

A Comparison of Two Genres: Lexical Bundles in the Discourse of Applied Linguistics

RAZIEH GHOLAMINEJAD

Shahid Beheshti University

r_gholaminejad@sbu.ac.ir

The present article is a corpus-based descriptive/comparative study of lexical bundles (LBs) in two university genres: textbooks (TBs) and research articles (RAs) on applied linguistics. It aims to identify the LBs used in the two genres, compare them on the basis of their functional type and frequency and explore how they are related to genre. To this end, four-word LBs were identified in two corpora drawn from applied linguistics TBs and RAs. The comparative analysis revealed that there are interesting differences between the two genres in terms of discourse functions: the occurrence of LBs in the TBs was lower than in the RAs; attitudinal/modality LBs occurred more frequently in the TBs than in the RAs; epistemic LBs occurred more frequently in the RAs than in the TBs; discourse organizers occurred more frequently in the RAs than in the TBs; and time, place and text reference LBs occurred almost twice as frequently in the RAs. The findings build on research into the variations of genres in terms of the use and functions of LBs in discipline-specific corpora.

Keywords: lexical bundle; comparative genre analysis; register analysis; register variation; applied linguistics discourse; corpus analysis

...

Una comparación entre dos géneros: paquetes léxicos en el discurso de la lingüística aplicada

Este artículo es un estudio descriptivo/comparativo de paquetes léxicos (LBs) en dos géneros universitarios: libros de texto (TBs) y artículos de investigación (RAs) en el campo de la lingüística aplicada. El objetivo es identificar los LBs utilizados en los dos géneros,

compararlos en base a su tipo funcional y frecuencia y explorar la relación entre los LBs y el género. A este fin, se identificaron LBs de cuatro palabras de dos corpus extraídos de TBs y RAs sobre lingüística aplicada. El análisis comparativo reveló diferencias interesantes entre los dos géneros en cuanto a las funciones discursivas: aparece menor cantidad de LBs en los TBs que en los RAs; los LBs actitudinales y modales son más frecuentes en los TBs que en los Ras; los LBs epistémicos ocurren con más frecuencia en los RAs que en los TBs; los organizadores del discurso aparecen en mayor cantidad en los RAs que en los LBs; y por último, los LBs de tiempo, espacio y referencia textual presentan casi el doble de frecuencia en los RAs. Los resultados se fundamentan en investigación sobre variaciones genéricas en el uso y las funciones de los LBs en corpus de disciplinas específicas.

Palabras clave: paquete léxico; análisis genérico comparativo; análisis de registro; variación de registro; discurso en lingüística aplicada; análisis de corpus

I. INTRODUCTION

Lexical bundles (LBs) are defined as “statistically the most frequent recurring sequences of words” (Hyland 2012, 150), are “highly sensitive to differences between text types” (Durrant 2015, 166) and make a considerable contribution to the coherence of a text (Hyland 2008b). They are nonidiomatic sequences with meanings that are transparently clear from their constituent parts (Biber et al. 2003). LBs can increase reading speed and comprehension as well as fluent nativelike production (Ellis et al. 2008; Hyland 2012). This is attributed to the fact that they are stored in the brain and then retrieved as whole unanalyzed chunks (Biber et al. 2004). The importance of LBs has to do with the fact that they can meet the most frequent communicative needs of language users by fulfilling pragmatic functions in discourse (Shin 2019). The identification of variations in the use of LBs has attracted the attention of researchers in recent decades. Fan Pan et al. observe that different genres and different groups of writers show “systematic differences” in their use of LBs (2016, 62). By showing that LBs behave in different ways in different academic disciplines, Ken Hyland challenges the widely held assumption that there exists a single set of core words that are useful for all students irrespective of their disciplines (2008b). Indeed, his argument in support of the disciplinary specificity of vocabulary underscores the necessity of developing specific language materials and delivering English courses specific to each group of students (Gholaminejad 2020a).

A number of earlier studies have focused on describing LBs across different genres (Biber et al. 2004; Biber 2006). Assuming the variability in LBs in different contexts of language use, the present study compares the occurrence of LBs in two academic genres—research articles (RAs) and textbooks (TBs) on applied linguistics—in order to explore how LBs are related to the genre in question. While TBs serve as a means of teaching knowledge, RAs typically report researchers’ interpretations of their results. Comparing the LBs occurring in these two different contexts of language use can thus lead to a deeper appreciation of genre variation in applied linguistics discourse. Drawing on Douglas Biber et al.’s functional taxonomy (2004), the present article examines which LBs are predominantly used in the two academic text types, taking into account that each genre has a different purpose and is targeted at different users.

To date, corpus-based studies on the identification of LBs in applied linguistics have been rare. Although Hyland’s interdisciplinary studies involve the identification of LBs in applied linguistics discourse compared to other disciplines (2008a, 2008b), little attention has been paid to the intradisciplinary variation in the occurrence of LBs in this field—that is, how the occurrence of LBs varies across different genres of applied linguistics. The current study is an attempt to fill this gap. More specifically, it aims to identify the four-word LBs that are unique to TBs and RAs on applied linguistics and those that are shared by the two genres, with the intention of determining the degree of similarity between the two genres in terms of their LB use, on the basis of Biber et al.’s taxonomy (2004). It also examines how the LBs and the genre they are present in are related.

2. LITERATURE REVIEW

2.1. Grammatical Structures of LBs

Researchers have examined the grammatical structures of LBs across genres and found that they vary depending on the genre in which they occur (Biber et al. 1999, 2004; Hyland 2008a, 2008b; Salazar 2014). Biber et al. examined the sequences in conversation and academic prose and found that these registers are dominated, respectively, by clausal and phrasal LBs (1999). Some years later, Biber et al. added two other registers to their previous corpora (2004). By comparing LBs across spoken registers—conversations and classroom teaching—and written ones—textbooks and academic prose—from six different disciplines—business, education, engineering, humanities, natural science and social science—they demonstrated that the two registers differ in the grammatical structures of the LBs they employ. Clausal verb-phrase based bundles—*is going to be*—occurred mostly in the spoken registers, whereas noun-phrase- and preposition-phrase- based bundles—*the beginning of the, in the case of*—were found mainly in the written register.

2.2. Discourse Functions of LBs

Research has also revealed that there is variation across genres in terms of the discourse functions of LBs. A number of functional taxonomies have been proposed to date for LBs (Biber et al. 2004; Hyland 2008a, 2008b; Salazar 2014). In the functional taxonomy designed by Biber et al. that is used in the present study, LBs are categorized into three groups (2004):

- Stance bundles can convey two kinds of meaning: epistemic and attitudinal/modality. Epistemic stance bundles comment on the knowledge status of the information in the ensuing proposition—*it is possible that, are more likely to*. Attitudinal/modality stance bundles express speaker attitudes towards the ensuing proposition—*it is important to, it is possible to*.
- Discourse organizers convey the relationship between prior information and that which is to come imminently—*on the other hand, as well as the*.
- Referential LBs make reference to physical or abstract entities or to the textual context itself. There are six subcategories:
 - a) Identification or focus LBs highlight the noun phrase following the bundle as being especially important—*one of the most*.
 - b) Tangible framing attributes are LBs that identify the specific size or shape of the head noun that follows—*the size of the, in the form of*.
 - c) Intangible framing attributes are LBs that describe abstract characteristics—*on the basis of, the extent to which, as a result of*.

- d) Place, time and text LBs denote a place, time or text that is not context dependent—*in the United States, at the same time.*
- e) Multifunctional reference LBs denote a place, time or text that is dependent on the particular context—*at the end of.*
- f) Specification of quantity LBs specify quantities or amounts—*the rest of the.*

In addition, Biber et al. found that genres vary in terms of the functional type of LBs they employ (2004). While conversations mostly use stance bundles, a greater number of referential bundles are used in the written register, due principally to the prominence of factual information in academic contexts (Conrad and Biber 2005).

Hyland proposes a modified version of Biber et al.'s 2004 taxonomy that narrows it down to LBs occurring in academic writing (2008b). In his classification, there are three major types of LBs:

- Research-oriented LBs help structure our experiences of the world and are subdivided into five subcategories:
 - a) Location LBs refer to time or place—*at the beginning of, in the present study.*
 - b) Procedure LBs show the way in which a study was conducted—*the purpose of the.*
 - c) Quantification LBs are used to identify quantity—*a wide range of.*
 - d) Description LBs refer to aspects of the object of the research or its context—*the structure of the.*
 - e) Topic LBs present the specific object of the research—*the currency board system.*
- Text-oriented LBs deal with the organization of the text and include:
 - a) Transition signals—*on the other hand.*
 - b) Resultative signals—*as a result of.*
 - c) Structuring signals—*in the present study.*
 - d) Framing signals—*in the case of.*
- Participant-oriented LBs focus on the writer or reader of the text and include:
 - a) Stance features—*are likely to be, may be due to.*
 - b) Engagement features—*it should be noted that.*

There have been criticisms of the above-mentioned taxonomies. Annelie Adel and Britt Erman (2012) and Ying Wang (2018), for instance, argue that they are still not fully developed and consist of subcategories that are not well defined or agreed upon in the literature. In spite of these reservations, Biber et al.'s 2004 taxonomy is used in the present study as it is the most comprehensive one developed thus far.

2.3. Comparison of LBs in Users' Productions

There has also been a recent surge of studies that have examined LBs with regard to language development by comparing the LBs produced by native versus nonnative speakers (Nekrasova 2009; Chen and Baker 2010; Adel and Erman 2012; Bychkovska and Lee 2017; Shin 2019) or between novice and expert writers (Cortes 2004; Chen and Baker 2010). Through comparing different language users' productions and identifying any deviations, these studies have shown that the LBs used by novice writers are functionally different from those produced by experts (Cortes 2002, 2004). In comparison to novice writers, expert writers produce more phrasal LBs—*the beginning of the, in the case of*—than nonphrasal LBs—*is going to be, if you want to*—and nonnative writers produce more phrasal LBs than native writers (Chen and Baker 2010). A number of researchers have also examined the use of LBs by nonnative writers across different proficiency levels (Staples et al. 2013; Chen and Baker 2016). Such studies have demonstrated that written production by nonnative speakers with lower proficiency levels tend to include not only more LBs overall (Staples et al. 2013) but also more features typical of conversation, while texts generated by more proficient writers are closer to academic prose (Chen and Baker 2016).

2.4. Comparison of LBs across Different Disciplines

There is some evidence for the existence of disciplinary variations in LBs, as seen in studies by Biber (2006), Hyland (2008b, 2012) and Philip Durrant (2015), where it was found that the use of LBs varies across different academic disciplines. Viviana S. Cortes compared the use of LBs in history and biology (2004); Hyland looked at engineering, microbiology, business and applied linguistics (2008a, 2008b); and Durrant made a detailed comparison of twenty-four different disciplines, grouped under humanities and social sciences and science and technology (2015). According to Cortes, hard sciences rely more on LBs than social sciences (2004). Durrant discovered that in science and technology LBs emphasized physical world and location, whereas in the humanities and social sciences they focused on abstract constructs (2015). Besides, he found that while LBs relating to cause and effect or referring to tables and figures commonly occurred in the latter (i.e., science and technology), LBs indicating relationships among entities or limitations were used frequently in the former (i.e., humanities and social sciences).

As the reviewed literature indicates, a number of comparative interdisciplinary studies have shown variation in the use of LBs across different academic disciplines. The present article contributes to the existing LB literature by examining the variation in the use of LBs across different genres—TBs and RAs—of the same discipline—applied linguistics.

3. METHODOLOGY

3.1. Corpora Compilation

To be able to compare TBs and RAs, access to two separate corpora drawn from the two genres was needed. For TBs, this study used the corpus of language teaching and applied linguistics textbooks developed by myself and Mohammad Reza Anani Sarab (2020). As shown in table 2, a total of 139 textbooks—available in both soft and electronic versions and published in the last three decades—were used to compile the corpus.¹ The corpus is balanced in such a way that it includes the same number of tokens for each subject area and each TB. Besides, it draws on TBs that applied linguistics students are commonly exposed to in the course of their studies.² As for the RA corpus, the websites of the major applied linguistics journals were visited and a random selection of RAs published over the last decade were downloaded from each journal, totaling 652 RAs that were then converted to TXT format.³ Table 1 shows the list of journals from which the RAs were downloaded and table 2 presents the main characteristics of the two corpora, the TB corpus (TBC) and the RA corpus (RAC).

TABLE 1. List of journals used as sources of RAs

Journals	
<i>Applied Linguistics</i>	<i>Journal of Second Language Writing</i>
<i>Assessing Writing</i>	<i>Journal of Psychoeducational Assessment</i>
<i>Assessment Quarterly</i>	<i>Language & Communication</i>
<i>Computers and Composition</i>	<i>Language</i>
<i>Contemporary Educational Psychology</i>	<i>Language Learning</i>
<i>Discourse & Communication</i>	<i>Language Sciences</i>
<i>Discourse & Society</i>	<i>Language Testing</i>
<i>Discourse Processes</i>	<i>Learning and Individual Differences</i>
<i>Discourse Studies</i>	<i>Learning and Instruction</i>
<i>ELT Journal</i>	<i>Lingua</i>
<i>English for Specific Purposes</i>	<i>Linguistics and Education</i>
<i>International journal of Bilingualism</i>	<i>Read Writ</i>
<i>International Journal of Corpus Linguistics</i>	<i>Reading Psychology</i>
<i>International Journal of Educational Research</i>	<i>RELC Journal</i>

¹ The whole list of textbooks is available in my PhD dissertation (Gholaminejad 2020c).

² A methodical procedure was followed for the selection of textbooks, which included checking the textbook offerings on major publishers' websites and the intended audiences in each case, and consulting the relevant syllabi. The final list was confirmed by experts.

³ The time span for TBs was three decades and for RAs one decade because access to TBs was limited compared to RAs.

Journals	
<i>Journal of English for Academic Purposes</i>	<i>Research in the Teaching of English</i>
<i>Journal of English Linguistics</i>	<i>Research on Language and Social Interaction</i>
<i>Journal of Language, Identity & Education</i>	<i>Scientific Studies of Reading</i>
<i>Journal of Neurolinguistics</i>	<i>System</i>
<i>Journal of Phonetics</i>	<i>TESOL Quarterly</i>
<i>Journal of Pragmatics</i>	<i>Text & Talk</i>
<i>Journal of Research in Reading</i>	<i>The Modern Language Journal</i>

TABLE 2. Description of the TB and RA corpora

	Number	Mean length (running words)	Total corpus size (running words)
TBC	139	77,562	10,781,118
RAC	652	9,000	5,626,767

3.2. Identifying and Comparing the LBs in the TB and RA Corpora

Following Biber et al., a frequency-driven approach was adopted to identify LBs in each of the two genres (1999, 2004). Only LBs made up of four words were included in the analysis. According to Tetyana Bychkovska and Joseph Lee, limiting the length of LBs to four words is a reasonable decision in that, compared to five-word LBs, they are more common in English (2017), and compared to three-word LBs, they are more manageable when it comes to checking the concordance lines for the purpose of classification (Chen and Baker 2010).

Given that the two corpora used in the present study were of different sizes, the raw frequencies of the retrieved LBs had to be converted to a normalized rate—per million words (PMWs). Since it has been argued that frequency cutoffs are arbitrary (Biber et al. 1999), normalized rates are usually set at a value of between 20 to 40 occurrences PMWs (Biber et al. 2004; Chen and Baker 2010). In this study the frequency threshold was set at 40 occurrences PMWs, using the criterion of “high-frequency LBs only” as established by Łukasz Grabowski (2015, 26). In order to identify the four-word LBs in the two corpora, the free program *kfNgram* was run on both separately and two sets of LBs were retrieved.⁴

Following Yu-Hua Chen and Paul Baker (2016) and Bychkovska and Lee (2017), the retrieved lists were manually checked for any topic-specific LBs, which are considered problematic in some studies. For example, Chen and Baker warn against including topic-specific LBs for analysis since their functions cannot be included in any of the prespecified

⁴ *kfNgram* (Fletcher 2007) is a program for linguistic research that generates lists of LBs from corpora.

categories in the taxonomy (2010). Hence, for the present study LBs specifically referring to applied linguistics, listed in table 3, were excluded from the analysis.

TABLE 3. Excluded topic-specific LBs in each corpus

RAs	TBs
<i>English for specific purposes</i>	<i>in the second language</i>
<i>in a foreign language</i>	<i>in the target language</i>
<i>in second language acquisition</i>	<i>language learning and teaching</i>
<i>in the teaching of</i>	<i>language teaching and learning</i>
<i>of English for academic</i>	<i>of second language acquisition</i>
<i>of spoken and written</i>	<i>of teaching and learning</i>
<i>of the English language</i>	<i>of the target language</i>
<i>of the variance in</i>	<i>the nature of language</i>
<i>of the words in</i>	<i>the study of language</i>
<i>of the writing process</i>	
<i>research on language and</i>	
<i>studies in second language</i>	
<i>the teaching of English</i>	
<i>the words in the</i>	
<i>times per million words</i>	
<i>word families in the</i>	

After classifying the retrieved LBs according to Biber et al.'s 2004 functional taxonomy, each of them was annotated manually by means of the concordance tool in *WordSmith Tools* (Scott 2018), which yielded concordance lists. The meaning and function of each LB was checked within the context by reading at least five random examples from the concordance lists. In some cases, further examples were read to determine the communicative function of the LB. Whenever clearly assigning a single specific function to an LB was not possible, the decision was made based on its "most common use" in the concordance lists (Biber et al. 2004, 384). Finally, the two lists of LBs generated from TBC and RAC were compared in order to determine the degree of difference in terms of frequency and functional types. The ultimate aim was to examine whether and how the differences are related to the nature of the genre.

4. RESULTS AND DISCUSSION

Despite the commonalities between the two genres in terms of LB frequency, interesting differences were observed. The total frequency of LBs occurring at least 40 times PMWs was 3,825 for the RAC and 3,191 for the TBC. The corpus analysis in the present study showed that LBs are less frequent in TBs than in RAs, which ties in with

Susan M. Conrad’s finding that the vocabulary used in RAs is rather restricted—since they resort to a limited set of fixed sequences of words—while the vocabulary in TBs is more varied (1996). In what follows, the LBs identified in the two corpora are presented classified according to functional types.

4.1. Attitudinal/Modality Stance

LBs conveying attitudinal/modality stance are shown in table 4. They occur more frequently in the TB corpus—364 PMWs—than in the RA corpus—209 PMWs. The majority expresses the ability to perform the action stated in the subsequent proposition. Among the attitudinal/modality LBs identified, only one expresses obligation—*it is important to*—which conveys the writer’s command for the reader to perform a given action.

TABLE 4. Attitudinal/modality stance LBs

Corpus	LB	Example
Unique to RAs	None	None
Unique to TBs	<i>it is possible to</i> <i>will be able to</i> <i>it is difficult to</i>	<u>It is possible to</u> extend this line of thinking into... (TBC) The assessee <u>will be able to</u> apply for... (TBC) <u>It is difficult to</u> interpret the immediate effect... (TBC)
Shared by both corpora	<i>it is important to</i> <i>can be used to</i> <i>to be able to</i>	<u>It is important to</u> emphasize/note (TBC and RAC) ...statistics <u>can be used to</u> verify qualitative... (RAC) <u>To be able to</u> analyze our examples, it is important to... (TBC)

The higher frequency of this type of LBs in TBs may indicate that TB writers have a stronger tendency to intrude into the discourse with their own attitudes, judgments and evaluations. According to Giovanni Parodi, in TBs particular attention is paid to the audience receiving academic instruction (2010), which according to my observations often involves the use of attitudinal/modality LBs. According to Biber and Conrad, RAs and TBs represent different types of relationships between writers and readers (2019): the former are addressed to other specialists in the field and the latter to novices. TB writers have more knowledge of the field and accordingly have authority over readers, whereas RA “readers are expected to already have expert background in the research area and so they are relatively equal in status to the writer” (Biber and Conrad 2019, 116). In other words, attitudinal/modality LBs occur more frequently in TBs because they are written by experts for novices, as opposed to RAs, written by specialists for other specialists.

4.2. Epistemic Stance

Epistemic stance LBs express degrees of certainty about the information in the ensuing

proposition. Their frequency the RAC—298 PMWs—is considerably higher than in the TBC—171 PMWs. Table 5 illustrates the LBs categorized under this functional type.

TABLE 5. Epistemic Stance LBs

Corpus	LB	Example
Unique to RAs	<i>it should be noted (that)</i> <i>to the fact that</i>	<u>It should be noted</u> that the correlation between... (RAC) In addition <u>to the fact</u> that many respondents highly valued... (RAC)
	<i>it is possible that</i>	<u>It is possible that</u> this usage pattern may... (RAC)
Unique to TBs	<i>is likely to be</i> <i>it is clear that</i>	...adaptation <u>is likely to be</u> accompanied by... (TBC) <u>It is clear that</u> in the absence of prior... (TBC)
	Shared by both corpora	<i>the fact that the</i>
<i>that there is a</i>		Let us accept <u>that there is a</u> correlation between... (TBC)

The higher frequency of epistemic LBs in the RAC suggests that RA writers explicitly clarify degrees of commitment or certainty with respect to their statements almost twice as often as TB writers. This is strongly to be expected from the situational differences between RAs and TBs. While TB writers commonly aim to inform readers of an already established piece of knowledge, according to Biber and Conrad, RA writers typically contribute new knowledge to their field and need to support their claims and convince readers that the knowledge presented has significance and scientific merit (2019). This is in line with the findings reported by John Swales, who claimed that RA authors use promotional elements to persuade readers about their findings (1990). At the same time, RA writers need to employ a wide range of lexical resources in order to clarify for the reader the extent to which their statements are grounded, which includes making extensive use of hedges to express doubt (Biber and Conrad 2019).

4.3. Discourse Organizers

Discourse organizer LBs occur more frequently in the RAC than the TBC—421 PMWs compared to 341. Only one topic introduction LB, exclusive to TBs, is found—*in this chapter we*. The rest of the discourse organizer LBs function as topic elaborators, establishing additive or contrastive associations between different propositions. Table 6 presents the LBs categorized under this functional type.

TABLE 6. Discourse organizer LBs

Corpus	LB	Example
Unique to RAs	<i>in addition to the</i>	<u>In addition to the</u> arguments raised above, there is also... (RAC)
Unique to TBs	<i>in this chapter we</i>	<u>In this chapter we</u> will examine... (TBC)
Shared by both corpora	<i>on the other hand</i>	<u>On the other hand</u> , many English-medium universities in ESL/EFL countries... (RAC)
	<i>as well as the</i>	protocol analysis, <u>as well as the</u> other nontraditional forms of assessment, is... (TBC)
	<i>on the one hand</i>	...language acquisition, <u>on the one hand</u> , and learning, on the other, are... (TBC) ...business discourse requires subject knowledge <u>on the one hand</u> and insights... (RAC)

The higher frequency of discourse organizers in the RAs is compatible with the finding presented by Conrad that connectors are more frequent in RAs than in TBs due to their impersonal style (1996). In the present study, discourse organizers were found to be the second most frequent type, after referential framing attributes. In the study by Hyland, the pervasiveness of this type of LB in the soft sciences, including applied linguistics, is also reported (2008b). He accounts for this by arguing that persuasion is interpretative in the soft sciences and knowledge is constructed through logical reasoning, which requires LBs to frame arguments by underscoring connections, limitations or case identification.

4.4. Referential: Identification or Focus

Identification or focus LBs put the focus on the noun phrase following the LB as especially important in introducing a discussion. They are slightly more frequent in the RAC—112 PMWs—than in the TBC—95 PMWs. As table 7 shows, no LB of this type occurs exclusively in one of the corpora and both TB and RA writers resort to similar LBs to highlight a notion, specifically *one of the most* and *is one of the*.

TABLE 7. Identification/Focus LBs

Corpus	LB	Example
Unique to RAs	None	None
Unique to TBs	None	None
Shared by both corpora	<i>one of the most</i>	PA is <u>one of the most</u> robust methods for... (RAC)
	<i>is one of the</i>	Design 5 <u>is one of the</u> most widely used... (TBC)

4.5. Referential: Framing Attributes

Framing attribute LBs contribute to the description of specific characteristics of the ensuing head noun by limiting the reader's attention to a certain aspect or condition of it. Such LBs occur in the RAC—1,497 PMWs—slightly more frequently than in the TBC—1,405 PMWs. Besides, while in the former this type of LB specifies details of the research conducted—*the results of the; the influence of the; the purpose of the*—in the latter they provide further explanations related to arguments—*the way in which; on the part of; in the same way*. Table 8 shows the LBs categorized under this functional type.

TABLE 8. Framing attribute LBs

Corpus	LB	Example
Unique to RAs	<i>the results of the</i>	<u>The results of the</u> study... (RAC)
	<i>the influence of the</i>	... <u>the influence of the</u> other languages... (RAC)
	<i>the purpose of the</i>	<u>The purpose of the</u> judging was... (RAC)
	<i>with respect to the</i>	<u>With respect to the</u> latter, the ratio of... (RAC)
	<i>of the use of</i>	advantage <u>of the use of</u> introspective methods... (RAC)
	<i>the relationship between the</i>	<u>The relationship between the</u> verbal and the visual... (RAC)
	<i>as a function of</i>	...online processes differ <u>as a function of</u> the type of... (RAC)
Unique to TBs	<i>the purpose of this</i>	<u>The purpose of</u> this study is to... (RAC)
	<i>for the purpose of</i>	...techniques were devised... <u>for the purpose of</u> giving practice in... (RAC)
	<i>the ways in which</i>	... <u>the ways in which</u> they communicate... (TBC)
	<i>the way in which</i>	<u>The way in which</u> language is used in... (TBC)
	<i>on the part of</i>	Learning thus remains a free act <u>on the part of</u> the user. (TBC)
	<i>in the same way</i>	...they will all act <u>in the same way</u> . (TBC)
	<i>the role of the</i>	<u>The role of the</u> classroom teacher... (TBC)
<i>in the field of</i>	<u>In the field of</u> applied linguistics, researchers... (TBC)	
<i>at the level of</i>	<u>At the level of</u> discourse, functional linguists are... (TBC)	
<i>in the sense that</i>	Videotaping provides more data than audiotaping <u>in the sense that</u> the researcher has... (TBC)	
Shared by both corpora	<i>in a way that</i>	KLM has acted <u>in a way that</u> is formally justified. (TBC)
	<i>in the form of</i>	Feedback is almost always <u>in the form of</u> some sort of score... (TBC)
	<i>in the case of</i>	<u>In the case of</u> the students in EFL Academic Writing courses, ... (RAC)

Corpus	LB	Example
Shared by both corpora	<i>on the basis of</i>	<u>On the basis of</u> the text, try to answer... (TBC)
	<i>in the context of</i>	This inventory has frequently been used <u>in the context of</u> English as a foreign... (RAC)
	<i>the extent to which</i>	<u>The extent to which</u> students are aware of... (TBC)
	<i>in terms of the</i>	<u>In terms of the</u> earlier examples, they examine... (RAC)
	<i>as a result of</i>	<u>As a result of</u> the negotiation, both participants end up... (TBC)
	<i>the use of the</i>	Presumably, <u>the use of the</u> term “material”... (RAC)
	<i>the nature of the</i>	<u>The nature of the</u> teacher’s contribution to... (TBC)
	<i>in the use of</i>	Disciplinary variation <u>in the use of</u> theme... (RAC)
	<i>in relation to the</i>	...question was always <u>in relation to the</u> learners. (TBC)
	<i>in the process of</i>	...language learning can be pushed <u>in the process of</u> genre learning... (RAC)
	<i>the meaning of the</i>	...conclusions about <u>the meaning of the</u> narrative... (TBC)
	<i>the degree to which</i>	... <u>the degree to which</u> the learners are engaged in the explanatory endeavor. (RAC)

The LBs functioning as referential framing attributes are the most frequent in the present study. Conrad and Biber (2005) and Biber and Federica Barbieri (2007) also report that referential LBs in general outnumber the rest of types in the written academic register, which they ascribe to the importance of presenting factual information in academic contexts. The higher frequency of LBs functioning as referential framing attributes in RAs is to be expected, since as Biber and Conrad point out, RAs discuss a specific research topic and therefore this type of LB is necessary to convey the exact identification of referents (2019). In contrast, TB writers tend to make general descriptions of the topic at hand.

4.6. Referential: Specification of Quantity

Quantification is a little more frequent in the TBC—320 PMWs—than in the RAC—300 PMWs—and, in addition, it is expressed by means of a greater variety of LBs in the former—4 LBs—than the latter—2 LBs. While TB writers use vaguer quantifiers—*a great deal of; the part of the; in a number of; in a variety of*—RA writers tend to use more precise ones—*the total number of; for each of the*. The LBs categorized under this functional type are shown in table 9.

TABLE 9. Specification of Quantity LBs

Corpus	LB	Example
Unique to RAs	<i>the total number of</i>	reduction in <u>the total number of</u> checklist items... (RAC)
	<i>for each of the</i>	<u>For each of the</u> four skills, Cohen's... (RAC)
Unique to TBs	<i>a great deal of</i>	...accumulating <u>a great deal of</u> evidence to support... (TBC)
	<i>the part of the</i>	<u>The part of the</u> brain that Broca identified as... (TBC)
	<i>in a number of</i>	<u>In a number of</u> influential books and papers, Halliday... (TBC)
Shared by both corpora	<i>in a variety of</i>	prominence is manifested acoustically <u>in a variety of</u> ways. (TBC)
	<i>a wide range of</i>	...enables findings to be tested across <u>a wide range of</u> subjects. (TBC)
	<i>a large number of</i>	...as might be expected, <u>a large number of</u> respondents marked... (RAC)
	<i>the rest of the</i>	Although <u>the rest of the</u> skills were fully covered... (TBC)
	<i>as part of the</i>	...should not be regarded <u>as part of the</u> English educational objectives... (RAC)

The greater occurrence of precise quantifying LBs is inevitable in RAs, where the clarity of the data presented is paramount. As mentioned earlier, in RAs, a specific research topic is dealt with, which requires linguistic resources that convey an exact identification of referents. On the other hand, TBs involve more general descriptions and place a greater emphasis on the explanation of concepts (Biber and Conrad 2019).

4.7. Place, Time and Text Reference

The LBs in this category, the majority involving text deictics, occur almost twice as frequently in the RAC—988 PMWs—than in the TBC—495 PMWs. Furthermore, while the RAC includes a wide variety of text deictics, the TBC contains only one (table 10). Multifunctional reference LBs—i.e., those that function differently depending on the context—are also present in both corpora, mostly referring to time or place—*at the end of*, *at the beginning of*. The only LB in the corpora referring to a particular place was *at the university of*. Table 10 shows the LBs categorized under this functional type.

Table 10. Place, time and text reference LBs

Corpus	LB	Example
Unique to RAs	<i>in the present study</i> <i>shown in table #</i> <i>of the present study</i> <i>at the university of</i> <i>table # shows the</i> <i>in table # the</i> <i>used in this study</i> <i>as shown in table</i> <i>as can be seen</i>	This was the procedure followed <u>in the present study</u> . (RAC) As <u>shown in Table 4</u> ... (RAC) The purpose of <u>the present study</u> is to investigate... (RAC) ...are <u>at the University of Michigan</u> ... (RAC) <u>Table 4 shows the</u> occurrences... (RAC) As illustrated <u>in Table 2, the</u> percentage of... (RAC) Courseware <u>used in this study</u> includes... (RAC) <u>As shown in Table 4</u> ... (RAC) <u>As can be seen</u> in the Appendix... (RAC)
Unique to TBs	<i>can be found in</i>	...AAVE <u>can be found in</u> many other English dialects... (TBC)
Shared by both corpora	<i>at the same time</i> <i>at the end of</i> <i>at the beginning of</i> <i>the end of the</i> <i>the beginning of the</i>	<u>At the same time</u> , however, it is not without problems. (TBC) ...to participate in the study, one at the beginning and one <u>at the end of</u> the semester... (RAC) Use Table 2.1 <u>at the beginning of</u> this chapter to help you. (TBC) ...when the beginning and <u>the end of the</u> period are compared. (RAC) ...after <u>the beginning of the</u> sound... (TBC)

The considerable difference between the two genres is predictable since RA writers commonly communicate research results to readers (Montesi and Owen 2008) and attempt to convince them by presenting—usually numerical—evidence to support their claims. They thus need to resort to many more deictic LBs to directly refer to appendices, tables or figures in the RA. In addition, Conrad (1996) showed that RAs have more informational characteristics than TBs, which are generally conveyed through the use of place, time and text references. Conversely, in TBs, established knowledge and existing views within a field are presented, which the writer is not typically required to support by referring to place, time or text evidence (Gholaminejad and Anani Sarab 2020)

5. CONCLUSIONS AND IMPLICATIONS

The corpus analysis in the present study shows that LBs occur frequently in applied linguistics discourse. Despite the commonalities in LB use between the two genres under study, considerable differences can be observed. LBs are more frequent in RAs than TBs, although attitudinal/modality stance LBs occur more frequently in TBs, which can be justified by the difference in the power relationships between writers and readers in these two genres. On the other hand, epistemic LBs are used more frequently by RA writers, who typically need to convince readers that the knowledge presented in the RA is significant. On their part, discourse organizers occur more frequently in RAs as a result of their more impersonal nature. Finally, referential LBs specifying quantity occur more frequently in TBs, although they are vaguer in nature than those used in RAs, which may stem from the more general nature of TBs. The three other types of referential LBs—identification or focus, framing attributes and place, time and text—are used more frequently in RAs. This can be ascribed to the fact that RA writers typically discuss a specific research topic, describe specific characteristics of referents or seek to support their claims by presenting evidence. In sum, the differences observed in this study can be accounted for by the differences in the nature, purpose and audience of each genre, and the underlying intentions of the authors.

The findings have implications for English for Academic Purposes (EAP). Firstly, as recommended by Hyland (2008b), the lists of LBs occurring in TBs and RAs compiled here can serve as realistic models for students of applied linguistics, through which they can recognize the specific characteristics of the genres they read and those they need to write. Writing an applied linguistics RA requires different ways of using language compared to writing a TB, because texts vary not only across disciplines, but also within a single one (Conrad 1996). In fact, writing RAs can be complicated for both native and nonnative students of applied linguistics (Wright 2019). In learning to write, they need to understand not only how meaning is conveyed, but also how language works in different genres and why. For these students, acquiring the knowledge necessary to produce different genres needs to be a part of the learning process since, according to Hyland, learning to use the LBs of a discipline contributes to increasing communicative competence in that discipline (2008a).

Secondly, the lists can serve as the basis for methodological practices including data-driven learning (Hyland 2008a). To do so, EAP instructors can ask students to study concordance lines in order to observe how LBs function in TBs and RAs. They can encourage students to pay attention to LBs and raise their awareness of the specific ways in which meaning is created in different genres. According to Chen-Yu Liu and Hao-Jan Howard Chen, explicit teaching of the functions of LBs is essential and needs to be contextualized to clarify the relationship between LB and surrounding items (2020). Finally, the LB lists presented here can be used as reference for the development of EAP materials for students of applied linguistics and offer insights for nonnative scholars who intend to write TBs in English in the field of applied linguistics.

WORKS CITED

- ADEL, Annelie and Britt Erman. 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.
- BIBER, Douglas. 2006. *University Language*. Amsterdam and Philadelphia: John Benjamins.
- BIBER, Douglas et al. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- BIBER, Douglas and Federica Barbieri. 2007. "Lexical Bundles in University Spoken and Written Registers." *English for Specific Purposes* 26 (3): 263-86.
- BIBER, Douglas and Susan M. Conrad. 2019. *Register, Genre and Style*. Cambridge: Cambridge UP.
- BIBER, Douglas, Susan M. Conrad and Viviana Cortes. 2003. "Towards a Taxonomy of Lexical Bundles in Speech and Writing." In Wilson, Rayson and McEnery 2003, 71-92.
- . 2004. "If You Look at...: Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25 (3): 371-405.
- BYCHKOVSKA, Tetyana and Joseph J. Lee. 2017. "At the Same Time: Lexical Bundles in L1 and L2 University Student Argumentative Writing." *Journal of English for Academic Purposes* 30: 38-52.
- CHEN, Yu-Hua and Paul Baker. 2010. "Lexical Bundles in L1 and L2 Academic Writing." *Language Learning & Technology* 14 (2): 30-49.
- . 2016. "Investigating Criterial Discourse Features across Second Language Development: Lexical Bundles in Rated Learner Essays, CEFR B1, B2 and C1." *Applied Linguistics* 37 (6): 849-80.
- CONRAD, Susan M. 1996. "Investigating Academic Texts with Corpus-Based Techniques: An Example from Biology." *Linguistics and Education* 8 (3): 299-326.
- CONRAD, Susan M. and Douglas Biber. 2005. "The Frequency and Use of Lexical Bundles in Conversation and Academic Prose." *Lexicographica* 20: 56-71.
- CORTES, Viviana S. 2002. "Lexical Bundles in Academic Writing in History and Biology." PhD diss., Northern Arizona University.
- . 2004. "Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology." *English for Specific Purposes* 23 (4): 397-423.
- DURRANT, Philip. 2017. "Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories." *Applied Linguistics* 38 (2): 165-93.
- ELLIS, Nick C., Rita Simpson-Vlach and Carson Maynard. 2008. "Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics and TESOL." *Tesol Quarterly* 42 (3): 375-96.
- FLETCHER, William. 2007. *KfNgram*. KWIC Finder.
- GHOLAMINEJAD, Razieh. 2020a. "What Do Iranian Undergraduate Students of Social vs. Natural Sciences Say about Their Language Needs?" *International Journal of Research in English Education* 5 (1): 104-15.

- . 2020b. "A Critical Review of the Word Classification System." *HOW Journal* 27 (2): 156-67.
- . 2020c. "Investigating Field-Dependency of Academic Words and Collocations: A Corpus Study of Language Teaching and Applied Linguistics Books." PhD diss., Shahid Beheshti University.
- GHOLAMINEJAD, Razieh and Mohammad Reza Anani Sarab. 2020. "The Academic Vocabulary and Collocations Used in Language Teaching and Applied Linguistics Textbooks: A Corpus-Based Approach." *Terminology* 26 (1): 82-107.
- GRABOWSKI, Łukasz. 2015. "Keywords and Lexical Bundles within English Pharmaceutical Discourse: A Corpus-Driven Description." *English for Specific Purposes* 38: 23-33.
- HYLAND, Ken. 2008a. "Academic Clusters: Text Patterning in Published and Postgraduate Writing." *International Journal of Applied Linguistics* 18 (1): 41-62.
- . 2008b. "As Can Be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27 (1): 4-21.
- . 2012. "Bundles in Academic Discourse." *Annual Review of Applied Linguistics* 32: 150-69.
- LIU, Chen-Yu and Hao-Jan Howard Chen. 2020. "Analyzing the Functions of Lexical Bundles in Undergraduate Academic Lectures for Pedagogical Use." *English for Specific Purposes* 58: 122-37.
- MONTESI, Michela and John Mackenzie Owen. 2008. "Research Journal Articles as Document Genres: Exploring Their Role in Knowledge Organization." *Journal of Documentation* 64 (1): 143-67.
- NEKRASOVA, Tatiana M. 2009. "English L1 and L2 Speakers' Knowledge of Lexical Bundles." *Language Learning* 59 (3): 647-86.
- PAN, Fan, Randi Reppen and Douglas Biber. 2016. "Comparing Patterns of L1 versus L2 English Academic Professionals: Lexical Bundles in Telecommunications Research Journals." *Journal of English for Academic Purposes* 21: 60-71.
- PARODI, Giovanni. 2010. "The Rhetorical Organization of the Textbook Genre across Disciplines: A 'Colony-in-Loops'?" *Discourse Studies* 12 (2): 195-222.
- SALAZAR, Danica. 2014. *Lexical Bundles in Native and Non-Native Scientific Writing: Applying a Corpus-Based Study to Language Teaching*. Amsterdam and Philadelphia: John Benjamins.
- SCOTT, Mike. 2018. *Wordsmith Tools* (version 5). Lexical Analysis Software.
- SHIN, Yu Kyoung. 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.
- STAPLES, Shelley et al. 2013. "Formulaic Sequences and EAP Writing Development: Lexical Bundles in the TOEFL iBT Writing Section." *Journal of English for Academic Purposes* 12 (3): 214-25.
- SWALES, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge UP.

- WANG, Ying. 2018. "As Hill Seems to Suggest: Variability in Formulaic Sequences with Interpersonal Functions in L1 Novice and Expert Academic Writing." *Journal of English for Academic Purposes* 33: 12-23.
- WILSON, Andrew, Paul Rayson and Tony McEnery, eds. 2003. *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Bern: Peter Lang.
- WRIGHT, Heidi R. 2019. "Lexical Bundles in Stand-Alone Literature Reviews: Sections, Frequencies and Functions." *English for Specific Purposes* 54: 1-14.

Received 24 December 2020

Revised version accepted 4 April 2021

Razieh Gholaminejad holds a PhD in English language teaching from Shahid Beheshti University, Iran. Among her research interests are corpus linguistics, English for Specific Purposes and academic discourse. In addition to publications in national journals, she has articles published by Routledge, Emerald and John Benjamins.