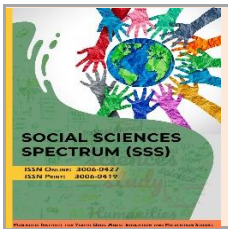

Social Sciences Spectrum

A Double-Blind, Peer-Reviewed, HEC recognized Y-category Research Journal

E-ISSN: 3006-0427 **P-ISSN:** 3006-0419
Volume 05, **Issue** 01, 2026
Web link:https://sss.org.pk/index.php/sss

Functional Variation in Lexical Bundles: A Corpus-Based Comparative Analysis of Native and Non-Native Web-Based English

Mairam Arif [1]
PhD Scholar, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan

Behzad Anwar [2]
Associate Professor, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan

Uswa Shahid [3]
PhD Scholar, Department of English, University of Gujrat, Gujrat, Punjab, Pakistan
Correspondence: uswaawan27@gmail.com

Article Information [YY-MM-DD]

Received 2026-01-13 **Revised** 2026-02-24 **Accepted** 2026-03-25

Citation (APA):

Arif, M., Anwar, B & Shahid, U (2026). Functional variation in lexical bundles: A corpus-based comparative analysis of native and non-native web-based English. *Social Sciences Spectrum*, *5*(1), 409-429. https://doi.org/10.71085/sss.05.01.506

Abstract

The current study compares the functional categories of lexical bundles in native and non-native web-based corpora. The main purpose of the study is to compare the frequency-based differences of lexical bundles in native and non-native corpora drawn from one of the largest web-based corpora. Drawing on interlingual differences, the study aims to unearth the different uses of lexical bundles by native and non-native users of English. A corpus-driven comparative approach is employed in the research to conduct the frequency-based differences. The data is obtained from one of the largest web-based corpora, GloWbe. AntConc (3.5.9) is used to derive 4-word lexical bundles from the corpus. Functional variations of lexical bundles presented by Biber et al. (2004) are used as a theoretical framework. The findings reveal that diversity in the use of lexical bundles exists among native and non-native speakers. Natives use lexical bundles in a more authentic, functional, balanced, and genuine way. At the same time, non-natives use the lexical bundles in a limited way in order to meet the needs of discourse functions only. Explicit instruction in lexical bundles offers a pedagogical advantage by enhancing students' ability to communicate effectively in academic discourse.

Keywords: Lexical Bundles, GloWbe, Functional Variations, Native and Non-native, Corpus-driven Approach, Web-based Corpus.

1. Introduction

In recent years, corpus-based research has broadened considerably, incorporating diverse varieties of English, especially British and Pakistani English (Shahid, Anwar & Amin, 2025). As learning English is a complex phenomenon, lexical bundles play a substantial role in attaining proficiency in English. Language, one of the most intricate and dynamic human communication systems, is influenced by social, cultural, and cognitive factors. In the context of discourse analysis, multi-word sequences—particularly lexical bundles have drawn more attention from researchers as essential components of linguistic organization. According to Biber et al. (2004), lexical bundles are recurrent word sequences that frequently occur in discourse but are not always idiomatic or fixed expressions. These lexical bundles are very significant in achieving communicative competence, i.e., the ability to use the English language correctly according to specific situations and contexts (Hymes, 1972). In short, communicative competence helps a speaker to achieve full communicative purpose.

The coherent and fluent communication of speakers is due to lexical bundles (Hyland, 2008a). It is important to study how lexical bundles help in understanding language structure and how they vary from speaker to speaker due to different linguistic backgrounds. Lexical bundles serve specific functions in discourse rather than simply being combinations of words. Information, stance, expression, organisation, and management of text are some examples of lexical bundles (Biber et al., 2004). Frequency distribution shows a rich source of information about language competency, and the functions depend on the genre and spoken proficiency (Cortes, 2004). Lexical bundles used in academic discourse often reflect discipline-specific norms and patterns of lexical preference (Noor & Anwar, 2019). It affects the organization and interpretation of information.

Lexical bundles shift the conventional, single-word analysis to a multi-word approach. It affects the language patterns used in different situations. Repetition is maintained by usage-based theories of language. The use of multiple words in a language indicates linguistic knowledge (Ellis, 2003). This concept helps the speaker to speak fluently and to use multi-word phrases in different discourses, either in complex discourse. Recent studies focus on second language use and how speakers employ multi-word expressions and different lexical bundles in different situations. This concept helps them to use language like native speakers (Biber, 2006). If a change is observed in the use of lexical bundles, it may focus on important perspectives in second language learning, such as proficiency and fluency. According to Tang and Zhang (2018), non-native speakers use a simpler, less frequent bundle, which affects how cohesive and natural the writing of speech is perceived.

By focusing on the frequency and function of lexical bundles, the researchers conducted a simple comparison of their frequencies. It examines the ways they use lexical bundles and their sequence of use. The dual focus is crucial because lexical bundles also serve communicative functions. For example, two lexical bundles may occur frequently in the same frequency pattern, but their discourse function is different and creates an effect on fluency and communicative purposes.

Current research used a functional classification of lexical bundles based on the framework by Biber et al. (2004). The functional bundles are classified according to their roles, and these can be divided into three major types, including stance bundles, discourse organisers, and referential expressions—a comprehensive investigation of lexical bundles on how the speakers use these bundles and their control of interaction. In the context of a second language, the functional role of lexical bundles is crucial. This research contributes to the growing body of knowledge and highlights the crucial use and structure of lexical bundles.

1.1 Research Questions

1. What is the distribution of frequency among different types of functional lexical bundles?
2. What is the influence of L1 on the usage of functional lexical bundles?
3. How do native and non-native speakers differ in their use of language functions, including stance, evaluation, and engagement?

This research identifies the differences in lexical bundles based on their function between native and non-native corpora. It highlights the broader issues of language competency and teaching; there are pedagogical implications that are shown by the results and discussion of the study. The findings of the research study could improve the theoretical model and enhance the understanding of lexical bundles and their exact usage.

2. Literature Review

Lexical bundles are the words that frequently occur together in a special language register (Bieber et al., 2004). Electrical bundles are used in both verbal and written discourse, and they are helpful for a deep study of language. The main purpose of this study is to compare the function variations in lexical bundles that emphasizes the results or outcomes in the field of discourse analysis and instructions of language. In the comprehension grammar of English, the idea of lexical bundles was coined by Biber et al. (2000). The recurrent word sequences that come frequently in a corpus are the lexical bundles, regardless of idiomatic and grammatical language (Biber et al., 2004).

Lexical bundles are central to discourse organisation and vary considerably across registers such as conversation, academic prose, and news writing. Lexical bundles such as stance expressions, discourse organisers, and referential expressions were categorised based on their functions. Researchers were better able to comprehend lexical bundles and how they are employed in texts to accomplish particular communicative goals thanks to this functional classification. The discussion about how this will impact the linguistic grazing ground was initiated by Swales (1997), focusing on the professional lives of students, teachers, and researchers from around the world. The terms monolingualism, linguistic hegemony, or imperialism have been used by some critical scholars to describe the spread of English in academia (Pennycook, 1994; Phillipson, 2008). On a very large scale, corpus linguistics has reshaped language by enabling researchers to examine authentic language use.

The recent research studies are conducted by Farhang Ju et al. (2024) on the introduction sections of research articles and on proximity construction in well-known TED talks by Wang and Csomay (2024). They examined the use of lexical bundles in different ways. Moreover, there is also a research study conducted by Kang et al. (2024), which shows the potential affordance of instruction-based lexical bundles in teaching propositions in the introduction section of journal articles. He demonstrated the pedagogical function of lexical bundles. In the introduction section of journal articles, Farhang et al. (2024) found that lexical bundles are linked to particular communicative purposes. They found that some bundles are more specific while others are more generic. Another significant finding was that, as compared to referential bundles, attitude bundles carried out a text-organising function.

Shin and Won (2024) analyzed that the generic distinction can be used to explain the variation in the utilization of lexical bundles. According to Saadatara et al. (2023), the forms and functions of lexical bundles in the writing section of the TOEFL exam reflect the preferences of highly successful test takers. So, the learners' proficiency had an equally proficient effect. The researchers

examined the low proficiency group of learners to employ less diversified lexical bundles with limited form and function.

In a corpus-driven research, Mbodj and Cortes (2025) advanced our understanding of lexical bundles in medical discourse by examining variations within disciplines across two registers. Their research covered a significant gap in the study of formulaic language. The research challenged preconceived thoughts about the consistency of academic discourse across fields and the depth of information. It clarified the adaptation of language resources in order to meet specific academic goals. Another study on lexical bundles was conducted by Alasmary (2022). In the study, the researcher compared two sub-corpora in the field of Psychology, focusing on four-word lexical bundles, and found through structural analysis that lexical bundles are used more frequently in the spoken register.

Comparing spoken and written discourse reveals functional differences in lexical bundles (Anwar & Maryam, 2017). Spoken language frequently contains bundles like *I do not know what* or *do you want to* that convey personal stance and interaction and real-time processing. Written academic texts, on the other hand, rely more on bundles that express logical relationships and arrange information, such as *on the other hand* or *as a result of*. So, it is clear that lexical bundles adjust according to their communicative needs. It was studied that in academic fields, different lexical bundles are used (Hyland, 2008b). The assessment and argumentation bundles are favoured by Arts and Social Sciences, but on the other hand, procedures and physical processes are favoured by Engineering texts. Different norms and objectives across special academic fields are reflected in functional variations in lexical bundles.

Another study was conducted about the use of lexical bundles by native and non-native speakers alters the language's functions. It has been studied that the language bundles for academic purposes written by native speakers and graduate students of Mexico were examined by Cortes (2004). She found that non-native speakers employ a smaller variety of bundles and, in their writings, overuse the simple stance bundle. It was observed that non-native writers use bundles improperly and repeatedly in academic discourse (Salazar, 2011). The focus of the study was on the relationships among three important points: the functions of flexible bundles, discourse awareness, and language proficiency.

Biber (2006) claimed that lexical bundles in different genres, such as academic, newspaper, and conversational, have different functions. He studied the effect of genre type on purpose and lexical bundles. Teachers can play a significant role in teaching students, which bundle can be used in different genres, making teaching students smooth and fluent. Academic writings can be enhanced by understanding the functions of bundle, structure, and meanings of text, Hyland and Tse (2007). Learners can gain a better understanding of how language is structured in authentic discourse by using corpus-based teaching materials that emphasize functional bundles in particular academic disciplines or genres. Humanities students may practice evaluative and argumentative sequences, while science students might concentrate on referential and procedural bundles.

These recurring word sequences are significantly influenced by communicative purpose, genre conventions, disciplinary norms, and the proficiency of the speaker or writer, according to research on functional variations in lexical bundles. While organizational and referential bundles are more commonly used in written and academic contexts, interactive and processing-related bundles are preferred in spoken language. Different fields' rhetorical conventions are reflected in disciplinary variation. Because learners frequently struggle with functional mastery, it is crucial to teach lexical bundles with both form and function in mind.

3. Research Methodology

This section of the research article explains the research approach and the methods used in the study. A mixed-methods research approach is used in the study to examine the most frequent lexical bundles in both native and non-native corpora. The data is analysed quantitatively and then interpreted qualitatively. The frequency and functional distribution of the lexical bundles that are found in both native and non-native corpora are the subject of the research study. The frequency-based difference of lexical bundles is presented in the form of charts and graphs. The main focus of the research is to compare the native and non-native corpora. It is methodologically appropriate because web-based corpora offer authentic and naturally occurring languages used in a variety of communicative settings. Mixed-method is used to compare the functional differences to provide the pedagogical dimensions and usage of lexical bundles according to different backgrounds.

3.1 Research Tool

Various data analysis tools are used in language teaching and Applied Linguistics. Over time, with the introduction of new tools, corpus-based analysis is conducted. In the research study, AntConc 3.5.9 is used. This software is used in the research because of its various capabilities, which include a concordance plot, tools for analyzing clusters, n-grams, a word list, a tool for displaying collocates, a keyword list, and a "KWIC" (keyword in context) format that illustrates the frequent usage of words and phrases in a corpus of texts. By employing the tool, 4-word lexical bundles are derived by processing data in the respective software along with their frequency and total number of words.

3.2 Theoretical framework

The study uses the theoretical framework of functional categories of lexical bundles presented by Biber et al. (2004. The framework explains the different lexical bundles based on their function. These include stance bundles, discourse organisers, and reference bundles. The framework is explained in the given table 1:

Table 1: *Classification of Lexical Bundles, Biber et al. (2004)*

Functional Classification of Lexical Bundles		
Stance Bundles	1. Epistemic Stance	I do not know what
	2.Attitudinal\Modality Stance	
	a) Desire	I do not want to
	b) Obligation\Directive	It is important to
	c) Intension\Prediction	I an't getting to
	d) Ability	To be able to
Discourse Organizers	1. Topic Introduction	If you look at
	2. Topic Elaboration/Clarification	On the other hand,
Referential Bundles	1.Identification\Focus	Of the things that
	2. Imprecision	Or something like that
	3. Specification of Attribution	
	a) Quantity of specification	There is a lot of
	b) Intangible framing	
	c) Tangible Framing	In the form of
		In the case of
	4.Time\place\text reference	
	a) Place reference	In the United States
	b) Time reference	At the same time
	c) Text deixis	As shown in the figure
	d) Multi-functional references	The beginning of the

Functional categories of lexical bundles are divided into three types: stance bundles, discourse organisers, and referential bundles. There are also other sub-categories of the main categories of lexical bundles. Stance bundles are the first major category of functional lexical bundles. These bundles show the stance of the speaker or writer towards any information. The first subcategory of stance bundle is epistemic stance, as it describes someone's desire to do something. There are two subcategories of stance bundles: epistemic stances and attitudinal stances. These types of bundles are about making judgments, expressing desires, and demonstrating abilities. The second type of lexical bundles under functional variation is discourse organizers. These are the bundles that structure information, make connections of the concepts, and also direct the readers and listeners to organise the flow of discourse. These bundles indicate the connection between various discourse elements, including elaboration, contrast, and sequence. The sub-categories of discourse organizers are topic introduction and topic elaboration. The third and important type is referential bundles. These lexical bundles are used to refer to something. It may refer to things, ideas, or some text. These bundles typically refer to abstract or physical things. These types of bundles also include subcategories, as mentioned with examples in the above table. The explanation clarifies the function of the framework used in the research study to compare native and non-native lexical bundles.

3.3 Research Sample

3.3.1 Global Web-Based English (GloWbe)

The corpus of Global Web-based English was created by Davies (2013). It is unique in that it allows you to compare different varieties of English. GloWbE is related to other corpora from English-Corpora.org, which are the most widely used corpora of English and which offer unparalleled insight into variation in English. GloWbE contains about 1.9 billion words of text from twenty different countries.

3.2 Data Scheme

The sample for this study is the non-native corpus, which contains data from the Pakistan, and the native corpus, which contains data from Britain. There will be a comparison based on frequency between the native and non-native corpus by using their functional categories.

Table 2: *Word Tokens and Types in both Native and Non-native Corpora*

Number of words in Native and Non-Native Corpora		
Native	Word Token	308201
	Word Types	21488
Non-Native	Word Token	133516
	Word Types	10254

There are two opposing trends in the size of corpora; none of the corpora in the world are compiled or found in the same size. On the one hand, the term "mega corpora" is used because these corpora are constantly expanding and contain hundreds of millions of words. In this research, the n-gram option is used to retrieve 4-word lexical bundles from the corpora. After fixing the range and frequency in the software, it provides the complete list of n-grams sorted by frequency. The file is processed, and 4-word lexical bundles are extracted from both native and non-native corpora. Below is the example image taken from the software while doing this whole process:

Figure1: *Procedure of Data Collection in AntConc*

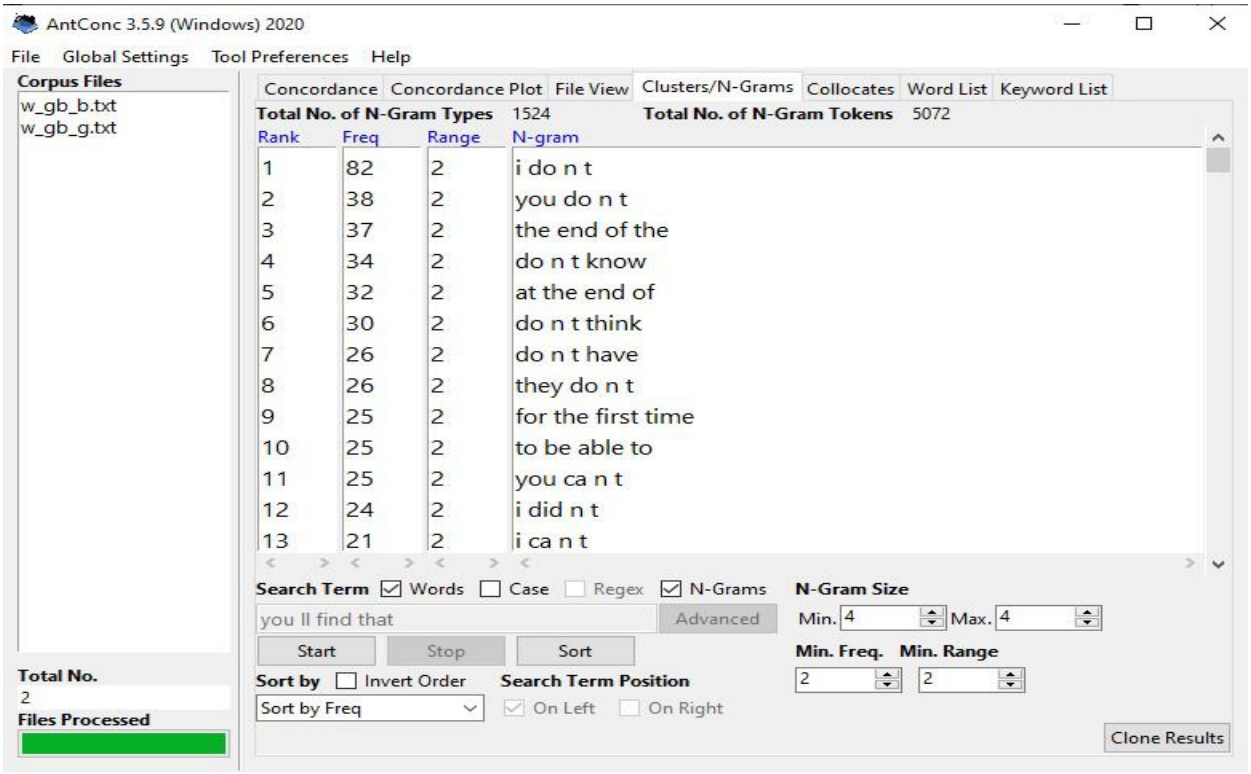

Figure 1 shows the number of files processed by the software, as well as the frequency, rank, and range of the 4-word lexical bundles derived using the software tool AntConc (3.5.9) for this process.

4. Data Analysis and Discussion

This section of the research deals with the data analysis and discussion. The section presents in detail quantitative and qualitative analysis of the 4-word lexical bundles. The bundles are classified according to the functional categories of lexical bundles presented by Biber et al. (2004). The study used the GloWbe corpus as a sample. It contains data from almost 20 countries worldwide. So, the sample that is selected for the research contains participants from native and non-native English-speaking countries. To identify differences among functional lexical bundles, the study used corpus analysis. By comparing lexical bundles and identifying the frequency-based differences, the research study provides insights into academic discourse competency and interlanguage development.

The frequency-based comparison of native and non-native functional lexical bundles is conducted through data normalization. Both native and non-native corpora have different sizes. So, data normalization is very necessary in order to make frequencies according to the dataset size. The frequency of each lexical bundle is normalized per 100 words because the sample is not much greater in size. In this way, the results are valid and can be compared with each other. Moreover, data normalization is very important because it shows the space of the lexical bundle among the larger data and the percentage of the space that is occupied by the specific expression.

The first major category of lexical bundles is stance bundles. Stance bundles include two types of functional categories: the first one is epistemic stance bundles, which highlight the probability, possibility, and the status of knowledge towards something. These are also central bundles. These

bundles support the speaker or writer to signal confidence and also express certainty towards specific things. The second subcategory of stance bundles is attitudinal or modality bundles. This subcategory expresses some desire, obligation, or intention towards something. These bundles express the writer's evaluative or affective attitude towards something. The table (3) below shows the frequency and normalized frequency of stance bundles that is found in native and non-native corpora:

Table 3: *Data Normalization of Stance Bundles in Native and Non-native Corpora*

Comparison of most frequent Stance Bundles among top-100 lexical bundles (4-word) in Native and Non-native Corpora GloWbe							
		Native			**Non-Native**		
Sr.	**Category**	**Lexical Bundles**	**Freq**	**Normalized Freq per 100 words**	**Lexical Bundles**	**Freq**	**Normalized Frequency per 100 words**
1.	**Epistemic Stance**	I don’t know that	4	0.4%			0%
	Attitudinal\Modality Stance						
1.		If you want to	14	1.4%	If you want to	2	0.2%
2.		Do you want to	5	0.5%	You want to be	2	0.2%
3.	**1.Desire**	You want to use	5	0.5%	Forum that you want	2	0.2%
4.		What we want to	4	0.4%	That you want to	2	0.2%
5.		Will be responsible for	2	0.2%	Of those who are	5	0.5%

6.	**2.Obligation\Direction**	You'll find that	2	0.2%	The best way to	4	0.4%
7.		Will come up with	2	0.2%	And there's no	3	03%
8.		When you add in	2	0.2%	Do not let your	3	0.3%
9.		To make the most	2	0.2%	From the section below	2	0.2%
10.		To keep in mind	2	0.2%			
11.		At the end of	32	3.2%	For the sake of	10	1%
12.		You will need to	12	1.2%	To be a part	4	0.4%
13.	**3.Intension\Prediction**	There will be a	6	0.6%	By saying that the	3	0.3%
14.		It is impossible to	6	0.6%	The views the	2	0.2%
15.		The rest of the	11	1.1%			
16.		to be able to	25	2.5%			

17.	**4. Ability**	will be able to	6	0.6%
18.		It is difficult to	5	0.5%

Table 3 presents the quantitative details of lexical bundles based on their functional categories of both native and non-native data. It is very important to mention that only 4-word lexical bundles are considered to compare native and non-native data. Normalized frequency is also mentioned in the table, which shows the space occupied by each lexical bundle. The first subcategory of stance bundles is epistemic stance, in which only one bundle is found, "I don't know that" as per the 100 most frequently used words. There is no epistemic stance bundle found in the non-native corpus. The second subcategory is attitudinal and modality stance. This subcategory is further divided into various expressions. As shown in the table 3, each is shown separately. Firstly, there appear desire bundles in the table. And the quantity of desire bundles is equal in native and non-native. However, there is a slight difference in normalized frequency. The first desire bundle *"if you want to*" appears the same in native and non-native corpora, but there is a difference in frequency. Secondly, there are obligation bundles that are found in both native and non-native corpora with different frequency patterns. In the same way, the other two categories, including intension and ability bundles, are found in the table. Each bundle within native and non-native corpora is mentioned with frequency and normalized frequency. Moreover, there are also lexical bundles that appear with 0. % frequency in both native and non-native corpora. For example, there are lexical bundles "*I don't know that" = 0.4%*, "*will be able to*" *= 0.6%* and also other many bundles with this type of frequency rate. So, zero frequency indicates that this specific lexical bundle lacks functionality in the discourse, and it is not among the high-frequency appearing bundles. But it really does not indicate that the lexical bundles with zero frequency is never occurred. It may be said that these types of lexical bundles are unable to reach the frequency threshold that is required to appear among the top-100 lexical bundles. This also does not indicate that the writer or reader is unable to use these types of stance bundles, but it may be suggested that the learner preferred other types of bundles to use. So, in the given corpora, these types of bundles lack functional prominence.

The second major category of functional lexical bundle is called a discourse organizer. A discourse organizer is used to organize the discourse in a professional academic way. It includes sub-categories that are named as: Topic Introduction and Topic Organization. Below is a table that highlights the major quantitative differences based on frequency between native and non-native corpora.

Table 4: *Data Normalization of Discourse Organizers in Native and Non-native Corpus*

Comparison of the most frequent Discourse Organizers among top-100 lexical bundles (4-word) in Native and Non-native Corpora GloWbe

		Native			Non-Native		
Sr.	**Category**	**Lexical Bundles**	**Freq**	**Normalized Freq per 100 words**	**Lexical Bundles**	**Freq**	**Normalized Frequency per 100 words**
1.	**1.Topic Introduction**	A look at the	6	0.6%			
2.		Have a look at	4	0.4%			
3.		To have a look	2	0.2%			
4.		It looks like it	2	0.2%			
5.		With a look at	2	0.2%			
6.		It seems likely that	2	0.2%			
7.		It seems as though	2	0.2%			
8.		Like the idea of	2	0.2%			
9.		On the other hand	14	1.4%			
10.		The other side of	11	1.1%			
11.		On the other side	7	0.7%			

12.		Other side of the	6	0.6%
13.	**2.Topic Elaboration**	At the other end	4	0.4%
14.		The other end of	4	0.4%
15.		Other end of the	3	0.3%
16.		In other words	3	0.3%
17.		Both sides of the	3	0.3%
18.		This side of the	3	0.3%
19.		To the side of	3	0.3%
20.		Either side of the	2	0.2%
21.		On both sides of	2	0.2%

Table (4) shows the comparison very clearly based on frequency. The results are diverse in nature and show a greater number of discourse organizers are used by native people. For example:

A look at the=0.6%

Have a look at=0.4%

To have a look=0.2%

It looks like it=0.2%

These bundles introduce the readers to the novel subject or introduce a specific topic and point of view. These types of bundles also provide a metadiscursive function to the discourse and provide an evaluative framework towards a specific point of view. The usage of discourse organizers implicates the capacity of the writer or reader to strongly and smoothly introduce the topic or issue, and it shows the strong orientation of the speaker and also discourse staging awareness of the speakers. Moreover, the usage of further inferential and perpetual frames shows the standards of advanced academic and informational sense of the user. Comparatively, non-native corpora show

the complete absence of discourse organizers. There is much diversity. The differences between native and non-native data highlight that variation in the discourse is a structuring strategy. The higher frequency containing bundle is *"on the other hand =1.4%"*. The other bundles comparatively with the frequency includes *"Other end of the=0.3%"* and "*it looks like it =0.2*%". the strong contrast of discourse organizers indicate that the native learners can handle the textual relationship more effectively. Natives are able to clearly identify alternatives and contrasting opposing points of view, and they can frequently engage in argument development and discussion.

The third major category presented by Biber et al. (2004) is referential expressions with many other sub-categories. The main function of referential expressions is to refer to something; it may be time, place, or another thing. By keeping in view the different nature of different places and things, the referential expressions are further subdivided into sub-categories. The frequency-based difference is shown in the given table below:

Table 5: *Data Normalization of Referential Expressions in Native and Non-native Corpora*

Comparison of most frequent Referential Expressions among the top-100 lexical bundles (4-word) in Native and Non-native Corpora GloWbe

			Native			**Non-Native**	
Sr.	**Category**	**Lexical Bundles**	**Freq**	**Normalized Freq per 100 words**	**Lexical Bundles**	**Freq**	**Normalized Freqper 100 words**
1.	**1.Identification\Focus**			0%			0%
2.		I'm not sure	10	1%	From the absence of	2	0.2%
3.		Or something like that	3	0.3%			
4.	**2. Imprecision**	And things like that	2	0.2%			
5.		Things like that I	2	0.2%			
6.		With the lack of	2	0.2%			
7.		To a lack of	2	0.2%			
8.		Due to a lack	2	0.2%			

9.	**3.Specification of attribution**						
10.	**a. Quantity of specification**	A lot of people	5	0.5%	Is one of the	5	0.5%
11.		A lot of things	5	0.5%	Was one of the	2	0.2%
12.		And a lot of	5	0.5%	Few in number and	2	0.2%
13.		Quite a lot of	5	0.5%	And there's no	2	0.2%
14.		And one of the	5	0.5%	Just a few of	2	0.2%
15		A lot of the	4	0.4%	And just a few	2	0.2%
16.		Got a lot of	4	0.4%			
17.		A lot of time	3	0.3%			
18.		A lot of businesses	2	0.2%			
19.		With at least one	7	0.7%			
20.		One of the things	4	0.4%			
21.		There were only two	2	0.2%			
22.		There is many p	6	0.6%			
23.	**b. Intangible Framing**	In the form of	10	1%	In terms of a	2	0.2%
24.		The form of the	3	0.3%			

25.		In search form	2	0.2%			
26.		Is the fact that	5	0.5%			
27.	**c. Tangible Framing**	In the case of	8	0.8%			0%
28.		Is not the case	3	0.3%			
29.		This is the case	3	0.3%			
30.		The case that the	2	0.2%			
31.		The way in which	5	0.5%			
32.		As an example of	4	0.4%			
33.		Is the fact that	5	0.5%			
34.	**4. Time\place\text reference**						
35.	**a. Place Reference**	At the university of	4	0.4%	Take the place	2	0.2%
36.		In the living room	3	0.3%	That the United States	4	0.4%
37.		At the house of	3	0.3%	In the United States	3	0.3%
38.		In the UK in	3	0.3%			
39.		In the country and	3	0.3%			
40.	**b. Time Reference**	For the first time	25	2.5%	Over the years and	2	0.2%

41.		At the same time	16	1.6%		
42.		A few years ago	7	1.7%		
43.		At the time of	7	1.7%		
44.		The first time i	5	1.5%		
45.		Over a period of	4	1.4%		
46.		A period of time	3	1.3%		
47.	**c. Text Reference**	Table shows that most	2	0.2%		0%
48.		From the selection below	2	0.2%		
49.	**d. Multi-functional References**			0%		0%

Referential bundles are very important in any spoken or written discourse. The main function of these bundles is to point out a specific object or place. The above table (5) clearly demonstrates the function-wise subcategories and their frequency-based differences between native and non-native corpora. Comparison reveals the systematic function between native and non-native corpora. Although referential bundles are used by both native and non-native corpora, native corpora show a greater functionality of referential bundles based on their usage. On the other hand, the non-native corpus exhibits a limited functionality of referential bundles.

Table 5 shows that the first category of referential bundles, which is identification/focus, is not found in both native and non-native corpora. This category is mentioned in table 5, but with empty boxes. The second major category of referential bundles that is shown in the table is imprecision. The use of imprecision bundles is different for both natives and non-natives. Natives employ imprecision as a useful tool in writing, while non-native speakers use it just for the sake of semantic needs. For example, it is used by natives as:

"*I'm not sure= 1%."*

And "*Or something like that=0.3%."*

But on the other hand, there is only one imprecision found in the non-native corpus that is *"From the absence of=0.3%".* The non-native usage of imprecision deficits framing with evolving interlanguage pragmatics. It indicates a more prepositional approach towards references.

The third category of referential bundles is called specification of attributions. It also includes three sub-categories that are mentioned in the above-mentioned table. The quantitative difference

is shown in the table. It shows that a higher range of this type of referential bundles is used by native speakers, while non-native speakers do not use these bundles at a much higher rate than native speakers. It indicates that natives heavily rely on the quantifying referential bundles. For example, the bundles used by natives are:

"*A lot of people=0.5%"*

"A lot of things=0.3%"

"A lot of time= 0.3%."

The frequent usage of these bundles by natives indicates their flexible quantification, audience-sensitive references, and discourse-driven approaches. On the other hand, the usage of referential bundles by non-natives is different from that of natives. They represent these bundles in a strict and rigid way. They use binary or categorical categorization, for example:

Is one of the=0.5%

Was one of the=0.5%

After these bundles, there are tangible and intangible referential bundles. Table 5 shows the comparative frequency and usage of bundles by both natives and non-natives. Natives use these bundles as an illustration of a critical perspective only. While non-natives continue to be confined to a more limited and conceptual framework. So, both the natives and non-natives imply these bundles from altogether different perspectives.

After these bundles, there are references that are shown in the table. These references are used to point out towards a specific time, place, or thing. The comparison shows the huge difference in using these references. Natives use the references towards their local place, for example, "In the UK in" and "at the university of" etc., while non-natives use these references to point out towards global entities like "*That the United States*". Then there come text reference bundles. Text reference bundles are used by natives. Only the native corpus has text reference bundles, which indicate advanced metadiscourse awareness and integration with academic practices. Lastly, there are multifunctional reference bundles that are neither found in the native corpus nor in the non-native corpus.

5. Conclusion

The research investigates the functional variations of lexical bundles presented by Biber et al. (2004). Functional variations include three types of bundles that are stance bundles, discourse organizers and referential expressions. mixed method strategy is used which employs both quantitative and qualitative analysis. The frequency measurement and distinctions are presented in the form of table while the discussion shows the qualitative differences. The analysis and discussion reveal that, overall, native speakers use lexical bundles more authentically. Non-natives usually rely on limited, irregular and frequently semantically set bundles. While natives used these bundles in a more balanced way and functionally integrated. These distinctions are qualitative rather than just quantitatively emphasizing basic variations and representation of meaning, stance, and reference bundles.

Natives use a wider range of epistemic and stance bundles, allowing for more complex articulation of assessment and ambiguity. On the other hand, non-natives exhibit a more restricted use in epistemic bundles. There are certain bundles that show 0 frequency, which indicates that these bundles do not come with higher frequency rates. It implies less use of probabilistic thinking and

less care in discourse. Then there are discourse organizers. In the native corpus, discourse organizer bundles are found frequently and prominently. These are used particularly for contrast, subject beginning, and topic introduction. On the other hand, non-natives rely on the sentence level while using these bundles. They do not use it to shape the argument; the usage in native corpora is altogether different. Lastly, in both corpora, the usage of reference bundles is seen, but the usage is different in their functionality. The non-natives are more constrained and less flexible in the usage of reference bundles. Whereas the native usage of these bundles is seen in a more diverse way and functionally integrated in the text. Native corpora highlight the advanced academic practices as compared to non-native corpora. Native corpora exhibit more authentic usage of time, place, and text references as compared to non-natives.

Theoretically, the findings support interlanguage theories of SLA and usage-based functions of lexical bundles. It claims that formulaic sequences acquire permanency through repeated usage. Non-native corpus highlights the limited usage and exposure that results in the under-representation of some functional lexical bundles. Furthermore, the findings of the study support the argument made by Biber et al. (2004) that lexical bundles have an important role in any discourse as they function as the building blocks of any discourse rather than just a stylistic representation.

Pedagogically, the research study explicitly emphasizes the significance of teacher lexical bundles. Especially, discourse organizers should be taken into consideration more. It is beneficial for the learner to learn the bundles as functional units. Therefore, it is important to include corpus-based lexical bundles in language learning and teaching within the EAP academic context. It may help to improve learners' confidence and language competence, with possible benefits for motivation. The future direction suggests making a generic base comparison, and there may be a comparison of different-length lexical bundles. There may be a longitudinal study based on the instruction and exposure of lexical bundles.

Conflict of Interest

The authors showed no conflict of interest.

Funding

The authors did not mention any funding for this research.

References

Alasmary, A. (2022). Academic lexical bundles in graduate-level math texts: A corpus-based expert-approved list. *Language Teaching Research, 26(1),* 99-123.

Anwar, B., & Maryam, S. (2017). Informational and interactional features of written and spoken Pakistani English. *Kashmir Journal of Language Research, 20*(1).

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2000). Longman Grammar of Spoken and Written English.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* John Benjamins.

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

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

Davies, M. (2013, September). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE).*

Ellis, N. C. (2003). *Constructions, chunking, and connectionism: The emergence of second language structure. The handbook of second language acquisition,* Blackwell.

Farhang-Ju, M., Jalilifar, A., & Keshavarz, M. H. (2024). Specificity and generality of lexical bundles in the rhetorical moves of applied linguistics research article introductions. *J. Engl. Acad.,* Purp. 69:101387. doi: 10.1016/J.JEAP.2024.101387

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes, 27*(1), 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International journal of applied linguistics, 18*(1), 41-62. *18*, 41–62. doi:10.1111/j.1473-4192.2008. 00178.x

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41(2),* 235–253.

Hymes, D. (1972). On communicative competence. *sociolinguistics, 269293,* 269-293.

Kang, S., Shin, Y. K., & Yoo, I. W. (2024). Using lexical bundles to teach prepositions to Korean EFL students. *J Second Lang Stud, 7,* 75–98. doi: 10.1075/jsls.00022.kan

Mbodj, N. B., & Cortes, V. (2025). Exploring Within-Discipline Variation of Lexical Bundle Use: An Example from Medical Research Articles and Medical Case Reports. *Revista Brasileira de Linguística Aplicada, 25(1),* e50012.

Noor, M., & Anwar, B. (2019). Structural and functional analyses of discipline-specific lexical bundles. *Pakistan Journal of Languages and Translation Studies, 7* (2).

Pennycook, A. (1994). Incommensurable discourses? *Applied linguistics, 15(2),* 115-138.

Phillipson, R. (2008). Lingua franca or lingua frankensteinia? English in European integration and globalisation. *World Englishes, 27(2),* 250-267.

Saadatara, A., Kiany, G., & Talebzadeh, H. (2023). Bundles to beat the band in high-stakes tests: pedagogical applications of an exploratory investigation of lexical bundles across band scores of the IELTS writing component. *J. Engl. Acad. Purp.*, 61:1012.

Salazar, M. (2011). A corpus-based study of lexical bundle use in a university EFL context. *Linguistics and Education, 22(4),* 310–323.

Shahid, U., Anwar, B., & Amin, M. W. (2025). Cleft Constructions in British and Pakistani English Newspapers: A Comparative Analysis. *Journal of Applied Linguistics and TESOL (JALT), 8*(2), 2385-2395.

Shin, Y. K., & Won, D. O. (2024). To what extent do L2 learners produce genre-appropriate language? A comparative analysis of lexical bundles in argumentative essays and speeches. *J. Engl. Acad. Purp*, 69:101389. doi: 10.1016/J.JEAP.2024.101389

Swales, J. M. (1997). English as Tyrannosaurus rex. *World Englishes, 16(3),* 373-382.

Tang, J., & Zhang, F. (2018). Lexical bundles and stance in L2 argumentative writing. Journal of Second Language Writing. *42*, 1-14.

Wang, W., & Csomay, E. (2024). Constructing proximity in popularization discourse: evidence from lexical bundles in TED talks. *Engl. Specif. Purp., 73,* 95–109. doi: 10.1016/j.esp.2023.10.003.