ExELL (Explorations in English Language and Linguistics)
12.2 (2024): 178-203
https://doi.org/10.51558/2303-4858.2024.12.2.178
Original scientific article

# A corpus-based analysis of the lexical bundles use in academic discourse

**Fadi Maher Al-Khasawneh**
*King Khalid University*
*Abha, Saudi Arabia*

## Abstract

This study aims to investigate the frequency of four-word lexical bundles (LBs) in research papers authored by faculty members from the English Department at King Khalid University. Additionally, it seeks to classify the functions of these bundles based on Biber's (2004) taxonomy. The study's corpus comprises 171 research papers published between 2016 and 2022. Lexical bundles were identified using three key criteria: frequency, range, and function. WordSmith 4.0 software was employed to extract and analyze the LBs from the corpus. The results reveal variation in the use of LBs, with referential bundles being the most common, followed by discourse organizers and stance bundles. These findings align with previous research in this area and offer valuable insights for studies on English for Academic Purposes (EAP). They may also benefit EAP instructors, curriculum developers, and policymakers.

**Keywords:** corpus; lexical bundles; academic discourse.

## 1. Introduction

Numerous academic studies have focused on the nature and use of formulaic language over the past several decades. Formulaic language is defined as discontinuous or continuous words that seem ready-made, grouped, and redeemed from memory at use (Wray, 2002). Similarly, Wood (2006) defines formulaic language as a constant series of words that play a role in communication and speech production and seem to be restored as single words. Formulaic language involves multi-word clusters that appear as single entities such as fixed expressions (e.g., *all of a sudden, one size fits all, happy new year)*, idiomatic expressions (e.g., *once in a blue moon, break a leg, under the weather)*, lexical bundles (e.g., *I agree with you, what do you think, is the way to),* collocations (e.g., *to save time, close a deal, strong smell),* and fillable slots (e.g., *either……or, neither…..nor)* (Sholkani, 2018).

Several studies (Kuiper, 2004; Taguchi, 2007; Wood, 2015; Wray, 2002) emphasize the significant role of formulaic language in pragmatic competence and speech fluency. They explain how formulaic language can help

learners sound more native-like, make fewer errors through ready-made expressions, and speak fluently and easily. Lexical bundles (LBs) are non-idiomatic, multi-word sequences that are structurally incomplete and fall under the broader category of formulaic language. Biber and Barbieri (2007) highlight that LBs comprise many formulaic expressions. Lexical bundles (LBs) are incomplete phrases that are semantically transparent, contributing to smoother communication (Biber & Barbieri, 2007). Unlike idioms, LBs can be expanded and show variation. For instance, an LB such as *as a consequence* can be extended into a longer form, such as *as a consequence of the*.

The use of LBs and their functions has been studied in different disciplines and registers. It has been shown that LBs are discipline-restricted to a certain degree. In other words, the LBs used in one discipline are not necessarily used in another (Sholkani, 2018). While it has been demonstrated that LBs are predominantly used in the written register, it is believed that there is a need for more focused studies that carefully explore the writers' inclination to use LBs along with the contextual meanings these bundles convey. The existing literature indicates that there is a lack of research focused on explaining the use and functions of LBs in the Saudi context. With this in mind, the current study explores the use of LBs and their functions in a large corpus compiled of recent research papers published by faculty members at the Department of English, King Khalid University, Saudi Arabia, from 2016-2022. Therefore, the present study seeks to address this gap by providing answers to the following research questions:

1. What are the most frequently used LBs in research papers published by English department faculty members at King Khalid University?
2. What is the functional distribution of LBs used in those research papers?

## 2. Literature review

### 2.1. *Lexical bundles*

There are many definitons of the term 'lexical bundles' (LBs). Biber et al. (1999) defined LBs as expressions of three or more words that frequently co-occur in a particular corpus, identified by specific distributional criteria and standardized frequency thresholds. Altenberg (1993) referred to LBs as recurrent word combinations, while Wood and Appel (2014) described them as multiword constructions. Unlike idioms and collocations, which often have fixed meanings and specific grammatical patterns, LBs are distinguished by their frequent, formulaic use in discourse, often functioning to structure the text or signal relationships between ideas. LBs across various disciplines are standard in English, especially in academic writing (Hyland,

2012). The use of LBs has been explored in various genres and registers across diverse academic disciplines in terms of its structure and function (e.g. Biber & Conrad, 1999; Conrad & Biber; 2005; Cortes, 2004; Hyland, 2008a, 2008b; Nasrabady, 2020; Zare & Valipouri, 2022).

The findings of previously mentioned studies revealed that all academic disciplines contain LBs but in different ranges; most of those bundles consist of adjectival phrases (e.g., is consistent with the), noun phrases (e.g., the results of the), anticipatory it (e.g., it should be noted), prepositional phrases (e.g., at the end of), and passive verbs (e.g., is based on). Hyland (2008b) attributed the diversity and similarity in using LBs in different academic disciplines to the purpose and audience for which those registers are written.

## 2.2. *Identification of lexical bundles*

According to Cortes (2004), LBs are identified using a frequency-based approach. She defined LBs as sequences of three or more words that occur in the same register. In Cortes's (2004) definition, LBs are identified according to three main components, namely, the occurrence of word sequence per million words (frequency), the purpose of the word cluster (function), and the distribution of LBs in a given text. The frequency component is essential in identifying LBs, and it is used by building a corpus or corpora and scanning it using specific corpus analysis software such as WordSmith software (Scott, 2007). This software lists the most frequent word clusters using a proper frequency cutoff.

The second component of LB identification deals with the distribution or the range of occurrence. The distribution is primarily related to the minimum number of written texts that should include the word sequences before considering it as an LB. Biber and Barbieri (2007) pointed out that LBs should occur in 10% of all texts to be considered LB. Wood (2015) emphasized the importance of the distribution criterion in eliminating the probability of using specific word sequences by a particular author much more than another author. In addition to the components above, functional properties are also considered a key component in identifying LBs, as they help identify those word sequences within a text.

## 2.3. *The structure of lexical bundles*

The structure of LBs is regarded as incomplete grammatical structure units or frames that link other structural components in a text. Biber (2004) found that only 15% of the LBs are regarded as complete clauses or phrases in texts, whereas less than 5% of the LBs in prosaic texts demonstrate complete

structural units. He also explained the difference between using LBs to bridge two structural units and prosaic texts and bundles in academic texts, which usually connect two phrases. Therefore, Biber (2004) suggested an LBs taxonomy, which included three primary types of bundles, namely noun and prepositional phrases [NP/PP] fragments, dependent clause [DC] fragments, and verb phrase [VP] fragments. These types of bundles differ according to their register. Biber's (2004) structural taxonomy of bundles can be further explained as follows:

**1. Lexical bundles of noun and prepositional phrase fragments**

a) Connector + noun phrase with of-phrase fragment
Examples: *a little bit of, one of the things*

b) Noun phrase with post-modifier fragment
Examples: *the way in which, those of you who*

c) Other noun phrase expressions
Examples: *and stuff like that, or something like that*

d) Prepositional phrase expressions
Examples: *at the end of the, of the things that*

e) Comparative expressions
Examples: *greater than or equal, as far as the*

**2) Lexical bundles of dependent clause fragments**

a) 1st/2nd person pronoun + dependent clause fragments
Examples: *you might want to, I want you to*

b) WH-clause fragments
Examples: *when we get to, what I want to*

c) If-clause fragments
Examples: *if we look at, if you have a*

d) (verb/adjective+) to-clause fragment
Examples: *want to do is, to be able to*

e) That-clause fragments
Examples: *that I want to, that there is a*

**3) Lexical bundles of verb fragments**

a) Connector + 1st/2nd person pronoun + VP fragment
Examples: *well I do not know, you do not have to*

b) Connector + 3rd person pronoun + VP fragment
Examples: *and this is a, it is going to be*

c) Discourse marker + VP fragment
Examples: *I mean you know, you know it was*

d) Verb phrase (with non-passive verb)
Examples: *have a lot of, is one of the*

e) Verb phrase with passive verb
Examples: *can be used to, is based on the*

f) Yes/no question fragments
Examples: *are you going to, do you want to*

g) WH-question fragments
Examples: *how many of you, what do you think*

## 2.4. *Functions of lexical bundles*

The classification of LB functions can be traced back to Altenberg (1993). He was the first to consider frequency as a significant criterion for identifying word clusters and categorizing them according to their functions. He manifested that recurrent word combinations are widely linked to specific situations, where their meaning may only be deduced from context. Altenberg (1993) claimed that each statement has a peculiar role in discourse, indicating how an utterance should be interpreted. Several researchers (Biber, 2004; Cortes, 2002) proposed different classifications of LB functions. Biber (2004) suggested a taxonomy for LBs functions that have been used in the previous research on LBs (Chen, 2008; Cortes, 2002; Hyland, 2008a, 2008b; Nasarbady et al., 2020; Taieb & Toumi, 2021; Zare & Valipouri, 2022). His taxonomy identifies three significant categories of LBs: stance expressions, discourse organizers, and referential expressions.

### 2.4.1. *Stance bundles*

Stance bundles show the writer's stance, attitude, or perspective towards the content. This collection of expressions can be divided into two subcategories: (1) attitudinal bundles that show the writer's feeling or attitude towards a forthcoming proposal and (2) epistemic bundles that exemplify knowledge status. These two subcategories can also be categorized into personal and impersonal bundles. Personal bundles are referred to the writer/speaker, while impersonal bundles show parallel meaning without being referred straightforwardly to the writer/speaker. According to Biber (2004), the vast majority of epistemic personal bundles in classrooms are personal, and the primary function of those bundles is to show uncertainty. In contrast, the impersonal epistemic bundles often indicate a degree of assurance rather than imprecision.

Modality/attitudinal are also considered as personal bundles. These bundles represent the writer/speaker's attitude towards an action or event. Attitudinal/modality bundles are also classified into four categories. First, desire bundles that represent desires or wishes or explore the desires or wants of other people—second, obligation bundles, which, in turn, can be personal or

impersonal. Third, prediction/intention bundles express the writer's/speaker's purpose in doing something, including personal and impersonal bundles. Finally, ability bundles show the speaker's/writer's ability to achieve something, including personal and impersonal bundles.

### 2.4.2. *Discourse bundles*

According to Biber and Barbieri (2007), discourse organizers align with text-oriented bundles, serving to structure a text, highlight its textual functions, and support the development of argumentation. These bundles are key in organizing and structuring ideas within a text and establishing connections between its various parts. Discourse bundles can be divided into two subcategories: (1) topic focus/introduction bundles, used to introduce new topics, and (2) topic elaboration bundles, used to expand on or clarify information about a given topic.

### 2.4.3. *Referential bundles*

The referential bundles category highlights key characteristics of an entity (Biber, 2004) and is divided into four subcategories: identification/focus bundles, imprecision indicators, bundles for specifying attributes, and text reference/place/time bundles. Identification/focus bundles pinpoint the noun phrase following them. For example, they may identify a specific group of students in focus, but they can also organize discourse by introducing topics or emphasizing key points.

Imprecision bundles signal that the reference is not exact or could encompass additional, similar references (e.g., *and so forth*, *something like that*). Bundles of specifying attributes describe the qualities or characteristics of the noun that follows the bundle. This subcategory includes three types: (1) tangible framing bundles, which describe physical form or size; (2) quantity specification bundles, which focus on amounts or quantities; and (3) intangible framing bundles, which describe abstract qualities. Lastly, text reference/place/time bundles indicate references to location, time, or specific parts of the text. Table 1 summarizes the functional taxonomy of lexical bundles (Biber, 2004).

Table 1: The functional taxonomy of lexical bundles (Biber, 2004)

| Category | Examples |
|---|---|
| **1. Stance Bundles** | |
| **1.1 Attitudinal Bundles** | |
| 1.1.1 Desire Bundles | *if you wish to, he would like to, I like to* |
| 1.1.2 Obligation Bundles | |
| 1.1.2.1 Personal | *you have to, I need to* |
| 1.1.2.2 Impersonal | *it is essential to* |
| 1.1.3 Prediction/intention Bundles | |
| 1.1.3.1 Personal | *I am not going to* |
| 1.1.3.2 Impersonal | *It is going to be* |
| 1.1.4 Ability Bundles | |
| 1.1.4.1 Personal | *he is able to* |
| 1.1.4.2 Impersonal | *it is possible to* |
| **1.2 Epistemic Bundles** | |
| 1.2.1 Personal | *I do not know if I do not think* |
| 1.2.2 Impersonal | *the fact that* |
| **2. Discourse Bundles** | |
| 2.1 Topic Focus/Introduction Bundles | *do you know that, have a look at* |
| 2.2 Topic Elaboration Bundles | *nothing to do with, on the other hand, in addition to* |
| **3. Referential Bundles** | |
| 3.1 Identification/Focus Bundles | *for those of you who* |
| 3.2 Imprecision Indicators | *and so forth, something like that* |
| 3.3 Bundles for Specifying Attributes | |
| 3.3.1 Tangible Framing Bundle | *in the shape of, the size of* |
| 3.3.2 Quantity Specification Bundles | *you have several merits, a small portion of* |
| 3.3.3 Intangible Framing bundles | *the nature of the, in terms of* |
| 3.4 Text Reference/Place/time Bundles | |
| 3.4.1 Place Reference | *in Saudi Arabia* |
| 3.4.2 Time Reference | *at the same time* |
| 3.4.3 Text Deixis | *as depicted/shown in the table* |

## 2.5. *Previous studies*

Several studies (Bal-Gezegin, 2019; Barbieri, 2018; Bychkovska & Lee, 2017; Chen & Baker, 2010; Malik et al., 2019; Nasrabady et al., 2020; Sholkani, 2018) have explored the use of lexical bundles in academic texts. Bal-Gezegin (2019) examined the usage of LBs in a corpus of published research articles across various academic disciplines. The corpus comprises numerous arti-

cles from high-impact journals, ensuring a diverse representation of research topics and methodologies. The findings reveal that referential bundles were the most frequently employed, underscoring their importance in establishing context and providing clarity in scholarly communication. Commonly identified bundles included phrases such as *the findings of this study* and *in this study*. The study also highlights the prevalence of discourse bundles, which serve to organize information and guide readers through the text.

Barbieri (2018) explored lexical bundles in informal written discourse, explicitly focusing on their functions in conversational contexts such as blogs and social media. The analysis conducted in this study was based on a corpus of informal texts, allowing for insights into the communicative purposes these bundles serve in everyday language use. The findings indicated that the bundles often express hedging, politeness, and emotional nuances, highlighting their role in managing interpersonal relationships in written discourse. Bychkovska and Lee (2017) investigated the use of lexical bundles in argumentative essays written by both native (L1) and non-native (L2) university students. The study analyzes a corpus of argumentative essays produced by L1 and L2 students, allowing for an in-depth exploration of lexical bundle usage in a specific genre. The findings revealed notable differences in the use of lexical bundles between L1 and L2 writers. L1 students frequently employed referential and stance bundles, which helped them establish context and express personal viewpoints. In contrast, L2 students were found to use discourse bundles more often, suggesting a tendency to prioritize coherence and structure in their writing.

Chen and Baker (2010) investigated the usage of lexical bundles in academic writing by both native (L1) and non-native (L2) English speakers. The researchers analyzed a corpus of academic essays from L1 and L2 students across various disciplines. The findings indicate that while both groups utilized lexical bundles, there were significant differences in their usage patterns. L1 writers used more referential bundles to connect ideas and establish context, whereas L2 writers frequently employed discourse bundles, which helped organize information and guide readers through their arguments. Malik et al. (2019) studied the use of lexical bundles in academic writing within the social sciences, focusing on both native (L1) and non-native (L2) scholars. The study utilizes a corpus of academic articles published in social science journals, encompassing a variety of topics to capture a comprehensive picture of lexical bundle use. The results revealed that referential bundles are the most frequently employed, reflecting their critical role in establishing context and conveying information in scholarly communication. The study highlighted differences between L1 and L2 writers. At the same time, both groups use referential bundles extensively; L2 writers are more

likely to utilize discourse bundles to enhance coherence and guide readers through their arguments.

Nasrabady et al. (2020) analyzed the use of lexical bundles in recently published research articles in the applied linguistics field. The researchers examined a corpus comprising 200 research articles from leading applied linguistics journals published over the past five years, ensuring a diverse range of topics and methodologies were represented. The analysis revealed that referential bundles were the most frequently used type, underscoring their importance in establishing context and linking new research to existing literature. The study emphasized the vital role of lexical bundles in facilitating effective communication among scholars and enriching the academic writing landscape in applied linguistics. Sholkani (2018) explored the use of lexical bundles within a selection of economics textbooks designed for first-year university students. The analysis was based on a corpus of several widely used economics textbooks, focusing on specific chapters introducing foundational concepts and theories. The results indicated that referential bundles were predominant, highlighting their role in establishing context and providing essential information to learners. Sholkani's findings suggest that using lexical bundles in educational materials is crucial for supporting students' comprehension of complex economic concepts and enhancing their academic literacy.

## 3. Methodology

### 3.1. *Compilation of the corpus*

The corpus of the current study comprised 1.378.482 words from 171 research articles published in the period from 2016-2022 by the faculty members at the Department of English Language at King Khalid University (KKU), Abha, Saudi Arabia. All the research articles are published in well-known journals indexed in Scopus, EBSCO, and DOAJ research databases. All research articles were downloaded from the college website as the list of published articles is constantly updated and accessible to all faculty members: https://clt.kku.edu.sa/en/Research-Publications. The researcher excluded all direct quotations, tables, figures, lists of references, and appendices since the focus was on the writers' use of lexical bundles.

### 3.2. *Identification of lexical bundles*

The present study focused on four-word lexical bundles due to their prominence in prior research, which found them manageable in size (Chen & Baker, 2010). According to Cortes (2013), although lexical bundles can range

from three to nine words, four-word bundles have been the focus of most studies because they appear more frequently. Biber et al. (1999) suggested that four-word bundles serve more diverse discourse functions compared to three- or five-word bundles.

Lexical bundles in this study were identified using three criteria: (1) frequency, (2) range, and (3) function. While there is no universally agreed-upon frequency threshold for identifying lexical bundles, previous research has used cut-off points ranging from 10 to 40 occurrences per million words (Chen, 2010; Cortes, 2004; 2013; Hyland, 2008; Wood & Appel, 2014). For this study, a minimum frequency of 40 occurrences per million words was set to ensure the inclusion of word clusters in the list.

The second criterion used to identify lexical bundles was 'the range.' This criterion helps to boost the frequency criterion since many authors prefer to use specific word clusters /phrases in their writing constantly. Biber (2004) indicated that the sequence of words should occur at least in five texts to be considered a lexical bundle. Biber's range helps lessen peculiar or individual uses of lexical bundles. The third criterion included in identifying lexical bundles was the function. This study included the word sequence that match the functions listed in the taxonomy of Biber (2004), namely stance bundles, discourse organizers, and referential bundles. The list of lexical bundles was identified using WordSmith Software 4 (Scott, 2007).

## 4. Results and discussion

### 4.1 *Frequency of lexical bundles*

This section presents the results obtained from the research conducted to answer the research questions of the present study. The frequency of lexical bundles used in research papers published by faculty members is tabulated and presented in the following table.

Table 2: The frequency of LBs used in the published research papers

| No. | Lexical Bundle | Frequency | Percentage | Number of Texts |
|---|---|---|---|---|
| 1. | On the other hand | 392 | 5.66 | 171 |
| 2. | The results of the | 288 | 4.16 | 169 |
| 3. | Is one of the | 224 | 3.23 | 166 |
| 4. | At the same time | 212 | 3.06 | 151 |
| 5. | At the end of | 186 | 2.69 | 148 |
| 6. | In the field of | 182 | 2.63 | 101 |
| 7. | The majority of the | 173 | 2.50 | 100 |
| 8. | Kingdom of Saudi Arabia | 171 | 2.47 | 97 |

| | | | | |
|---|---|---|---|---|
| 9. | As shown in table | 161 | 2.32 | 90 |
| 10. | In the context of | 160 | 2.31 | 88 |
| 11. | In the case of | 159 | 2.29 | 83 |
| 12. | As a result of | 158 | 2.28 | 81 |
| 13. | In the form of | 156 | 2.25 | 76 |
| 14. | In the present study | 155 | 2.24 | 68 |
| 15. | It is important to | 154 | 2.22 | 68 |
| 16. | The fact that the | 149 | 2.15 | 64 |
| 17. | In terms of the | 148 | 2.13 | 63 |
| 18. | On the basis of | 146 | 2.11 | 61 |
| 19. | At the beginning of | 146 | 2.11 | 59 |
| 20. | In the current study | 141 | 2.03 | 59 |
| 21. | In addition to the | 139 | 2.00 | 57 |
| 22. | To be able to | 134 | 1.93 | 52 |
| 23. | It is possible to | 131 | 1.89 | 50 |
| 24. | The nature of the | 131 | 1.89 | 48 |
| 25. | The purpose of the | 130 | 1.87 | 47 |
| 26. | The total number of | 127 | 1.83 | 47 |
| 27. | In this study the | 125 | 1.80 | 46 |
| 28. | It is necessary to | 119 | 1.72 | 45 |
| 29. | In the absence of | 116 | 1.67 | 42 |
| 30. | I would like to | 112 | 1.61 | 40 |
| 31. | Is more likely to | 98 | 1.41 | 37 |
| 32. | An increase in the | 92 | 1.33 | 37 |
| 33. | As shown in figure | 92 | 1.33 | 36 |
| 34. | An example of the | 90 | 1.30 | 34 |
| 35. | Is because of the | 88 | 1.27 | 34 |
| 36. | Has to do with | 86 | 1.24 | 33 |
| 37. | One of the important | 86 | 1.24 | 33 |
| 38. | A number of studies | 77 | 1.11 | 33 |
| 39. | If you want to | 68 | .98 | 30 |
| 40. | The percentage of the | 67 | .98 | 30 |
| 41. | A change in the | 62 | .89 | 29 |
| 42. | The rest of the | 60 | .86 | 29 |
| 43. | The number of the | 60 | .86 | 29 |
| 44. | The way in which | 57 | .82 | 27 |
| 45. | In foreign language classrooms | 57 | .82 | 27 |
| 46. | There is a need to | 56 | .81 | 26 |
| 47. | The size of the | 55 | .79 | 26 |
| 48. | The value of the | 54 | .78 | 26 |
| 49. | They would be able | 53 | .76 | 25 |
| 50. | Which is one of | 53 | .76 | 25 |

| | | | | |
|---|---|---|---|---|
| 51. | It is expected to | 52 | .75 | 23 |
| 52. | There are many reasons | 49 | .70 | 22 |
| 53. | One part of the | 48 | .69 | 19 |
| 54. | In this part of | 46 | .66 | 18 |
| 55. | That is one of | 46 | .66 | 16 |
| 56. | As revealed in table | 46 | .66 | 14 |
| 57. | I don't know what | 44 | .63 | 12 |
| 58. | A decrease in the | 44 | .63 | 9 |
| 59. | You don't have to | 43 | .62 | 6 |
| 60. | Anticipate that these findings | 42 | .60 | 6 |
| 61. | Is not going to | 40 | .58 | 5 |
| 62. | Or something like that | 40 | .58 | 5 |
| 63. | In other studies of | 40 | .58 | 5 |
| | **Total** | **6916** | **100** | |

As depicted in Table (2), a variety of lexical bundles have been used in the corpus of this study and it contained a total of (63) four-word LBs. The frequency of the used LBs ranged between 40-392 occurrences. The following figure shows the five most frequently used LBs.

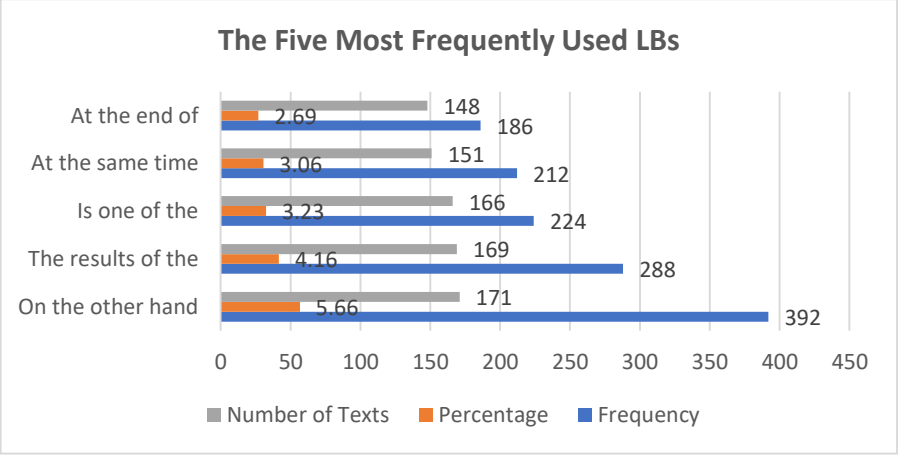


Figure 1: The five most frequently used LBs

As shown in Figure 1, the most frequently used LBs in the corpus were *on the other hand* with 392 occurrences which constitute (5.66%) of the overall number of used LBs appeared in all articles (171) texts, followed by *the results of the* with 288 occurrences (4.16%) in (169) texts, *is one of the* with 224 instances (3.23%) in (166) texts, *at the same time* 212 (3.06%) appeared in (151)

texts, and *At the end of* which appeared 186 times (2.69%) in 148 texts. Examples of the most frequently used LBs are as follows:

*.....**On the other hand**, in implicit intertextuality, the allusion is more oblique, for example, through similarities such as genre or style.....*

*…..**The results of the** PLS prediction have been considered…..*

*…..Learning words through flashcards **is one of the** simplest techniques to…..*

*…….....a compromise between addressing the demands of language knowledge and language skills and **at the same time** the proper ideological orientation…..*

*…..Moreover, the translator could not keep a similar rhyme and the same sound **at the end of** each line of the TL…..*

On the other hand, the least frequently used LBs were *in other studies of, or something like that,* and *are not going to* appear 40 times (.58%) in only five texts for each, followed by *anticipate that these findings* with 42 incidents (.60%) in (6) texts, and finally *you don't have to* with 43 occurrences (.62%) in (6) texts. Examples of these LBs are the following:

*…..Test takers **in other studies of** performance on cloze tasks have been observed…..*

*…..**or something like that** lesson plan…..*

*…..The poem **is not going to** lament the loss…..*

*…..**We anticipate that these** findings will provide guidelines for…..*

*…..**You don't have to** pay for these drinks…..*

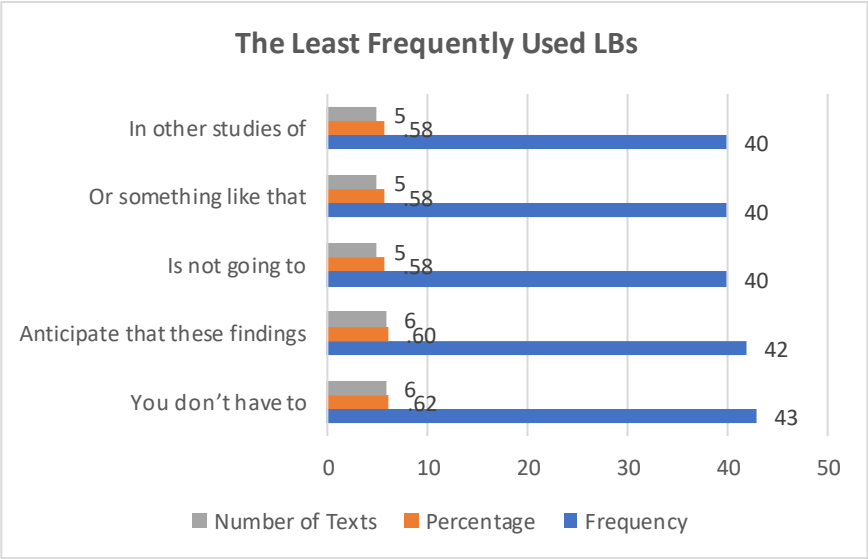


Figure 2: The five least frequently used LBs

### 4.2 *Functional distribution of lexical bundles*

The lexical bundles examined in this study were categorized based on Biber's (2004) functional taxonomy. The table below presents the distribution of the main categories and subcategories of lexical bundle functions.

Table 3: Functional distribution of LBs

| | | | |
|---|---|---|---|
| **1. Stance Bundles** | | | |
| **Attitudinal Bundles** | | | |
| **Desire Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 1. | I would like to | 112 | 1.61 |
| 2. | If you want to | 68 | .98 |
| | ***Total*** | **180** | **2.59** |
| **Obligation Bundles** | | | |
| **Personal Obligation Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 3. | You don't have to | 43 | .62 |
| **Impersonal Obligation Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 4. | It is necessary to | 119 | 1.72 |
| 5. | There is a need to | 56 | .81 |
| 6. | It is important to | 154 | 2.22 |
| | ***Total*** | **372** | **5.37** |
| **Intention/Prediction Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 7. | It is expected to | 52 | .75 |
| 8. | Anticipate that these findings | 42 | .60 |
| 9. | Is not going to | 40 | .58 |
| | ***Total*** | **134** | **1.93** |
| **Ability Bundles** | | | |
| **Personal Ability Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 10. | They would be able | 53 | .76 |
| **Impersonal Ability Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 11. | It is possible to | 131 | 1.89 |
| 12. | Is more likely to | 98 | 1.41 |
| 13. | To be able to | 134 | 1.93 |
| | ***Total*** | **416** | **6.01** |
| **Epistemic Bundles** | | | |
| **Personal Bundles** | | | |

| | *Lexical Bundle* | *Frequency* | *Percentage* |
|---|---|---|---|
| 14. | I don't know what | 44 | .63 |
| | **Impersonal Bundles** | | |
| | *Lexical Bundle* | *Frequency* | *Percentage* |
| 15. | The fact that the | 149 | 2.15 |
| | ***Total*** | **193** | **2.78** |
| | ***Overall Use of Stance Bundles*** | **1295** | **18.68** |
| | **2. Discourse Bundles** | | |
| | **Topic Focus/Introduction Bundles** | | |
| | *Lexical Bundle* | *Frequency* | *Percentage* |
| 1. | In the present study | 155 | 2.24 |
| 2. | In the current study | 141 | 2.03 |
| 3. | In the context of | 160 | 2.31 |
| 4. | In this part of | 46 | .66 |
| 5. | On the basis of | 146 | 2.11 |
| 6. | At the beginning of | 146 | 2.11 |
| 7. | The purpose of the | 130 | 1.87 |
| | ***Total*** | **924** | **13.33** |
| | **Topic Elaboration Bundles** | | |
| | *Lexical Bundle* | *Frequency* | *Percentage* |
| 9. | On the other hand | 392 | 5.66 |
| 10. | In addition to the | 139 | 2.00 |
| 11. | An example of the | 90 | 1.30 |
| 12. | Is because of the | 88 | 1.27 |
| 13. | Has to do with | 86 | 1.24 |
| 14. | That is one of | 46 | .66 |
| | ***Total*** | **841** | **12.13** |
| | ***Overall Use of Discourse Bundles*** | **1765** | **25.46** |
| | **3. Referential Bundles** | | |
| | **Identification/Focus Bundles** | | |
| | *Lexical Bundle* | *Frequency* | *Percentage* |
| 1. | There are many reasons | 49 | .70 |
| 2. | The results of the | 288 | 4.16 |
| 3. | One of the important | 86 | 1.24 |
| 4. | One part of the | 48 | .69 |
| 5. | The majority of the | 173 | 2.50 |
| 6. | A number of studies | 77 | 1.11 |
| 7. | Which is one of | 53 | .76 |
| 8. | In this study the | 125 | 1.80 |
| 9. | Is one of the | 224 | 3.23 |
| | ***Total*** | **1123** | **16.23** |
| | **Imprecision Indicators** | | |
| | *Lexical Bundle* | *Frequency* | *Percentage* |

| | | | |
|---|---|---|---|
| 10. | Or something like that | 40 | .58 |
| **Bundles for Specifying Attributes** | | | |
| **Tangible Framing Bundle** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 11. | In the form of | 156 | 2.25 |
| 12. | The size of the | 55 | .79 |
| | ***Total*** | **211** | **3.04** |
| **Quantity Specification Bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 13. | The total number of | 127 | 1.83 |
| 14. | The value of the | 54 | .78 |
| 15. | The percentage of the | 67 | .98 |
| 16. | A change in the | 62 | .89 |
| 17. | The rest of the | 60 | .86 |
| 18. | An increase in the | 92 | 1.33 |
| 19. | A decrease in the | 44 | .63 |
| 20. | The number of the | 60 | .86 |
| | ***Total*** | **566** | **8.16** |
| **Intangible Framing bundles** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 21. | In the case of | 159 | 2.29 |
| 22. | As a result of | 158 | 2.28 |
| 23. | In the absence of | 116 | 1.67 |
| 24. | The way in which | 57 | .82 |
| 25. | In terms of the | 148 | 2.13 |
| 26. | The nature of the | 131 | 1.89 |
| | ***Total*** | **769** | **11.08** |
| **Text Reference/Place/time Bundles** | | | |
| **Place Reference** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 27. | In other studies of | 40 | .58 |
| 28. | Kingdom of Saudi Arabia | 171 | 2.47 |
| 33. | In foreign language classrooms | 57 | .82 |
| 29. | In the field of | 182 | 2.63 |
| | ***Total*** | **450** | **6.50** |
| **Time Reference** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 30. | At the same time | 212 | 3.06 |
| 31. | At the end of | 186 | 2.69 |
| | ***Total*** | **398** | **5.75** |
| **Text Deixis** | | | |
| | ***Lexical Bundle*** | ***Frequency*** | ***Percentage*** |
| 32. | As shown in table | 161 | 2.32 |

| | | | |
|---|---|---|---|
| 33. | As shown in figure | 92 | 1.33 |
| 34. | As revealed in table | 46 | .66 |
| | ***Total*** | **299** | **4.31** |
| | ***Overall Use of Referential Bundles*** | **3856** | **55.88** |
| | ***Overall Frequency of LBs*** | **6916** | **100** |

### 4.2.1. *Overall distribution of LB functions*

The findings of this study showed that the total frequency of lexical bundles used in the corpus was 6,916. These LBs were functionally categorized into three main types: stance bundles, discourse bundles, and referential bundles (Figure 3).

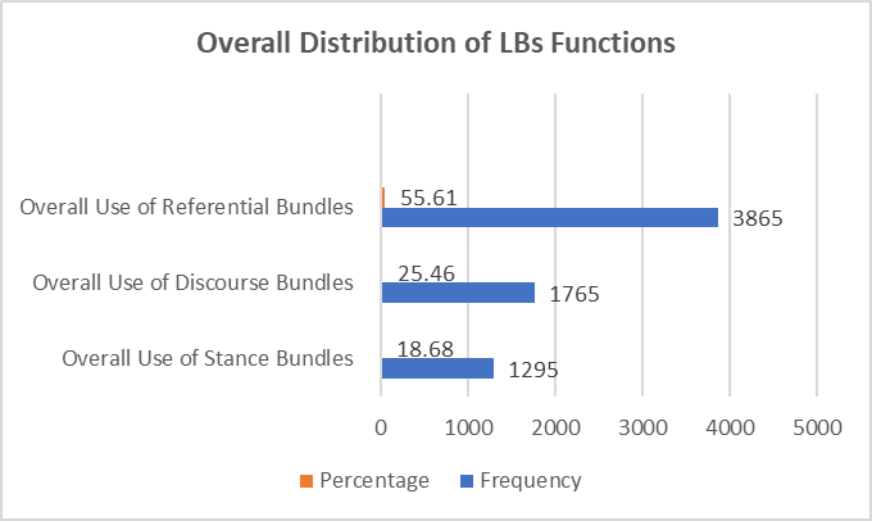


Figure 3: Overall distribution of LB functions

The above figure showed that most of the LBs were used for referential functions (N=3856, Per=55.88), followed by discourse functions (Freq=1765, Per=25.46) and stance functions (Freq=1295, Per=18.68).

### 4.2.2. *Distribution of stance function*

As mentioned earlier, stance bundles exemplify the writer's attitude, stance, or perspectives toward the content. Stance bundles can be categorized into epistemic bundles, ability bundles, intention/prediction bundles, obligation bundles, and desire bundles. The following figure depicts the distribution of stance bundles used in this study's corpus.

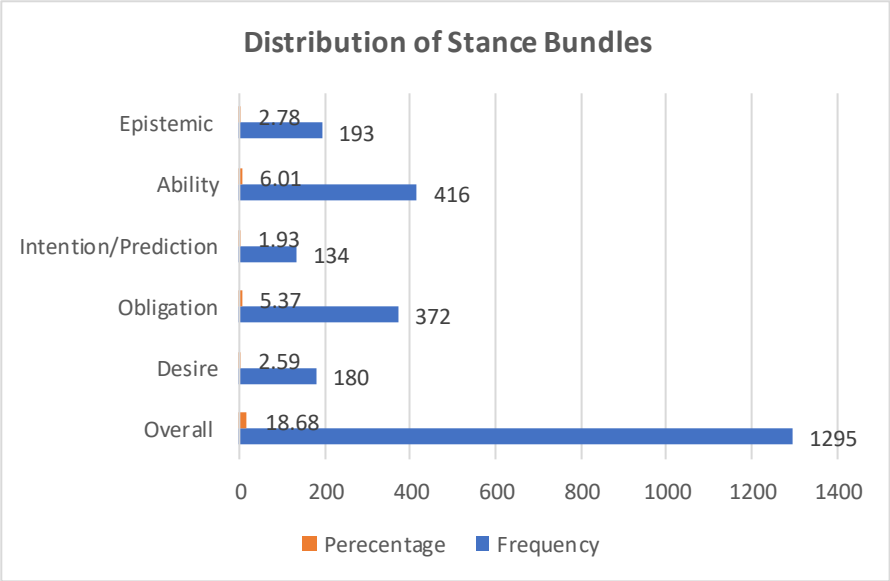


Figure 4: Distribution of stance bundles

Stance bundles have been used 1295 times with an overall percentage (18.68%). The results showed that the most frequently used stance bundles were the ability ones, which are used to show the writer's ability to achieve something (Freq=416, Per=6.01), followed by obligation bundles (Freq=372, Per=5.37), epistemic bundles (Freq=193, Per=2.78), desire bundles (Freq=180, Per=2.59), and intention/prediction bundles (Freq=134, Per=1.93). The following are examples for each category of stance bundles.

1) **Ability Bundles**

   a) Personal Ability
   ... that **they would be able** to successfully master similar situations…

   b) Impersonal Ability
   ...The translator adopted foreignization, which **is more likely to** cause misunderstanding…

2) **Obligation Bundles**

   a) Personal Obligation
   ...**You don't have to** pay for these drinks…

   b) Impersonal Obligation
   ...Thus, **it is important to** study how the individual psychological traits…

**3) Epistemic Bundles**

a). Personal Epistemic

...**I don't know what** this girl will be like when she is fifteen…

b) Impersonal Epistemic

...the lack of significant association might arise from **the fact that the** two groups' division is not clear...

**4) Desire Bundles**

…**I would like to** continue to learn English even after I leave this school/college…

**5) Intention/Prediction Bundles**

…and **it is expected to** contain diverse translation problems...

### 4.2.3. *Distribution of discourse function*

Discourse bundles are defined as expressions used to organize textual functions and to develop text argumentations. Discourse bundles help in organizing the flow of ideas and identifying the connection between the text portions. The distribution of discourse bundles is illustrated in the following figure.

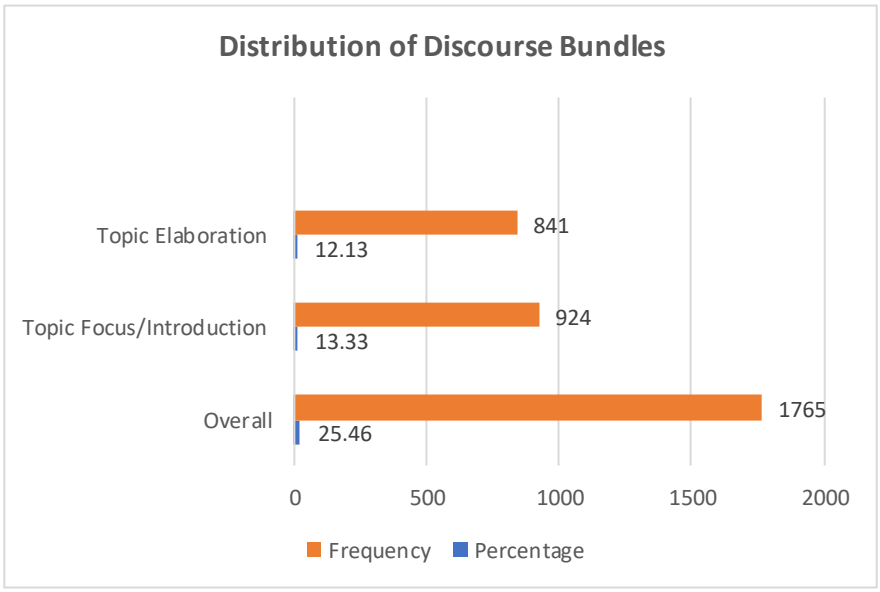


Figure 5: Distribution of discourse bundles

As shown in Figure 5, discourse bundles have been used 1765 times with an overall percentage (25.46%). The use of Topic/focus/introduction bundles

was more frequent (Freq=924, Per=13.33) than topic elaboration (Freq=841, Per=12.13). The following are examples for each category of discourse bundles.

**1) Topic Focus/Introduction Bundles**

... of the relationship between TEI and L2 resilience **in the present study** is similar to the…

...on the reality of introverted learners **in the context of** classroom presentations…

**2) Topic Elaboration Bundles**

...Foreignizing translation, **on the other hand**, is not always favored as…

…**In addition to the** questionnaire, classroom observation as well as interviews…

### 4.2.4 *Distribution of referential function*

The major role of referential bundles is to identify some attributes of a written text. This type of bundle consists of four subcategories: identification/focus bundles, imprecision indicators, bundles for specifying attributes, and text reference/place/time bundles. Figure 5 shows the distribution of referential function as used in the corpus of the present study.

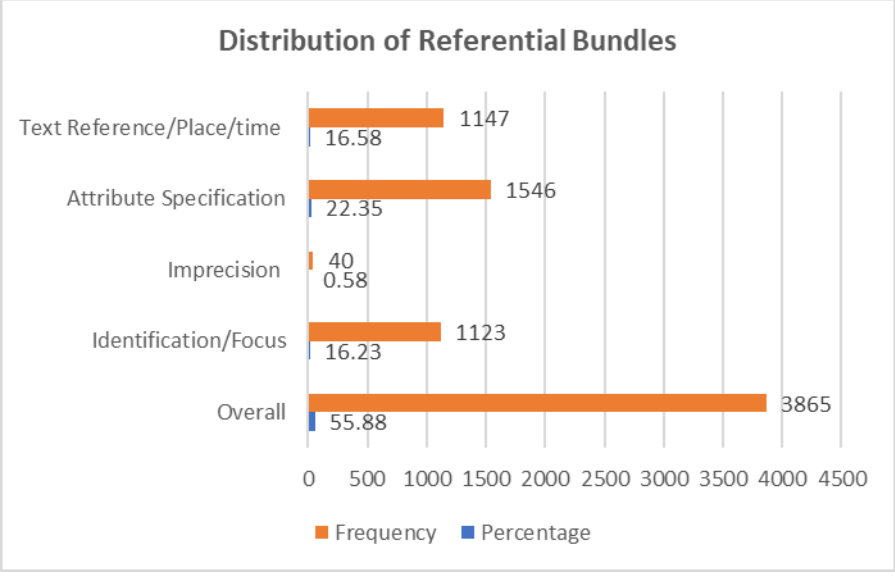


Figure 6: Distribution of referential bundles

As illustrated in Figure 5, the referential function was used 3856 times with an overall percentage (55.88%). Attribute specification bundles were the most frequently used function (Freq=1546, Per=22.35), followed by text reference/place/time (Freq=1147, Per=16.58), identification/focus (Freq=1123, Per=16.23), and imprecision function (Freq=40, Per=.58). The following are examples for each category of referential bundles.

**1) Identification/Focus Bundles**

…**The result of the** table is parallel with other studies which…
…For the majority of the students, the idea of…

**2) Imprecision Bundles**

…to be aware of the phrase **or something like that**…

**3) Attribute Specification Bundles**

a) Tangible Framing

…One of the major problems was **the size of the** class…

b) Intangible Framing

...can be created underlyingly in Turkish **as a result of** morphological operations…

c) Quantity Specification Bundles

…**The total number of** enrolled female students in the…

**4) Text Reference**

a) Place

... occupied a very prestigious position **in the field of** American literature…

b) Time

...**At the same time**, it develops teamwork, organizational, leadership, and problem-solving skills…

c) Text Deixis

...**as shown in Table** A in the online supplementary material…

The results of this study show a diverse use of LBs in research manuscripts published by faculty members. A closer look at the overall use of LBs across the functional categories, it can be noticed that *referential bundles* form the largest proportion of the used bundles (Freq= 3856, Per= 55.88%), followed by discourse bundles (Freq= 1765, Per= 25.46%), and stance bundles (Freq= 1295, Per= 18.68%). This finding is consistently congruent with Bal-Gezigen (2019), Malik et.al., (2019), Nasrabady et.al., (2020), and Sholkani (2018) who also found extensive use of referential bundles in their analyses. The high frequency of using *referential bundles* in the present study could be attributed to several reasons. One possible reason could be due to their ex-

treme importance in establishing context and making clear arguments. Faculty members may prefer using these bundles to ensure that readers can understand the reasoning throughout their research papers. The high frequency of using *referential bundles* could also be considered a sign of an academic environment that appreciates and values accuracy through connecting new arguments to existing research, which in turn contributes to ongoing scholarly conversations.

It is worth noting that the results of this study would also vary from those of some previous studies. Chen and Baker (2010) indicated that most Chinese learners of English tend to use discourse bundles more frequently than other categories of LBs. reported that discourse organizers are used more frequently in English for Academic Purposes (EAP) corpora. This difference emphasizes the effect of educational background and writing styles in different cultures. For instance, the emphasis on rhetorical strategies that prioritize engagement and flow in Chinese academic writing may lead to a greater reliance on discourse bundles, which serve to connect ideas and facilitate reader understanding. A study conducted by Barbieri (2018) revealed the pervasive use of stance bundles in blogs, while discourse organizers were the least used bundles. The varied distribution of LB function in the previous studies implies that 'genre' plays a significant role in using LBs and writing conventions.

Furthermore, it is essential to consider how the genre-specific demands of research articles may lead to an increased reliance on referential bundles compared to other types of writing. Research articles often require a formal tone and a structured presentation of arguments, which can naturally result in a greater emphasis on referential language. This is particularly relevant in the social sciences, where establishing connections with previous research is vital for validating new contributions to the field. The findings from this study underscore the necessity for researchers to adapt their language use according to the conventions of their discipline, which may contribute to the observed predominance of referential bundles.

On the other hand, Bychkovska and Lee (2017) and Hernandez (2013) reported that the use of LBs may vary between native writers when compared to non-native ones. This implies the association between the use of LBs and the writers' competence in the English language. However, this association seems to be complex and needs further investigation (Malik et.al 2019). This claim could interpret the high frequency of using referential bundles in the corpus of this study; the writers tend to use academic jargon composed of LBs when producing academic discourse since it constitutes their linguistic repertoire. Another interpretation of the high use of referential bundles in this study could be attributed to the referential nature of languages and academic discourse. The research manuscripts in the field of social sciences

refer to textual bundles very frequently due to their extreme importance in any formal discourse (Malik et.al, 2019).

However, it is essential to recognize that the relationship between LB usage and writing proficiency is complex and warrants further investigation (Malik et al., 2019). The current study's results could further indicate that the referential nature of academic discourse necessitates frequent use of textual bundles, which serve to clarify arguments and solidify connections with existing literature. This reflects the inherent expectations of academic writing, where clarity and precision are paramount.

The implications of this study extend beyond individual writing practices. They suggest a need for targeted pedagogical interventions that focus on teaching the use of LBs, particularly referential bundles, to enhance academic writing skills among both native and non-native speakers. Understanding the functional significance of these bundles can empower writers to construct more coherent and persuasive arguments, thereby improving their contributions to scholarly discourse. In summary, while this study demonstrates a predominant use of referential bundles, it also underscores the significance of genre, writer competence, and the communicative purposes of academic writing in influencing the distribution and function of lexical bundles. The findings highlight the importance of continued research into lexical bundle usage across various contexts and genres to develop a comprehensive understanding of their role in effective academic communication.

## 5. Conclusion

The main objective of the current study was to explore the use of LBs in a corpus of research articles published in the field of English language and literature. The findings of this study provided evidence of the use of LBs by writers/authors in the field of the English language. These findings will assist novice authors in producing more effective academic texts. The present analysis attempted to find out the most frequently used lexical bundles in (1.378.482) word sample of published research papers. The results showed various LBs indicative of their mental lexicon and writing proficiency. The study also found that referential bundles were significantly overused when compared to other types of bundles (i.e., discourse organizers and stance bundles). This might be ascribed to the growing demand for publishing and abundant resources for developing academic discourse.

Based on these findings, the present study recommends further analysis of the use of LBs, especially in other genres, to get more insights into the use of LBs in different genres. The results obtained from those investigations can be used to design more effective materials and shape the specific discourse of academic writing in the field of English language and literature. It is also

recommended to conduct classroom-based research to explore how to control the use of LBs and their functions. Such studies would help to know whether teaching LBs can be successfully achieved and the fluency-related benefits that can be gained from the instruction.

### *Acknowledgment*

The author extends his appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through the Large Group Research Project under grant number (RGP2/48/45).

## References

Altenberg, Bengt (1993). Recurrent verb-complement constructions in the London-Lund Corpus. Aarts, Jan, Pieter de Haan, Nelleke Oostdijk, eds. *English language corpora: Design, analysis and exploitation*. Amsterdam/Atlanta: Rodopi, 227–245.

Bal-Gezegin, Betül (2019). Lexical bundles in published research articles: A corpus-based study. *Journal of Language and Linguistic Studies 15*(2): 520–534. https://doi.org/10.17263/jlls.586188

Barbieri, Federica (2018). *I don't want to and don't get me wrong*: Lexical bundles as a window to subjectivity and intersubjectivity in American blogs. Kopaczyk,m Joanna, Jukka Tyrkkö, eds. *Applications of Pattern-driven Methods in Corpus Linguistics*. Amsterdam/Philadelphia John Benjamins, 251–276. https://doi.org/https://doi.org/10.1075/scl.82.10bar

Biber, Douglas (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, Douglas. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers.* Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/scl.23

Biber, Douglas, Federica Barbieri (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, Douglas, Susan Conrad (1999). Lexical bundles in conversations and academic prose. Hasselgård, Hilde, Signe Oksefjell (Eds.) *Out of Corpora: Studies in Honor of Stig Johansson.* Amsterdam: Rodopi, 181–189.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan (1999). *Longman Grammar of Spoken and Written English.* Harlow: Pearson Education Ltd.

Bychkovska, Teytana, 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. https://doi.org/10.1016/j.jeap.2017.10.008

Chen, Lu (2008). *An investigation of lexical bundles in electrical engineering introductory textbooks and ESP textbooks.* Doctoral dissertation, Carleton University. https://doi.org/10.22215/etd/2008-07668

Chen, Yumin, Paul Baker (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology 14*(2): 30–49.

Conrad, Susan, Douglas Biber (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica 20*: 56–71. https://doi.org/10.1515/9783484604674.56

Cortes, Viviana (2002). Lexical bundles in Freshman composition. In Reppen, Randi, Susan Fitzmaurice, Douglas Biber, eds. *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 131–145. https://doi.org/10.1075/scl.9.09cor

Cortes, Viviana (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes 23*(4): 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Cortes, Viviana (2013). The purpose of this study is to: connecting lexical bundles and moves in research article introductions. *English for Academic Purposes 12*: 33–43.

Hernández, Sandra (2013). Lexical bundles in three oral corpora of university students. *Nordic Journal of English Studies 12*(1): 187–209. https://doi.org/10.35360/njes.281

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

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

Hyland, Ken (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics 32:* 150–169. https://doi.org/10.1017/s0267190512000037

Kuiper, Koenraad (2004). Formulaic language and the lexicon (review). *Language 80*(4): 868–871. https://doi.org/10.1353/lan.2004.0209

Malik, Moazzam Ali, Hafiz Zeeshan Fazal, Hassan Moavia (2019). Formulaic language in social sciences: A functional analysis of lexical bundles in native and non-native academic discourse. *Pakistan Social Sciences Review 3*(1): 234–249. https://doi.org/10.35484/pssr.2019(3-i)17

Nasrabady, Peyman, Majid Elahi Shirvan, Seyed Ehsan Golparvar (2020). Exploring lexical bundles in recently published papers in the field of applied linguistics. *Journal of World Languages 6*(3): 175–197. https://doi.org/10.1080/21698252.2020.1797992

Öztürk, Yasemin, Gul Köse (2016). Turkish and native English academic writers' use of lexical bundles. *Journal of Language and Linguistic Studies 12*(1): 149–165.

Scott, Michael (2007). *Oxford Wordsmith Tools: Version 5.0*. Released June 2007 from http://www.lexically.net.

Sholkani, Ali (2018). *Lexical Bundles in University First Year Economics Textbooks*. Doctoral dissertation, Carleton University.

Taguchi, Naoko (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research 11*(4): 433–457. https://doi.org/10.1177/1362168807080962

Taieb, Amal, Nabil Toumi (2021). Exploring conventionalized phraseology in advanced academic writing of Tunisian linguistics students: A lexical bundles analysis. *Arab Journal of Applied Linguistics* 6(2): 1–30.

https://mail.arjals.com/ajal/article/view/315/166
Wood, David (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review 63*(1): 13–33. https://doi.org/10.3138/cmlr.63.1.13
Wood, David (2015). *Fundamentals of Formulaic Language: An Introduction*. London: Bloomsbury Publishing.
Wood, David, Randy Appel (2014). Multiword constructions in first-year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes* 15: 1–13. https://doi.org/10.1016/j.jeap.2014.03.002
Wray, Alison (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511519772
Zare, Javad, Leila Valipouri (2022). Lexical bundles in chemistry research articles. *Frontiers in Psychology 13*: 1–13. https://doi.org/10.3389/fpsyg.2022.906641

## Author's Address

Fadi Maher Al-Khasawneh
Department of English Language, College of Languages and Translation, King Khalid University,
Saudi Arabia
Email: fadialkhasawneh1983@gmail.com
Orcid: https://orcid.org/0000-0003-3484-591X

Received: April 27, 2024
Accepted for publication: September 30, 2024