in L2 corpus. In studies conducted by Hyland (2008a, 2008b), it can be observed that NP and PP were essential features of academic writing, which is parallel to our findings. In the same vein, Parkinson & Musgrave (2014) identify PPs as "the most common postmodifiers in the twentieth-century academic prose" (p. 49).

Another notable difference between the two corpora was that VP was the fourth frequented category in L1 corpus represented with ten types and 2.99 frequency counts per 10.000 words. However, it was utilized in L2 with 38 types and 12.43 per 10.000 words as the third frequently applied category. In (S5), the author tried to describe a certain point in the study.

(S5) Thus, to find out whether this difference is statistically significant, a binomial test was run. (L2-3)

Among the sub-structures of VP, the most paramount difference between L1 and L2 was observed in the use of passive verbs. Along with the scarce usage rate of the other sub-categories of LBs of VP by L1, passive verbs did not constitute a common LB in L1 corpus while they were the most common sub-structure of VP LBs among L2. Due to this fact, we observed -796.03 LL value indicating that it was applied significantly more frequently by L2 against L1.

It is quite clear that the utilization of passive verbs does not only allow researchers to present the propositional content and their claims objectively but to also distance themselves from their text and leave readers alone with texts, which increases the possibility of reader's approval towards their claims as seen in (S6). According to Biber et al. (1999), passive forms present an ambiguous attribution of stance. In addition, Hyland (2008b) explains that passive bundles not only assist readers through the text but also make the basis of claims clear for readers.

(S6) This can be seen in Figure 3.5 below, which re-plots the results of the simulations seen in Figure 3.3 to show the probability of a coda cluster being realized as a cluster. (L2-12)

As for clause-related category, we found 14 and 57 types with 4.17 and 24.39 frequency counts per 10.000 words in L1 and L2 corpora respectively, accounting for -2003.54 LL value, which revealed a rather statistically significant difference between the two corpora in terms of structural dispersion of LBs. The majority of clause-related structures were utilized through the use of anticipatory it in both corpora, but it was significantly more frequent among L2 compared to L1 as the LL value was observed to be -683.86, indicating a significant difference between the two groups. This sub-category usually indicates that the focus is on the evaluation of the propositional content with a disguise of the writer from the text (Hewings & Hewings, 2002; Hyland, 2008b). The examples (S7) and (S8) clearly display writers' claims without explicitly signaling themselves.

(S7) [...] because of the rhetorical traditions students mentioned above, and, it is important to note, material circumstances. (L1-30)

(S8) As such, it can be argued that not having a child is suggested as a desirable attribute for a woman who wants to get married. (L2-28)

The last structural category, others, was at the bottom of the list in both corpora. It was observed with an occurrence of 0.61 and 1.14 times per 10.000 words in L1 and L2 with -21.39 LL value, which was quite low compared to other structural categories. The writer directed readers on how the sentence should be interpreted in (S9).

(S9) Therefore, it can be expected that the lack of these features will have a negative effect on the quality of writing. (L1-26)

Now that we figured out the use of structural units of LBs in the two corpora, we can move on to the functional analysis of LBs found in the data. Table 4 displays the comparison of functional units of LBs by showing the frequencies, the types, and the LL values between the two corpora. From the table, we can see that text-oriented function was the most dominant category, followed by research-oriented and participant-oriented in the two corpora, which is in line with Hyland (2008a) and Lu & Deng (2019).

According to Table 4, the text-oriented category emerged 20.45 and 56.11 times per 10.000 words with 54 and 121 types in L1 and L2, respectively, which was proven to be statistically significant with -2263.48 LL. Framing was the most frequently utilized sub-category in the two corpora with 21 and 43 types in L1 and L2, respectively, indicating that L2 employed a variety of this sub-category compared to L1. In L1, LBs, with the transition function, occurred as the second most frequently employed category with 20.45 times per 10.000 words and 15 types. Structuring was the third and resultative was the least frequently employed sub-category among L1. In L2, however, resultative was the second most employed category with 14.19 occurrences per 10.000 words and we observed 31 resultative LBs. Structuring also had the same number of occurrences with a frequency of 12.70 per 10.000 words. Transition was at the bottom of the list with 16 types and a frequency of 12.39 per 10.000 words. The functions of text-oriented category were illustrated in sentences (S10) - (S13):

(S10) In "American Callings" I too explore a specific, limited strand of American identity: the identity of liberal, cosmopolitan American citizens in a position to be able to extend humanitarian aid to someone else. (L1-1; framing)

(S11) This fact that himself is bound as a result of the semantic computation derives Principle A for LSOR anaphors. (L1-56; resultative)

(S12) The results are presented in Table 41. (L2-10; structuring)

(S13) Section 4.2 presents the qualitative evidence recorded after the informal interviews that were carried out throughout the study and the observations made as well as the open-ended questions forwarded in the questionnaires, and discusses these. (L2-11; transition)

Table 4 The functional distribution of LBs in the two corpora

		Type				Frequency				LL Ratio (p<0.05)
		L1		L2		L1		L2		
		N	%	N	%	N	N/10000	N	N/10000	
Research-oriented	Description	5	4.85	55	20.00	170	1.36	2799	19.80	-2492.30
	Location	3	2.91	9	3.27	272	2.18	794	5.62	-205.12
	Procedure	5	4.85	12	4.36	199	1.60	681	4.82	-221.58
	Quantification	8	7.77	19	6.91	328	2.63	1083	7.66	-355.92
	Topic	2	1.94	-	-	97	0.78	-	-	---
	Sub-total	23	22.33	95	34.55	1066	8.56	5357	37.89	-2613.04
Text-oriented	Framing	21	20.39	43	15.64	978	7.85	2380	16.83	-439.56
	Resultative	8	7.77	31	11.27	336	2.70	2006	14.19	-1118.48
	Structuring	10	9.71	31	11.27	359	2.88	1795	12.70	-871.82
	Transition	15	14.56	16	5.82	875	7.02	1752	12.39	-197.93
	Sub-total	54	52.43	121	44.00	2548	20.45	7933	56.11	-2263.48
Participant-oriented	Engagement	4	3.88	9	3.27	139	1.12	518	3.66	-187.37
	Stance	22	21.36	50	18.18	886	7.11	3048	21.56	-998.59
	Sub-total	26	25.24	59	21.45	1025	8.23	3566	25.22	-1185.35

The dominant employment of text-oriented bundles by two groups of academic authors suggests that doctoral students have an awareness of their audience and positioning their claims in the norms of their disciplines. This functional category is also an effective means for writers to stamp themselves as competent academics, which is mainly achieved by transition bundles (Hyland, 2008a). The high concentration on framing bundles also shows that both groups of writers tend to "situate arguments by specifying limiting conditions" (Hyland, 2008a, p. 49). In this way, they may limit the truth of the propositional content and their claims and reduce the risks of readers' objection to their claims. Although L1 focused more on "establishing additive and contrastive links" with the use of transition bundles (Hyland, 2008a, p. 49), L2 give more importance to establishing a cause-effect relationship with the use of resultatives. The striking difference between the two corpora in terms of type and frequency of structuring might be an indicator of L2 refer more on the organization of the text against L1.

Research-oriented bundles showed statistically significant difference (LL= -2613.04) between the two corpora as L1 used them 8.56 times per 10.000 words with 23 types while LBs with research-oriented function were represented with 95 types and 37.89 occurrences per 10.000 words by L2. Among L1, we found quantification as the most frequently employed sub-category, followed by location, procedure, and description. However, the topic occurred with low-frequency counts. The description was calculated as the most frequent sub-category in L2. The relatively less common sub-categories were quantification, procedure, and location. Topic was not observed in L2. With the use of research-oriented bundles, both L1 and L2 tend to refer to their research rather than its presentation. While L1 demonstrate

their competence by controlling the quantification of the material sources with the use of quantification bundles, the heavy use of description bundles made us to state that L2 emphasize how they conduct the research. Extracts in (S14) - (S17) below demonstrate the sub-categories of research-oriented bundles.

(S14) The majority of the cross-linguistic data presented in this dissertation has focused on breathy voiced sonorants as rare (L1-33; quantification)

(S15) The rules of grammar do not necessarily apply to idioms as they do to other expressions, and knowledge of the literal meanings of the individual words does not always assist in determining the meaning of the idiom. (L1-45; description)

(S16) Specifically, then, threateners strengthen their stance through an implicit and/or explicit control over the victim and an emphasis on the victim's active participation in the threat; they weaken their stance by demonstrating a level of compassion for the victim and saving face through the use of more polite language. (L1-24; procedure)

(S17) Another problematic issue at this point is that from their responses or their enthusiasm during the interviews, it was observed that the learners participated in the current study might have overstated their feelings about these activities. (L2-10; location)

These findings are in line with Hyland's (2008a) study in which text-oriented features are more frequently utilized than research-oriented ones in doctoral dissertations. In Güngör & Uysal's (2016) study, text-oriented features were more frequented than research-oriented features in L2 research articles while the vice versa was observed in L1 research articles. It is likely that the employments of LBs may vary based on genre differences. As Hyland (2008b) states, the use of LBs is shaped by genres.

Consistent with the findings of Lu & Deng (2019) and Hyland (2008a), both L1 and L2 preferred to use fewer participant-oriented features. This category occurred 8.23 and 25.22 times per 10.000 words with an LL value of -1185.35, indicating a statistically significant difference across the two corpora at  $p < 0.01$  level. Participant-oriented LBs were represented by 26 types among L1, while there were 59 types among L2. Both groups had the tendency of employing more stance bundles than engagement bundles in their doctoral dissertations, contrary to Hyland (2008a), who observed that 70% of this category consisted of engagement markers in doctoral dissertations, reflecting a reader-focused text. The overuse of stance bundles may reflect that doctoral students are comfortable with aligning themselves with a personal evaluation of the content and their claims, as shown in (S18). In (S19), the writers tried to draw the readers' attention to a certain claim.

(S18) While the preservation of the [h] in these cases is somewhat unexpected given that the epenthetic vowel is unstressed, it is clear that resyllabification is not forced. (L1-12; stance)

(S19) Moreover, it should be noted that it was assumed in this study that the three universities in question follow more or less the same curriculum. (L2-16; engagement)

In order to find out the most common LBs in the two corpora, we calculated LBs that occurred over 40 times per million words, as Hyland (2008a) suggested. The results revealed that 26 LBs in L1 and 93 LBs in L2 were calculated as the most frequently recurring LBs in the corpora. Table 5 illustrates these particular LBs, and the ones in bold were used in both corpora. A total of 16 LBs were shared by L1 and L2. Seemingly, either corpus did not share similar LBs, which is in line with Salazar (2014), who observed that the most frequently employed LBs by L2 were not preferred by L1 English academic writers. The most common first eight LBs in L1 corpus occurred in L2 corpus with varying frequency counts and orders. However, on the other hand was the most frequented bundle in both corpora, which is consistent with Hyland (2008a).

Table 5 The most frequent LBs in the two corpora

L1	L2		
<p><b>on the other hand</b>  <b>in the case of</b>  <b>at the end of</b>  <b>at the same time</b>  <b>as well as the</b>  <b>the fact that the</b>  <b>in terms of the</b>  <b>it is important to</b>  in the context of  over the course of  the ways in which  <b>as a result of</b>  <b>that there is a</b>  the extent to which  <b>the results of the</b>  <b>to be able to</b>  <b>at the beginning of</b>  in the United States  <b>the rest of the</b>  in this section I  <b>in addition to the</b>  in this chapter I  <b>to the fact that</b>  are more likely to  it is possible that  in the same way</p>	<p><b>on the other hand</b>  the participants in the  <b>the results of the</b>  as opposed to the  the end of the  <b>in terms of the</b>  <b>at the end of</b>  the evaluation of the  the distribution of the  it is seen that  significant difference  between the  <b>as a result of</b>  <b>the fact that the</b>  <b>at the beginning of</b>  the difference  between the  the participants stated  that  in the use of  <b>to the fact that</b>  <b>in addition to the</b>  it is possible to  <b>as well as the</b>  one of the most  <b>that there is a</b>  of the present study  <b>in the case of</b>  is one of the  was found to be  the total number of  a statistically  significant difference  the analysis of the  as shown in table</p>	<p>in relation to the  in the present study  the number of the  in line with the  with the help of  can be seen in  in other words the  to find out whether  it was found that  that there is no  in terms of their  <b>at the same time</b>  majority of the  participants  native speakers of  English  most of the  participants  the majority of the  there is no difference  <b>to be able to</b>  are presented in table  the name of the  the use of the  due to the fact  in the form of  it can be concluded  it can be said  it can be seen  with the use of  as can be seen  with regard to the  <b>the rest of the</b>  with respect to the</p>	<p>of English language teaching  experimental and control  groups  of the fact that  it is observed that  are given in table  as seen in the  findings of the study  <b>it is important to</b>  the students in the  were found to be  in the light of  in accordance with the  in the target language  according to the results  number of the participants  on the basis of  when compared to the  it is clear that  it should be noted  that most of the  the nature of the  the role of the  for the first time  the most frequently used  there is a significant  as one of the  more than half of  of the study is  one of the participants  the relationship between the  that is to say</p>

The structural distribution of the most frequently employed LBs revealed that there were 16 PP-related, five NP-related, three clause-level, and two VP-related bundles in L1. Out of 93 LBs in L2, however, the distribution of structural categories was as follows: 32 PP-related, 30 NP-related, 22 clause-level, eight VP-related, and one others. The structural distributions of the both groups regarding the most pervasively used LBs in both corpora exhibited that PP related LBs were the most frequently preferred one. As for the functional distribution of the most common LBs, we found that text-oriented function was the most commonly utilized one in L1 and L2 corpora represented with 14 and 43 types successively. Participant-oriented and research-oriented functions occurred in 6 types of LBs in L1, while these two functions emerged in 22 and 28 types of LBs in L2, respectively.

## 5. CONCLUSION

Academic writing has long been a great question of interest in some particular fields, revealing that it does not simply refer to the presentation of propositional content but reflecting authors' stance and interacting with readers through the use of some linguistic features, one of which is LBs. Drawing on a detailed analysis of two specialized corpora, the main goal of the current study was to assess the use of LBs by L1 of English and Turkish L2 of English in doctoral dissertations. LBs in this study were investigated in terms of their structural and functional units by using Ant-Conc software, and LL statistics were run to examine whether there was a statistical difference between these two groups of authors.

Overall we observed a greater reliance on LBs by L2, which might be an indication of their less confidence in shaping their texts. In addition, the small amount of common LBs in the two corpora can be interpreted as an indication of the culturally-shaped production of doctoral dissertations in the Turkish academic community. Hyland (2004) states that disciplinary communities have an influence on rhetorical practices of arguing the claims and engaging readers into texts. Whether or not they are conscious choices or unreflective practices, authors follow a community-oriented employment of linguistic devices to construct their stance, texts, and interaction with their readers. In this study, we clearly can see a discrepancy of LB usage in the L1 academic community when compared with the L2 community in that community-oriented use of LBs in the L2 academic community, which does not share much with the L1 academic community. However, it may reduce the credibility of these academic authors in today's global academic world, adhering to L1 English speakers' linguistic norms.

Differences existed in the use of LBs by the two groups of authors in terms of structural categories. While L1 preferred more PP bundles, L2, however, employed more NP related formulaic sequences. Having said this, though, we can observe that both groups used substantially more text-oriented bundles compared to research-oriented LBs, which may be interpreted as an attempt to create a reader-friendly academic text and thus communicate with readers, as had also been well observed by Hyland (2008a). The two corpora included fewer participant-oriented bundles. Suffice to add that stance bundles were more common than engagement bundles in our corpora. The least variety of engagement bundles showed that both L1 and L2 tended to create a distance-relationship with their readers. It is likely that they were reluctant to put themselves as a member of the discipline in the leans of their readers who are highly qualified professionals of the discipline. On the contrary, readers' refusal of the claims is a major problem for writers to be dealt with. Therefore, in order for writers to persuade their readers about the truth of their claims, they need to signal their stance and engage the readers in their texts.

Awareness of common LBs in a particular discipline is a prerequisite of gaining communicative competence in that discipline. Thus, it may be of help to assist learners in acquiring these features of academic genres (Hyland, 2008a). In the implementation of learner corpora as a teaching material to a curriculum, several points need to be taken into account. Firstly, LBs that are driven from the recent academic genres should be applied in the classroom with a particular purpose. It is also important to build a knowledge of LBs incrementally with a focus on context. Since each academic genre has its own linguistic conventions, teachers of English using the language for academic purposes should emphasize the existing diversity of LB usage based on genres (Byrd & Coxhead, 2010).



Pang (2010) claims that students taking general English academic writing courses do not have an academic identity in their discipline in addition to being unaware of the expected language skills in their disciplines. LBs are disciplinary-based and it is essential for academic writing teachers to highlight dissimilar ways in different disciplinary environments. Corpus-informed lists and concordances can be used as teaching materials to assist learners in identifying common LBs specific to their disciplines, and through consciousness-raising activities, they can be encouraged to apply such identified entities in their own texts (Hyland, 2008b). It is also possible to offer a genre-based and disciplinary-specific instruction of English in academic writing courses. In such courses, students' awareness of cultural and universal norms of academic genres should be raised. To illustrate, in MA and PhD dissertations, postgraduate students may employ more cultural norms since their readers are specifically their supervisors and academicians, yet for a piece of research to be published in an international journal, they should be aware of universal linguistic norms since the potential readers would be academicians from different cultural and linguistic backgrounds. Most importantly, academic students should respond well to the expectations of academic genres with the appropriate use of linguistic devices to marshal their arguments convincingly; construct their stance, and negotiate with their readers through their academic texts.

Despite the contribution this study provides by identifying LB employment by L1 and L2 of English in doctorate dissertations, it is, however, limited to such bundles in a specific genre and field. Therefore, it would be too daring to generalize the results to all academic genres and disciplines. As for implications for further studies, a program evaluation of a professional development course for teachers to adapt corpus-linguistics and its outcomes into their teaching might yield fruitful results. The main problem in the teaching of academic writing, particularly faced by L2 English teachers, seems that they mostly use course-books that do not provide authentic materials for students to learn appropriate linguistic conventions. In this sense, an analysis of learner corpora and specialized corpora consisting of academic genres will offer effective teaching materials to be utilized in academic writing classes. More corpus-based studies should also be conducted to identify unrecognized linguistic and cultural features that may affect and shape L2 English speakers' academic writing.

#### REFERENCES

- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- 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. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97-116. <https://doi.org/10.1016/j.jeap.2006.05.001>
- Biber, D., & Conrad, S. (2009). *Register, genre and style*. Cambridge: Cambridge University Press.
- 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., 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., Johanson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Edinburg: Pearson Education.
- Burke, S. B. (2010). *The construction of writer identity in the academic writing of Korean ESL students: A qualitative study of six Korean students in the U.S.* (Doctoral Dissertation). Indiana University of Pennsylvania, Pennsylvania.
- 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(1), 38-52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- 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(5), 31-64.
- Chen, Y.H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49. Retrieved from <http://lt.msu.edu/vol14num2/chenbaker.pdf>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Demirel, E. T., & Hesamoddin, S. A. (2013). Lexical bundles in research article acknowledgments: A corpus comparison. *Hacettepe University Journal of Education*, 28(28-2), 457-468.
- Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English*, 38(2), 37-58. Retrieved from <http://hdl.handle.net/11222.digilib/126942>
- Genç, B., & Bada, E. (2010). English as a world language in academic writing. *The Reading Matrix*, 10(2), 142-151. Retrieved from [http://www.readingmatrix.com/articles/sept\\_2010](http://www.readingmatrix.com/articles/sept_2010)
- Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.
- Güngör, F. (2016). *Cross linguistic analysis of lexical bundles in L1 English, L2 English and L1 Turkish research articles* (Unpublished doctoral dissertation). Gazi University, Ankara.
- 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.
- Hewings, M., & Hewings, A. (2002). "It is interesting to note that...": a comparative study of anticipatory 'it' in student and published writing. *Journal of English for Specific Purposes*, 21(4), 367-383. [https://doi.org/10.1016/S0889-4906\(01\)00016-3](https://doi.org/10.1016/S0889-4906(01)00016-3)
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091-1112. [https://doi.org/10.1016/S0378-2166\(02\)00035-8](https://doi.org/10.1016/S0378-2166(02)00035-8)
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133-151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- Hyland, K. (2008a). 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. (2008b). 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. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32(1), 150-169. <https://doi.org/10.1017/S0267190512000037>
- Hyland, K., & Jiang, F. K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407. <https://doi.org/10.1075/ijcl.17080.hyl>
- Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia-Social and Behavioral Sciences*, 70, 622-628. <https://doi.org/10.1016/j.sbspro.2013.01.101>
- Liu, L., & Li, L. (2016). Noun phrase complexity in EFL academic writing: a corpus-based study of postgraduate academic writing. *Journal of Asia TEFL*, 13(1), 1-71.
- Loewen, S., & Plonsky, L. (2016). *An A-Z of applied linguistics research methods*. New York, NY: Palgrave Macmillan.
- Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36. <https://doi.org/10.1016/j.jeap.2019.03.008>
- Murray, M., & Moore, S. (2006). *The handbook of academic writing*. Berkshire: Open University Press.
- Muşlu, M. (2018). Use of stance lexical bundles by Turkish and Japanese EFL learners and native English speakers in academic writing. *Gaziantep University Journal of Social Sciences*, 17(4), 1319-1337.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283-304. <https://doi.org/10.1075/bct.17.03nes>
- Pan, F., Reppen, R. & Biber, D. (2016). Comparing patterns of L1 versus L2 academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21(1), 60-71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal*, 47(1), 10-11.
- 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>
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Philadelphia: John Benjamins Publishing Company.
- Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, 40, 1-14. <https://doi.org/10.1016/j.jeap.2019.04.004>
- Shin, Y. K., Cortes, V., & Yoo, I. W. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, 39, 29-41. <https://doi.org/10.1016/j.jslw.2017.09.004>
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1-14. <https://doi.org/10.1016/j.esp.2018.09.001>