QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

A CORPUS-BASED COMPARISON OF LEXICAL BUNDLES FOUND IN RESEARCH PAPERS OF BRITISH AND PAKISTANI RESEARCHERS PUBLISHED IN THE FIELD OF ENGLISH LINGUISTICS

Muhammad Sumair Zahid,
Dept. of English Linguistics, IUB
Dr. Riaz Hussain,
Chairperson, Dept. of English Linguistics, IUB

Abstract

The art of persuasive writing is partially contingent upon the formulaic use of language. Lexical bundles, as formulaic sequences, play a crucial role in constituting persuasive and coherent discourse. Research on lexical bundles seems to have grabbed the attention of EAP instructors and researchers. Although it seems that less attention has been paid to the recurrent expressions found in the research discourses of non-native speakers. The current study compares the use of lexical bundles in research papers by British and Pakistani researchers in the discipline of English Linguistics, focusing on frequency, structure, and function. This study employed mix-method approach, incorporating a self-developed corpus of 1,086 research papers. This corpus was further classified into two sub-corpora, COBRA and COPRA. The corpus was further analyzed by using AntConc (version 4.2.4, Anthony, 2023), which generated two lists of four and five-word target bundles consisting of 158 bundles in COBRA and 229 in COPRA. For structural and functional analysis, Biber et al.'s (1999) and Hyland's (2008a) taxonomies were used. The findings indicated that both corpora were dominated by PP-based, NP-based, and VP-based bundles, with research-oriented bundles being the most frequent functional category. While British writers depicted the preference for NP + of bundles in locative and descriptive contexts, Pakistani writers overused them in procedural, quantification, and topical contexts. This study plays a significant role in understanding the formulaic use of language in research writing by focusing on cross-cultural variations in research discourse.

Keywords: Lexical Bundles, Recurrent Expressions, Corpus-Based Comparison, Academic Writing, Scientific Writing, English for Academic Purposes (EAP), Research Papers, and Corpus Linguistics

Introduction

English, a lingua franca, is primarily used in economic, cultural, and technological spheres (Abdelrady et al., 2025, 2026; Ma et al., 2024, 2025; Marline and Xu, 2018). The fact that 1.5 billion people speak English, with only 400 million native speakers (Dyvik, 2023; Ramzan et al., 2020, 2023, 2025), indicates that there are more non-native users than native ones. Swales (2004), Xu and Tang (2020), and Selvi (2021) denote that this impact can be simultaneously witnessed in the academic world (Akram & Li, 2024), giving ESL learners an opportunity to access international universities, conferences, and scholars. Today, the majority of international journals are published in English, and it is central to the academic exposure and the existence of the academic world (Flowerdew, 1999). According to Jeyaraj (2014), the majority of journals in the Scopus database are in English, which shows the significance of the English language for the dissemination of research and sharing of knowledge worldwide.

Although English enables the international involvement of scholars, it also poses a challenge to non-native scholars. Kuteeva (2023, p. 81) aptly observes that "English serves both as a bridge and a fence in the context of international knowledge exchanges". Even advanced learners often face difficulties in academic writing due to limited proficiency in genre-specific vocabulary, prefabricated structures, and word combinations. These challenges can compromise the clarity and persuasiveness of their work, regardless of subject expertise. Nevertheless, English remains the preferred medium of publication, and the number of papers written by the researchers who use English as a second language continues to grow (Hyland,

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

2006). Publications in international journals have become an important indicator of institutional and individual success, as most of the international and national universities prefer such standards to strengthen global rankings (Gonzales et al., 2014). Turning away from English is therefore not a viable solution; instead, non-native researchers must focus on enhancing their mastery of grammar, vocabulary, word combinations, and discourse conventions. However, grammar may be relatively easy to learn; mastering vocabulary, multiword expressions, and discourse conventions remains far more challenging because these vary across registers, disciplines, and genres (Biber et al., 1999, 2004; Hyland, 2008b; Chen & Baker, 2010; Ramzan et al., 2025).

For academic discourse, precise vocabulary and well-structured expressions are crucial in presenting arguments convincingly. Researchers employ technical terms, multiword units, and discourse conventions to make their writing engaging and credible, but non-native scholars often struggle with these features, particularly in the use of lexical bundles. Several studies addressed this issue earlier: Dontcheva-Navratilova (2012) highlighted the lack of attention to the linguistic needs of non-native writers; Gungor and Uysal (2016) and Ucar (2017) showed that non-native scholars tend to overuse or misuse lexical bundles compared to native researchers; and Nasrabady et al. (2020) confirmed the complexity of lexical bundles for second-language users. Likewise, highlighting the importance of genre-specific vocabulary for ESL learners, Bao and Liu (2022) also indicated several challenges across various disciplines. Corpus linguistics is one method to effectively address such challenges and provides systematic data-driven insights into the real use of language. It promotes both research and pedagogical innovations as well as assists learners to develop native like proficiency by emphasizing how language is used in particular situations. The discipline has gained extensive recognition, especially in the research on the application of lexical bundles to academic writing. Lexical bundles, also regarded as recurrent multiword expressions, are not simply word strings but essential building blocks responsible for bringing cohesion, coherence, and meaning in academic texts.

Coxhead and Byrd (2007) argue that these bundles can be significant for three reasons: they constitute a structural support of writing, the repetition needed to facilitate fluency and effective composition, and they combine vocabulary and grammar, which offers a lexico-grammatical structure of discourse. Recent studies (i.e., Siahposh & Varghaei, 2022) also contribute that lexical bundles assist writers in organizing writing, framing arguments, presenting findings, and grabbing the interest of readers, thus improving discourse. Notably, they have different functions and structures in different registers, genres, and disciplines, which highlights the necessity of teaching lexical bundles in academics. It is therefore essential for learners to master the balanced use of lexical bundles. These expressions in literature, frequently labelled as formulas, routines, or lexical phrases, are responsible for providing coherence, fluency, and contextual meaning in discourse. Examples include *"at the same time," "on the other hand,"* and *"as a result of"* (Cortes, 2004).

Lexical bundles are not only central in organizing arguments but also responsible for writing scholarly texts in compelling and interesting ways. Despite their prevalence in the oral discourse, lexical bundles also affect written academic texts, which increase clarity, cohesion, and disciplinary appropriateness. Professionally, academic writing is often judged by accurate use of bundles, and it differentiates a novice learner from a skilled writer. In the case of ESL and EFL students, it is especially crucial to master the use of lexical bundles because their proper use leads to native-like performance in written and oral communication. Cortes (2004)

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

observes that high-level learners, like university students, are supposed to learn how to proficiently use these recurrent expressions to become members of academic communities.

Based on these findings, it is critical to educate EAP learners and novice researchers about lexical bundles. In the case of non-native writers, especially in Pakistan, where publication in international journals is the measure of academic visibility, the ability to use lexical bundles effectively can be the determining factor. This paper thus investigates the significance of corpus linguistics in the development of native-like writing ability with special emphasis on lexical bundles and places the focus in the context of the problem that Pakistani researchers have to contend with. Although the present literature has discussed the lexical bundles in diverse registers and genres, little has been done in relation to their application in research papers, especially in comparative studies. Few studies investigate the variation between native and non-native researchers in English Linguistics, which leaves a gap in the literature. In an effort to fill this gap, the current study will compare the use of lexical bundles in research papers written by British and Pakistani scholars in order to determine the similarities and differences. This paper, besides showing how linguistic and cultural factors condition research discourse, helps to gain a more accurate insight into how ESL scholars can improve their academic writing. The results are likely to have useful implications for the Pakistani researchers, the developers of the EAP course, its instructors, and ESL learners who aim to enhance their involvement in the international academic communities.

Research Questions

1. Are there any similarities and differences found in the usage of lexical bundles by British and Pakistani researchers in the discipline of English Linguistics?
2. How do the lexical bundles vary in British and Pakistani research papers in terms of structure and function?

Significance of the Study

The present research holds significance as it will offer new findings to the existing research on lexical bundles in various contexts. Also, the research has practical implications for Pakistani scholars, educators, students, and curriculum developers by enhancing the readability, coherence, and persuasiveness of research writing as well as by enlightening the teaching of academic writing using corpus-based methods. The research can also assist Pakistani researchers in gaining more opportunities to be published in reputable journals by attracting attention to the trends of lexical bundle usage in internationally acknowledged studies. On the whole, it aims at developing the research culture in Pakistan and increasing the opportunities for global scholarly engagement.

Literature Review

The studies on lexical bundles contribute to the substantial evidence of their prominence in oral and written discourse. Altenberg (1998), Biber et al. (1999), Erman and Warren (2000), Biber et al. (2004), Cortes (2013), Biber and Barbieri (2007), Dontcheva-Navratilova (2012), Biber and Reppen (2015), Biber and Gray (2015), and Lorenna et al. (2020) demonstrated the wide occurrence of lexical bundles in everyday communication and academic language. Altenberg (1998), for instance, examined spoken texts in the London-Lund Corpus and demonstrated that lexical bundles are essential recurring patterns in daily conversation. Similarly, Biber et al. (1999) showed that three- and four-word bundles accounted for 28% of conversational language and 20% of academic prose. Erman and Warren (2000) further reported that lexical bundles constituted 58.6% of spoken discourse and 52.3% of written discourse in the LLC and LOB corpora. Following the line of inquiry, numerous studies Biber et al. (2004), Hyland, (2008a, 2008b), Mizumoto (2009), Ahmad et al. (2022), Amjad et al. (2021), Chen and Baker

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

(2010), Salazar (2011), Adel and Erman (2012), Gungor and Uysal (2016), Bychkovska and Lee (2017), Shin et al. (2018), Alamri (2017), Jiang and Hyland (2022), Akram & Abdelrady, (2023), (2025), Al-Adwan et al., (2022), Bao and Liu (2022) and Aziz (2022) investigated lexical bundles in academic and research contexts. Hyland (2008a, 2008b) explored bundles across disciplines. Later research extended this focus to research writing. Mizumoto (2009), for instance, analyzed a 31-million-word corpus from ten international applied linguistics journals to identify common lexical bundles. Chen and Baker (2010), Salazar (2011), and Adel and Erman (2012) extended the analysis to research papers and dissertations, emphasizing structural and functional variations. Gungor and Uysal (2016) investigated the dissimilarities between native and non-native writers in the use of bundles in research writing. Ucar (2017) compared three-word bundles in native and non-native Turkish papers with corpus-based methods, and Alamri (2017) linked genre-specific and corpus-driven approaches to study bundles in research papers, and Nasrabady et al. (2020) investigated their structural and functional roles in applied linguistics papers. Based on previous results, Varghaei and Khodadadi (2022) compared the use of lexical bundles in Iranian and international medical paper abstracts, and Shirazizadeh and Amirfazlian (2022) investigated their presence within theses, research papers, and textbooks. All these studies note the differences in frequency, structure, and use of lexical bundles in both native and non-native research discourses. But very little focus has been placed on Pakistani academic writing, especially the research articles, as compared to native English research articles. This study, therefore, aims to address this gap by examining similarities and differences in the use of lexical bundles between Pakistani and British researchers in English Linguistics.

Research Methodology

This study employed a mixed-method approach to compare the use of lexical bundles in research papers published by British and Pakistani researchers in the field of English Linguistics. According to Johnson and Christensen (2014), a mixed-method design integrates both qualitative and quantitative approaches, while Biber et al. (1998) propose that corpus-based studies often combine both approaches. Following that, the quantitative analysis was initially used to examine the frequency of lexical bundles (Biber, 2006), whereas qualitative analysis provided insight into the structural and functional roles of lexical bundles. The present study compiled two lists of meaningful lexical bundles and compared them in terms of frequency, structure, and function.

Description of Corpus

The present study employed a specialized corpus that was further classified into two sub-corpora, consisting of 519 British research papers and 567 Pakistani research papers. All of these research papers were published in reputable journals during 2019 and 2023 in the field of English Linguistics. Journal selection followed Nwogu's (1997) criteria, emphasizing quality, accessibility, and relevance to the discipline. JCR and HEC recognized Journals were preferred only. Purposive sampling was used for the selection of the research papers. The corpus covered three major domains, including Applied Linguistics, Stylistics, and Discourse Analysis. Table 1 further illustrates the sub-disciplines of the three domains of English Linguistics.

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

Table 1: Sub-disciplines in the Three Domains.

Applied Linguistics	Discourse Analysis	Stylistics
English language teaching	Critical discourse analysis	Literary stylistics
Language testing	Conversation analysis	Cognitive stylistics
Second language acquisition	Pragmatics	Comparative stylistics
Corpus linguistics	Discursive Psychology	Corpus stylistics

The overall corpus comprised 6.59 million tokens from 1,086 open-access research papers. Sub-corpus I, the Corpus of British Research Articles (COBRA), included 3.97 million tokens across 519 texts, while Sub-corpus II, the Corpus of Pakistani Research Articles (COPRA), contained 2.62 million tokens across 567 texts. Although the sub-corpora differ in size, such a difference is common in corpus-based research. According to Cortes (2004), Biber and Barbieri (2007), Chen and Baker (2010), Salazar (2011), Johnston (2017), Gil and Caro (2019), and Amjad (2022), these differences can be resolved by normalization of data. The description of the corpus is given in Table 2.

Table 2: Description of the Corpus

Sub-Corpus I			
Discipline	COBRA		
	No. of Texts	Tokens	Percentage
Applied Linguistics	182	1197718	30.13%
Stylistics	162	1060477	26.68%
Discourse Analysis	175	1716711	43.19%
Total	519	3974906	
Sub-Corpus II			
Discipline	COPRA		
	No. of Texts	Tokens	Percentage
Applied Linguistics	206	821239	31.30%
Stylistics	182	872386	33.26%
Discourse Analysis	179	929592	35.44%
Total	567	2623217	

Compilation of Corpus:

For the compilation of the corpus, IMRD's conventional structure was followed. The main body of research articles was included. All texts were manually processed, with extra information such as authors' names, affiliations, journal details, footnotes, tables, figures, references, and appendices excluded to maintain consistency across texts. The papers were cleaned, formatted, and converted into plain text files to make them compatible with corpus analysis tools. The compiled corpus was then processed using AntConc 4.2.4, a concordance and text analysis tool widely employed in corpus linguistics. This allowed for the identification and extraction of lexical bundles based on set frequency thresholds. The Corpus was also annotated using TagAnt 2.05 for structural analysis. The corpus was specifically designed to enable a systematic comparison of the frequency, structure, and function of lexical bundles in British and Pakistani research discourses, ensuring representativeness and reliability of findings.

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

Extraction of Lexical Bundles

After establishing the criteria for the identification of lexical bundles, i.e., size, frequency, and range of dispersion, both sub-corpora were processed in AntConc (version 4.2.4). Using the N-gram function, bundles were extracted in the form of frequency lists. Following Biber et al. (2004), the dispersion threshold was set at 2% of the total number of texts, ensuring that only recurrent sequences across multiple texts were included. Since the two sub-corpora differed in size, applying the same absolute number of texts as the range could have skewed the results; therefore, a normalized range was applied to maintain balance. In this study, only four- to six-word lexical bundles that occurred at least 15 times per million words and appeared in a minimum of 2% of the texts were considered for analysis. Table 3 indicates the normalized frequency and range applying to both sub-corpora.

Table 3: Normalized Frequency and Range

	Corpus	Token Size	Normalized Frequency (15 times PMW)	Number of Texts	Normalized Range (2% of total texts)
1.	Sub-corpus I (COBRA)	3,974,906	60	519	10
2.	Sub-corpus II (COPRA)	2,623,217	39	567	11

After the generation of lists, excessive bundles, i.e., overlapping bundles, short bundles, subsumptions, and irregular bundles were exempted using Zahid and Hussain's (2025) exclusion criteria. The implementation of all methodological steps confirmed that the extracted bundles represented frequently, repeatedly, and extensively found fragments, making them reliable markers of academic discourse. After minor modifications, Biber et al.'s (1999) and Hyland's (2008a) taxonomies were adapted for structural and functional classification of lexical bundles.

Results and Discussion

After extracting the lexical bundle lists from both sub-corpora, the data were organized by viewing frequency and range. Following the research objectives, the comparison was conducted at structural and functional levels to identify similarities and differences in the use of lexical bundles in British and Pakistani research articles. The comparison unfolded both shared and distinct lexical bundles between the two sub-corpora. The most frequent lexical bundle in the Corpus of British Research Articles (COBRA) was *"at the same time",* appearing with the frequency of 559 in 250 texts, followed by "*on the other hand"*, which occurred 510 times across 255 research papers. Comparatively, in the Corpus of Pakistani Research Articles (COPRA), the most frequently appearing bundle was "*on the other hand*", occurring 592 times in 285 texts, followed by "*with the help of"*, which appeared 343 times in 161 texts. It is important to note that the bundle *"at the same time"* also appeared in COPRA with a frequency of 230, specifying some overlap in usage patterns. Table 4 outlines a comparison of the top 25 frequently used target bundles retrieved from both of the lists. The bundles that were commonly found in both of the lists are bold.

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

Table 4: Comparison of Frequent Bundles

Sub-corpus I: COBRA				Sub-corpus II: COPRA			
Sr.	**Type**	**F**	**R**	**Sr.**	**Type**	**F**	**R**
1	***at the same time***	559	250	**1**	***on the other hand***	592	285
2	***on the other hand***	510	255	**2**	*with the help of*	343	161
3	***in the context of***	505	186	**3**	***in the form of***	333	178
4	***in the case of***	393	173	**4**	*is one of the*	258	176
5	***as well as the***	339	198	**5**	***at the same time***	230	140
6	***the ways in which***	323	113	**6**	***on the basis of***	228	117
7	***at the end of***	320	160	**7**	*the analysis of the*	227	147
8	*the use of the*	287	146	**8**	***as well as the***	214	132
9	***the end of the***	277	156	**9**	***in the context of***	213	130
10	*on the one hand*	258	139	**10**	*in the field of*	203	130
11	***it is important to***	254	156	**11**	*to find out the*	190	115
12	***on the basis of***	254	133	**12**	*the findings of the*	189	105
13	***in the form of***	249	142	**13**	***(through) + the use of the***	179	90
14	*in terms of the*	245	151	**14**	***at the end of***	172	100
15	*the extent to which*	224	114	**15**	*the results of the*	172	100
16	*as a result of*	221	147	**16**	***the end of the***	154	95
17	*the inmate in case*	209	10	**17**	*one of the most*	152	107
18	***through the use of***	197	107	**18**	*of the view that*	149	77
19	*in relation to the*	182	114	**19**	*in the process of*	142	86
20	*the way in which*	177	95	**20**	*that there is a*	137	98
21	*at the time of*	171	88	**21**	***in the case of***	135	86
22	*at the end of the*	171	103	**22**	***the ways in which***	134	38
23	*the fact that the*	164	122	**23**	*is based on the*	129	95
24	*can be used to*	161	104	**24**	***it is important to***	129	93
25	*inmates in cases and*	152	10	**25**	*in the light of*	128	86

Structural Analysis of Lexical Bundles

This comparative study of the British and Pakistani research discourses maintains that there exist similarities and differences in the application of lexical bundles in terms of structures. These results reveal that, despite the fragmentary nature of lexical bundles, they take certain and repetitive structural forms. The findings indicate that prepositional phrase (PP)-based, noun phrase (NP)-based, and verb phrase (VP)-based bundles are collectively over 74 percent and 79 percent of all the bundles of the Corpus of British Research Articles (COBRA) and Corpus of Pakistani Research Articles (COPRA), respectively. Table 5 highlights the structural distribution of bundles in detail.

Table 5: Structural Distribution

Category	Structures	Sub-Corpus I COBRA		Sub-Corpus II COPRA	
		Types	Types %	Types	Types %

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

1. NP-Based Fragments	Noun Phrase + of phrase fragment	27	17.08	44	19.21
	Noun phrase + other post-modifier fragments	10	6.32	11	4.8
	Other NP expressions	4	2.53	1	0.43
2. PP-Based Fragments	Prepositional Phrase + of Phrase Fragments	32	20.26	40	17.5
	Other Prepositional Phrase Fragments	34	21.52	44	19.21
3. VP-Based Fragments	Passive + Prepositional Phrase Fragments	4	2.53	9	3.93
	Other Passive Fragments	0	0	4	1.75
	Noun + Verb Phrase Fragments	1	0.63	12	5.24
	Other Verbal Fragments	1	0.63	4	1.75
	Copula Be + Adjective/Noun Phrase Fragments	4	2.53	12	5.24
4. Verb/Adjective/Noun + to/that Phrase Fragments	Verb/Adjective + To Phrase Fragments	3	1.9	2	0.87
	Noun or/and Verb + That Phrase Fragments	1	0.63	7	3.06
5. Adverbial/Adjectival Phrase Fragments	Adverbial Phrase Fragments	11	7	9	3.93
	Adjectival Phrase Fragments	1	0.63	3	1.31
6. Anticipatory It, That/There/To Fragments	Anticipatory it + Verb or Adjectival Phrase Fragments	10	6.32	14	6.11
	That/ There Expressions	3	1.9	6	2.62
	To + Verb/Noun Fragments	4	2.53	2	0.87
7. Other expressions	Conjunction and/but Fragments	3	1.9	2	0.87
	Clauses	1	0.63	1	0.43
	Other Expressions	4	2.53	2	0.87
Total		158	100	229	100

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

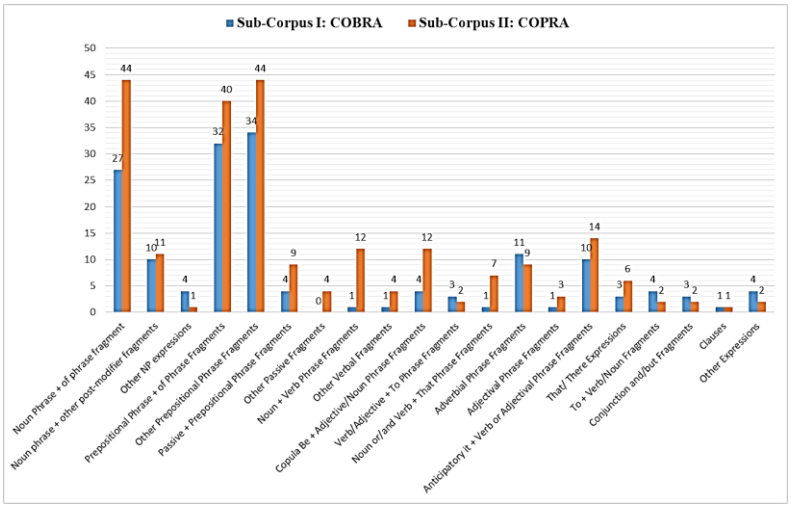

Figure 1. Structural Distribution of Lexical Bundles

The most common structural type in COBRA is PP-based bundles, which constitute 41.78% of all lexical bundles. COPRA, on the other hand, has a marginally smaller proportion of PP-based bundles (36.71%). The second most common in COBRA are NP-based bundles (25.93%), though they are somewhat less prevalent in COPRA (24.44%). Although Pakistani researchers (19.21%) employ NP + of phrase fragments more often than British researchers (17.08%), indicating that they prefer nominal fragments when writing scholarly texts. On the other hand, VP-bundles are more common in COPRA (17.91%) than in COBRA (6.32%), which reveals that Pakistani scholars are more likely to use verb phrase fragments in the context of the research. On the whole, these results demonstrate that although both groups share common structural preferences, Pakistani researchers have a higher propensity for verbal expression, and British researchers have a greater focus on prepositional and nominal fragments.

Additionally, structural analysis of lexical bundles shows that the majority of PP-based and NP-based fragments are variations of the frames of “the N of the, in the N of, at the N of, and it is Adj to/that”, which form almost half of the bundles of both sub-corpora. The frame “the N of the” comes out as the most frequently used one in both corpora, with 18 (11.39%) and 29 (12.64%) hits respectively in the Corpus of British Research Articles (COBRA) and the Corpus of Pakistani Research Articles (COPRA). This indicates that Pakistani scholars employ this structural pattern more often than the British one. The second most common frame “in the N of” upholds 9.49% of the bundles in COBRA and 8.3% in COPRA, which makes British researchers relatively more inclined to use it. The third significant frame, which is “it Adj to/that”, covers 4.43% of bundles in COBRA and is the least common in COPRA, indicating that it is not severely used among Pakistani researchers. These results indicate that British scholars show a more balanced and mixed structural application of lexical bundles, and Pakistani scholars show a greater tendency to rely on a concrete nominal structure like “the N of the”. The identified patterns are similar to existing studies that have also found out that the most common frames in academic writing are “the N of the” and “in the N of” (Biber et al., 2004; Hyland, 2008b; Chen and Baker, 2010; Johnston, 2017; Hyland and Jiang, 2018; Taieb et al., 2022). It can be notably observed that the preference for certain structural types reflects different stylistic and disciplinary conventions in academic discourse. These results can be used to influence the pedagogical practice by enabling novice researchers of Pakistan to improve their academic writing by using lexical bundles in an appropriate and balanced manner.

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

The structural comparison of lexical bundles also reveals significant insights into their semantic properties. Most target bundles in both sub-corpora are statements of abstract relations between propositions, including time, place, comparison, similarity, and text structure. The most common of these are PP-based bundles that are used in various sections of research articles to convey information, structure content, cite studies, and provide descriptions of procedural aspects. The relationships of context, domain, and discipline are also manifest in many of these bundles, and it is indicative of their central role in creating coherence and cohesion in academic discourse. Moreover, the descriptive and procedural information about participants or the research process is presented in a significant part of lexical bundles, which emerge as prominent structural features in both corpora.

These findings support existing literature on lexical bundles (Biber et al., 2004; Byrd & Coxhead, 2010; Simpson-Vlach & Ellis, 2010; Salazar, 2011; Johnston, 2017; Amjad, 2022; Aziz, 2022), which specify that even though lexical bundles are fragmentary, they follow certain systematic structural and semantic patterns. In the present research, these tendencies assist in revealing the unique features of the British discourse on research in the field of English Linguistics. This structural evidence sets the stage for the functional analysis of the following section, which highlights the pedagogical eminence of learning functional categories in EAP.

Functional Analysis of Lexical Bundles

The comparative analysis of British and Pakistani research discourses reveals both similarities and differences in the functional use of lexical bundles. These variations demonstrate that language operates through smaller units where lexical bundles exhibit distinct structural and functional patterns, each contributing uniquely to meaning-making in academic writing. While a structural pattern provides form, its function determines communicative purpose; hence, a single structure may serve multiple functions depending on the register, genre, or disciplinary context. For novice researchers and ESL learners, understanding and mastering this interplay is crucial for achieving native-like academic proficiency.

In terms of distribution, research-oriented bundles emerge as the dominant category across both sub-corpora. In the Corpus of British Research Articles (COBRA), they account for 46.2% of all lexical bundles, while in the Corpus of Pakistani Research Articles (COPRA), they constitute 50.6%. Text-oriented bundles make up 41.8% in COBRA and 42.8% in COPRA, whereas participant-oriented bundles are the least frequent, comprising 12% in COBRA and 6.6% in COPRA. These figures suggest that Pakistani researchers tend to overuse research- and text-oriented bundles, though a closer examination of their subcategories reveals subtle differences in usage and functional application. Table 6 illustrates the functional distribution of bundles in detail.

Table 6: Functional Distribution

Category	Function	COBRA		COPRA	
		Type	%	Type	%
1. Research-oriented bundles ***(structure activities and experiences of the world)***	Location	12	7.6	10	4.37
	Procedure	26	16.46	53	23.12
	Quantification	13	8.23	16	6.99
	Description	17	10.76	33	14.4
	Topic	5	3.16	4	1.75
2. Text-oriented bundles	Transition Signals	9	5.7	9	3.92

QUALITATIVE RESEARCH JOURNAL FOR SOCIAL STUDIES

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

(organize and present the text as a message/argument)	Resultative Signals	6	3.79	19	8.28
	Structuring signals	13	8.22	30	13.08
	Framing Signals	38	24.05	40	17.44
3. Participant-oriented bundles *(focus on the writer/reader of the text)*	Stance features	9	5.7	11	4.9
	Engagement features	10	6.33	4	1.75
	Total	158	100	229	100

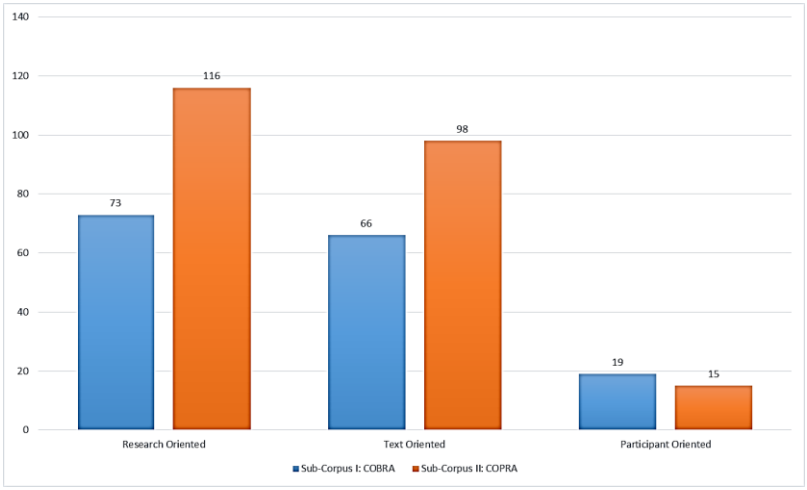

Figure 2. Functional Distribution of Lexical Bundles

The functional classification of COBRA reveals that research-oriented bundles dominate four of the top seven functions: procedure (26 types, 7.6%; 2116 tokens, 11.07%), description (17 types, 10.76%; 1525 tokens, 7.98%), quantification (13 types, 8.23%; 1408 tokens, 7.37%), and location (12 types, 7.6%; 2116 tokens, 11.07%). Among text-oriented bundles, framing signals represent the most frequent function, with 38 types (24.05%) and 5152 tokens (26.95%), followed by structuring signals with 13 types (8.22%) and 1141 tokens (5.97%). Participant-oriented bundles contribute a smaller but notable share, mainly through engagement markers (10 types, 6.33%; 1011 tokens, 5.29%). Collectively, these seven functions account for over 80% of the total lexical bundles (types and tokens) in COBRA.

In contrast, the functional classification of COPRA indicates a broader functional spread, with research-oriented bundles dominating four of the top eight functions: procedure (53 types, 23.12%; 3868 tokens, 22.1%), description (33 types, 14.4%; 1741 tokens, 9.94%), quantification (16 types, 6.99%; 1265 tokens, 7.22%), and location (10 types, 4.37%; 983 tokens, 5.61%). Text-oriented bundles also occur frequently, particularly framing signals (40 types, 17.44%; 3725 tokens, 21.27%), structuring signals (30 types, 13.08%; 1865 tokens, 10.7%), and resultative signals (19 types, 8.28%; 1487 tokens, 8.5%). Within participant-oriented bundles, stance markers appear with 11 types (4.9%) and 776 tokens (4.4%). Altogether, these eight functions represent more than 90% of the total lexical bundles identified in COPRA.

Building upon the previous section, an evaluation of the functional distribution across sub-categories reveals that framing signals are the most frequently used bundles in British research discourses, accounting for 24.05% of the target bundles. In contrast, procedural bundles

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

dominate Pakistani research discourses, constituting 23.12% of the total. Procedural bundles appear as the second most frequent category in the British corpus (16.46%), while framing signals occupy the second position in the Pakistani corpus (17.44%).

Beyond these functional distributions, notable structural variations are also observed between the two corpora. In COBRA, research-oriented bundles frequently employ NP + of fragments (procedure 11%, location 17%, quantification 30%, description 47%, and text 20%) and PP + of fragments (procedure 39%, location 42%). Text-oriented bundles mainly consist of PP fragments (transition 55.55%, resultative 33%, structuring 69.23%, framing 65.8%), while participant-oriented bundles often follow the anticipatory it + verb/adjective phrase structure, accounting for 40% of engagement and stance markers. In COPRA, a similar yet slightly varied distribution is evident. Research-oriented bundles also commonly use NP + of fragments (procedure 19%, location 10%, quantification 31.25%, description 36.36%, text 25%) and PP + of fragments (procedure 26.4%, description 21%, location 60%). Text-oriented bundles involve the use of PP fragments (transition 66.66%, resultative 10%, structuring 33.33%), whereas participant-oriented bundles are based on anticipatory it + verb/adjective phrases, which take into account 50% of the engagement and stance markers.

Such patterns demonstrate that various functional categories operate with different frequencies and structures, which proves the interdependence of the function and form. Research-oriented bundles are mostly used to demonstrate the research procedure, process, and purpose of research, as well as to indicate proportion, time, and place. Text-oriented bundles can be used for cohesive and interpretive purposes: to organize, relate, and place textual information into a specific context. Participant-oriented bundles, on the other hand, ensure the reader engagement in the text and maintain the relationship between the reader and the writer, which increases the authorial credibility and persuasiveness.

Still, having partial knowledge of academic writing conventions does not ensure proficiency. The lack of linguistic competence of most non-native or new researchers limits their capacity to carry out these functions efficiently. However, a similar approach to realizing these functions is provided by the use of lexical bundles, especially those that are aligned with certain structural patterns, i.e., NP/PP + of fragments and Anticipatory it frames. Mastering such structural and functional patterns may effectively contribute to EAP pedagogy, providing new researchers with linguistic strategies to improve their writing competency and ability to produce disciplinary discourse.

Conclusion

The present study conducted a comparative corpus-based lexical and functional analysis of research articles written by British and Pakistani scholars in the discipline of English Linguistics. The investigation aimed to identify the most frequent lexical bundles, their structural realizations, and their functional roles across the two corpora, COBRA and COPRA. The findings reveal that while both British and Pakistani researchers employ a range of lexical bundles to perform research-oriented, text-oriented, and participant-oriented functions, the distribution, structure, and frequency of these bundles vary across the two linguistic contexts.

In the British corpus (COBRA), framing signals were the most dominant functional category (24.05%), whereas procedural bundles (23.12%) were frequent in the Pakistani corpus (COPRA). These differences emphasize the influence of local academic conventions and linguistic preferences on the functional shaping of discourse. Structurally, both corpora demonstrated a strong tendency toward NP + of and PP + of fragments in research-oriented bundles and PP fragments in text-oriented bundles, while anticipatory it + verb/adjective phrase patterns were recurrent in participant-oriented bundles. These structural–functional

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

correspondences demonstrate that lexical bundles are not random but serve specific rhetorical and communicative purposes within academic discourse.

Overall, the results indicate that British researchers display more balanced and conventional use of lexical bundles across functional categories, whereas Pakistani researchers tend to overuse research- and text-oriented bundles. These patterns reflect both developmental and contextual differences in academic writing practices. The findings further suggest that the ability of non-native researchers is hampered while performing complex discourse functions due to restricted lexical repertoire and limited linguistic exposure.

Pedagogical Implications of the Study

Lexical bundles can be difficult for newbie researchers and ESL users (Cortes, 2004; Karabacak & Qin, 2013; Ucar, 2017; Bao & Liu, 2022; Amjad, 2022). Additionally, the appropriate use of lexical bundles can play a significant role in academic writing. New instructors can benefit from this research. Pedagogically, the study highlights the importance of raising awareness of lexical bundle usage in English for Academic Purposes (EAP) instruction. Novice and non-native researchers can learn to employ functionally and structurally suitable lexical bundles, thus enhancing coherence, cohesion, and disciplinary legitimacy in their writing by integrating corpus-based insights into academic writing pedagogy. Researchers may incorporate this comparative framework in other fields or genres to further explore cross-cultural academic discourse practices, linguistic patterns, and their implications for EAP curriculum design in the future.

References

Abdelrady, A. H., Akram, H., Mohamed, A. T. A., Abaalkhail, G. F., Osman, E. H. E., Mohamed, G. T. A. H., & Suliman, W. M. M. (2026). *World Journal of English Language, 16*(4), 311-321.

Abdelrady, A. H., Ibrahim, D. O. O., & Akram, H. (2025). Unveiling the Role of Copilot in Enhancing EFL Learners' Writing Skills: A Content Analysis. *World Journal of English Language, 15*(8), 174-185.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes, 31(2),* 81-92.

Ahmad, N., Akram, H., & Ranra, B. (2022). In quest of Language and National Identity: A Case of Urdu language in Pakistan. *International Journal of Business and Management Sciences*, *3*(2), 48-66.

Akram, H., & Abdelrady, A. H. (2023). Application of ClassPoint tool in reducing EFL learners' test anxiety: an empirical evidence from Saudi Arabia. *Journal of Computers in Education*, 1-19.

Akram, H., & Abdelrady, A. H. (2025). Examining the role of ClassPoint tool in shaping EFL studensts' perceived E-learning experiences: A social cognitive theory perspective. *Acta Psychologica, 254,104775*.

Akram, H., & Li, S. (2024). Understanding the Role of Teacher-Student Relationships in Students' Online learning Engagement: Mediating Role of Academic Motivation. *Perceptual and Motor Skills,* 00315125241248709.

Al-Adwan, A. S., Nofal, M., Akram, H., Albelbisi, N. A., & Al-Okaily, M. (2022). Towards a sustainable adoption of e-learning systems: The role of self-directed learning. *Journal of Information Technology Education: Re-search, 21,* 245-267.

Alamri, B. M. (2017). *Connecting Genre-Based and Corpus-Driven Approaches in Research Articles: A Comparative Study of Moves and Lexical Bundles in Saudi and International Journals*. https://digitalrepository.unm.edu/educ_llss_etds/81

ISSN Online: 3006-4686, ISSN Print: 3006-4678
Volume No: 03 Issue No: 02 (2026)

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.101–122). Oxford: Oxford University Press.

Amjad, M. (2022). English for official purposes: Exploring lexical bundles in the corpus of official documents of school and higher education departments in Pakistan (Unpublished doctoral thesis). The Islamia University of Bahawalpur, Pakistan.

Amjad, M., Hussain, R., & Akram, H. (2021). Structural and functional taxonomies of lexical bundles: an overview. *Harf-o-Sukhan*, *5*(4), 358-367.

analysis in EAP/ESP: Countering criticisms against corpus-based methodologies.

Anthony, L. (2023). *AntConc* (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Aziz, S. (2022). *Use of lexical bundles in academic writing in English by expert writers, native students, and non-native students in Applied Linguistics* (Unpublished doctoral thesis). University of Essex, England.

Bao, K & Liu, M. (2022). A Corpus Study of Lexical Bundles Used Differently in Dissertations Abstracts Produced by Chinese and American PhD Students of Linguistics. *Frontiers in Psychology*. 13. 893773. 10.3389/fpsyg.2022.893773.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97-116.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes, 26*(3), 263-286.

Biber, D., & Gray, B. (2015). *Grammatical complexity in academic English: Linguistic change in writing.* Cambridge University Press

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson Education Limited.

Bychkovska, T. & Lee, J.J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purpose*. 2017, 30, 38–52.

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

Chen, Y & Baker, P. (2010). Lexical Bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30–49.

Chen, Z., & Ramzan, M. (2024). Analyzing the role of Facebook-based e-portfolio on motivation and performance in English as a second language learning. *International Journal of English Language and Literature Studies*, *13*(2), 123-138.

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

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129-147. https://doi.org/10.1016/j.jslw.2007.07.002

Dontcheva-Navratilova, O. (2012). Lexical Bundles in Academic Texts by Non-native Speakers. *Brno Studies in English*. 38. 37-58. 10.5817/BSE2012-2-3.

Dyvik, E. H. (2023). The most spoken languages worldwide, 2023, *Statista*. Available at: https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/ (Accessed: 07 November 2023).

English for Specific Purposes, 24(3), 321–332.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text,* 20(1), 29-62

Flowerdew, J. (2005). An integration of corpus-based and genre-based approaches to text

Gil, N. N., & Caro, E. M. (2019). Lexical bundles in learner and expert academic writing. Bellaterra *Journal of Teaching & Learning Language & Literature*, 12(1), 65-90.

Gonzales, L. D., Martinez, E., & Ordu, C. (2014). Exploring faculty experiences in a striving university through the lens of academic capitalism. Studies in Higher Education, 39(7), 1097-1115.

Güngör, F & Uysal, H.H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*. 9. 176. 10.5539/elt.v9n6p176.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. New York: Routledge

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.

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing?. *International Journal of Corpus Linguistics*, 23(4), 383-407.

Jeyaraj, J. (2014). *English as 'THE' Publication Language of Science Research: Problems & Prospects.*

Jiang, F. K., & Hyland, K. (2022). "The datasets do not agree": Negation in research abstracts. English for Specific Purposes, 68, 60-72.

Johnson, R. & Christensen, L. (2014). Educational Research: Quantitative, Qualitative, and Mixed Approaches Fifth Edition.

Johnston, K. M. (2017). Lexical Bundles in Applied Linguistics and Literature Writing: A Comparison of Intermediate English Learners and Professionals. [An Unpublished Master of Arts (TESOL) thesis]. Portland State University, Portland. Retrieved from: https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=4491&context=open_access_etds.

Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. Procedia-Social and Behavioral Sciences, 70, 622-628. https://doi.org/10.1016/j.sbspro.2013.01.101

Kuteeva, Maria. (2023). Knowledge flows and languages of publication: English as a bridge and a fence in international knowledge exchanges. *Journal of English for Research Publication Purposes.* 4. 80-93. 10.1075/jerpp.22008.kut.

Lorenna, M., Fitriati, S., & Widhiyanto, W. (2020). The Comparison of Lexical Bundles in EFL Teachers' Talk between Non-native and Native English Teachers. *English Education Journal*, 10(1), 69-75. https://doi.org/10.15294/eej.v10i1.33830

Ma, D., Akram, H., & Chen, I. H. (2024). Artificial Intelligence in Higher Education: A Cross-Cultural Examination of Students' Behavioral Intentions and Attitudes. *The International Review of Research in Open and Distributed Learning*, *25*(3), 134-157.

Ma, D., Akram, H., & Li, S. (2025). Assessing the role of physical activity in shaping students' academic motivation: the mediating role of mental health. *BMC Public Health*.

Marlina, R. & Xu, Z. (2018). "English as a Lingua Franca" in The TESOL encyclopedia of English language teaching. eds. J. I. Liontas and M. DelliCarpini. 1st ed (Hoboken, NJ, USA: John Wiley & Sons). 10.1002/9781118784235.eelt0667.

Mizumoto, A. (2015). Corpus-based analysis of lexical bundles: Its potential applications in English language teaching. *Journal of the Japan Society for Speech Sciences*, 16, 30-34.

Nasrabady, P, Shirvan, M.E. & Golparvar, S.E. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages.*

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16(2), 119–138.

Ramzan, M., Akram, H., & kynat Javaid, Z. (2025). Challenges and Psychological Influences in Teaching English as a Medium of Instruction in Pakistani Institutions. *Social Science Review Archives*, *3*(1), 370-379.

Ramzan, M., Awan, H. J., Ramzan, M., & Maharvi, H. (2020). Comparative Pragmatic Study of Print media discourse in Baluchistan newspapers headlines. Al-Burz, 12(1), 30-44.

Ramzan, M., Azmat, Z., Khan, M. A., & un Nisa, Z. (2023). Subject-Verb Agreement Errors in ESL Students' Academic Writing: A Surface Taxonomy Approach. In *Linguistic Forum-A Journal of Linguistics*, 5(2), 16-21.

Ramzan, M., Javaid, Z. K., Fatima, N., & Kamran, M. (2025). Apprehension And Challenges In Lab-Based Language Education Using Online Applications. *Educational Research And Innovation*, *5*(4), 106-120.

Salazar, D. J. L. (2011). *Lexical bundles in scientific English: A corpus based study of native and non-native English writing* (Unpublished doctoral thesis). Universitat De Barcelona.

Selvi, B. (2021). English as the Language of Science: The Role of English in the Academic World. *The Academic Elegance*, 8(16). 1-21.

Shin, Y.K., Cortes, V. & Yoo, I. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing.* 39. 29-41. 10.1016/j.jslw.2017.09.004.

Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*. 49. 100946. 10.1016/j.jeap.2020.100946.

Siahpoosh, Hossein & Varghaei, Elshan. (2022). Bundle-Driven Stance and Engagement Metadiscourse Types in Spoken and Written Medical Discourse.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.

Swales, J. (2004). Research genres: Exploration and applications. Cambridge: Cambridge University Press.

Taieb, A., Montassar B., & Toumi, N. (2022). Exploring conventionalized phraseology in advanced academic writing of Tunisian linguistics students: a lexical bundles analysis. *Arab Journal of Applied Linguistics*, 6(02), 1–30. https://doi.org/10.1234/ajal.v6i02.315

Uçar, S. (2017). A Corpus-based Study on the Use of Three-word Lexical Bundles in the Academic Writing by Native English and Turkish Non-native Writers. *English Language Teaching*. 28(10) 28. 10.5539/elt.v10n12p28.

Varghaei, E., & Khodadadi, G. (2022). Comparing Lexical Bundles in Medical Research Article Abstracts of Iranian and Foreign Journals. *GEMA Online Journal of Language Studies,* 22(3). 86-102. 10.17576/gema-2022-2203-05.

Xu, W. & Tang, Yi. (2020). A Comparative Study of Linguistic Features of English Agricultural Journal Abstracts Written by American and Chinese Scientists. *International Linguistic Research,* 3(4). 94-111. 10.30560/ilr.v3n4p94

Zahid, M. S., & Hussain, R. (2025). Corpus for Extracting Lexical Bundles: Reviewing Methodologies and Applications. Journal of Applied Linguistics and TESOL, 8(1). 1731-1750.