ARTICLE

Sindh Journal of Linguistics (2024)
Volume 3, Issue 1–pp. 121-134
eISSN: 2957-9309
Published Annually by

DOI: 10.58921/sjl.v3i1.22

Analyzing the Functions of Lexical Bundles for Teaching Academic Writing to Graduate Students

Tanveer Hussain Mangnejo[1]
Akhtar Abbas[2]

Abstract
Academic writing skills contribute to academic performance. One of the critical elements of good academic writing is to know the use of the strings of the words that usually occur together. These strings, known as lexical bundles, help understand and reproduce the academic discourse. Some research has been conducted to examine the use and functions of lexical bundles in spoken and written discourses (Chen & Chen, 2020). However, there is scarce research to analyse the functions and use of lexical bundles that can help novice researchers to write research articles as per academic standards in the Pakistani context. This study aims to investigate the use and distribution of lexical bundles in academic writing of Pakistani researchers. The corpus of 90201 words was developed from 12 research papers by Pakistani researchers. The list of the lexical bundles by Simpson-Vlach and Ellis (2010) was used as a source to interpret the lexical bundles. We found the referential expressions were the highest lexical bundles in research articles of Pakistani authors, followed by the expressions of ability and possibility and hedges in the list. Based on these findings, we argue that Pakistani researchers are aware of various types of lexical bundles and their functions in academic writing. It is suggested that the high-frequency bundles can help the learners improve using modal words as hedges in academic writing. As per the findings of this study, it is recommended that a large corpus may be built for better results in the future.
Keywords: academic writing, lexical bundles, corpus, frequency, function

[1] Assistant Professor at Institute of English Language & Literature, Shah Abdul Latif University Khairpur & PhD Scholar at Air University Islamabad
[2] Associate Professor, Department of English, University of Sialkot, Sialkot

Introduction

The recurrent expressions, mostly called lexical bundles, and their functions have been studied extensively across the globe. They play an important role in speech and writing in general and academic writing in particular. This study has provided a list of the lexical bundles commonly found in Pakistani research articles. The lexical bundles are multi-word combinations often occurring together as a unit that serves pragmatic and discoursal purposes (Chen & Baker, 2010). These units are very complex from the terminological perspective. Wray (2003) also comments on the complexity of the terminology. He explained that some scholars use these terms as synonyms and others use these terms for different notions. For example, phrasicon, formulaic sequence, clusters and recurrent word combinations, n-grams, and lexical bundles are used for a similar concept. However, Gray (2016) distinguishes these lexical bundles based on some defining characteristics, such as their commonality based on the distributional criteria, their nature to often bridge two units, and their non-idiomatic meaning. Biber et al. (1990) define lexical bundles as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status (p.990)". Lexical bundles are a form of extended collocations (Biber et al., 1999). These writers differentiated the lexical bundles in a spoken language, such as "do you want me to, I said to him,…" in academic prose, such as "in the case of the, there was no significant…: (p.989). Some recent research has illustrated that lexical bundles are present in spoken and written academic discourse (Wright, 2019). Academic discourse is one of the areas of interest recently studied by researchers (Bal-Gezegin, 2019).

Academic writing is a very complex term. Writing from classroom assignments to term papers comes under the umbrella of academic writing. Arnaudet and Barrett (1984) suggest that academic writing should objectively express ideas and not include personal opinions. Lester (1993) argues that academic writing should be written in the third person voice, and personal references such as "i, we" must be avoided. However, Hyland (2000) believes that various disciplines have different conventions about the objectivity and use of language. He argued that academic writing is not a separate, undifferentiated body of mass that must be viewed as a single entity but based on subject-specific practices (Hyland, 2002).

Use of Lexical Bundles across Genres

Lexical bundles are groups of words that commonly co-occur in any given genre and are investigated by the researchers by using statistics of the words rather than relying on the researchers' intuitions (Cortese, 2004). Neely and Cortes (2009) defined lexical phrases as the chunks of language use that mostly co-occur. For the study of lexical bundles, the researchers arbitrarily set out the cut-off criteria (Nesi & Basturkmen, 2006). Most of the studies have explored three-word lexical bundles in academic writing. Another way to avoid the researcher's idiosyncratic choices is to look for the dispersion plot for the lexical bundles (Hyland, 2008). Several studies have studied lexical bundles across various genres, such as research articles, lectures, thesis, and so on (Chen-Yu Liu, 2020).

Lexical Bundle and their Discourse Function

The academic registers and genres have lexical bundles that perform a particular sort of function in them. Biber et al. (2004) categorise the lexical bundles into taxonomies based on their functions in academic discourse. Biber et al.'s taxonomy has been adopted by many researchers (Chen-Yu Liu, 2020). This taxonomy is divided into four main types: Stance markers (the speaker's attitudes and opinions), discourse organisers (that outline the texts and introductions and elaborate on the topics of the interests), referential bundles (that highlight time and place and specify focus and quantity) and special conversational functions(such as politeness, report). This research has investigated the use of referential expressions and stance markers proposed by Simpson-Vlach and Ellis in research articles by Pakistani authors. We have delimited the scope of the study only to referential bundles by Simpson-Vlach and Ellis. Even the Simposon-Vlach and Ellis' taxonomy has not been fully used in this study due to a shortage of time. The other taxonomies are also important; many studies have been conducted using them. Based on the above literature review and background study of the topic, this study aims to identify the most frequent lexical bundles in Pakistani scholars' academic writing and highlight their discourse functions. Furthermore, the current study also focuses on the dispersion of Pakistani Scholars' lexical bundles in academic writing.

Methods

The researchers developed this study's corpus by collecting twelve articles written by Pakistani researchers in Applied Linguistics. The papers were published in HEC-recognized Y-Category Journals from 2018 to 2020. The research articles were selected purposively. Purposive sampling "refers to a group of non-probability sampling techniques in which units are selected because they have characteristics that you need in your sample. In other words, units are selected on purpose" (Nikolopoulou, 2023, what is purposive sampling section). The researchers selected articles that met their research objectives, i.e., the first author of this study selected three journals from social sciences and articles related to applied linguistics, which were between 5000 to 9000 words and published from 2018 to 2020. Though purposive sampling is not used to show representativeness, its characteristics signal the diversity of selection that adequately represents the sample selected for the study. The four-word expressions must meet the threshold level to be counted as a lexical bundle. Some previous studies set 10, 20, and 40 frequency threshold levels at least in five and 10 texts per million words (Biber et al., 2004; Hyland, 2008). This study used a more conservation threshold of 3 occurrences in two corpus sections per million words after normalisation. Perez-Paredes (2020) argues that "there is no small or big corpus" (p. 51). The context and the purpose of building a corpus matter. Therefore, this study used a small specialised corpus that may be normalised to a 1000000 base. The normalisation process divides the raw frequency by the total number of words in the corpus, multiplying it with the normed corpus. In this case, the raw frequency of 3 is divided by the number of the tokens in the corpus, i.e., 90000, and it is multiplied by 1000000. The resulting normed frequency will be a relative frequency per million words. Therefore, in this study, the normed frequency occurrences are

3/90000*1000000= 33, thirty-three, which are the same cut-off levels (Cortes, 2004). The common "normalised bases include number of words or lemmas per 1,000, 10,000, 100,000 or 1,000,000 words" (Perez-Paredes, 2020, p. 55).

We followed the norms of corpus cleaning by removing all unnecessary material from the research articles, such as headers, footers, page numbers, and citations. The cleaned corpus was uploaded to AntConc 3.5.7, a corpus research tool (Anthony, 2018). We generated the lexical bundles and manually selected those related to our study. We also generated the dispersion plot of the lexical bundles found in this study. However, we have presented only two figures for dispersion plots generated through the AntConc 3.5.7 corpus tool.

Table 1

Specialised Corpus of Pakistani Research Articles from Applied Linguistics

S.No	Word Count of each paper
1	6,025
2	5,917
3	8,478
4	7,218
5	6,996
6	7,989
7	8,261
8	7,945
9	8,369
10	8,233
11	7,981
12	6,789
The total number of tokens in the corpus	90201

The researchers created the corpus from three research papers published in Y-Category HEC-recognized Pakistani Journals. All the selected papers were from the discipline of English Linguistics. The total tokens of the corpus were 90201, and the frequency of the lexical bundles was extracted from the corpus.

The frequency and distribution of lexical bundles were also explored in this study. This study used four-word lexical bundle lists. The inclusion criteria by Biber et al. (2004) was that if a lexical bundle occurs 40 or more than 40 times in the corpus is included. However, this study included fewer occurrences of the lexical bundles that are a minimum of 3 times out of the whole corpus. Simpson-Vlach and Ellis' (2010) functional taxonomy has been used to interpret

the study results. We used this taxonomy because it served our intention to teach lexical bundles in academic writing. In addition, this taxonomy is also detailed or extensive in forms and functions performed in discourse. We used two group categories from his list to compare our results. These two categories include Referential expressions and Stance markers. However, their list includes three categories, the third being discourse organisers. We excluded this category for the next research project to be conducted in detail.

Figure 1

Functional Taxonomy of Lexical Bundles by Simpson-Vlach & Ellis 2010 (Liu & Chen, 2020)

/

Referential bundles	Specification of attributes
Chart Area	- Tangible framing attributes - Tangible framing attributes - Quantity - Quantity Identification and focus
	Vagueness markers
	Deictics and locatives
	Contrast and comparison
Stance expressions	Epistemic stance
	Obligation and directive
	Intention/volition, prediction
	Expressions of ability and possibility
	Evaluation
	Hedges
Discourse organizers/Discourse organizing functions	Topic introduction and focus
	Topic elaboration
	- Non-causal
	- Cause and effect
	Metadiscourse and textual reference
	Discourse markers
Special Conversational functions	

Following are the research questions that this research has answered:

1. What are the most frequent four-word lexical bundles present in the academic writing of Pakistani Scholars?
2. What discourse functions do these lexical bundles perform in the academic writing of Pakistani Scholars?
3. How are these lexical bundles distributed in the corpus of academic writing of Pakistani Scholars?

Results and Discussion

Answering research questions one and two for the high frequency and discourse functions of the four-word lexical bundles and their similarity with Simpson-Vlach and Ellis' list, 78 high-frequency lexical bundles were extracted from the corpus of twelve research articles. These 78 lexical bundles were then compared with Simpson-Vlach and Ellis' list. It was found that Simpson-Vlach and Ellis' list was more comprehensive and detailed, including three, four, and five words in lexical bundles. However, their list includes seventy-four-word lexical bundles,

and our study found 78 four-word lexical bundles even in a very limited corpus compared to the corpus used by Simpson-Vlach and Ellis in their study. These findings are very interesting and meaningful because four-word lexical bundles have a variety of functions, are more complete syntactic structures than three-word bundles and are more understandable semantically. These findings suggest that our lists of bundles may be more helpful for pedagogical purposes in teaching academic writing to undergraduates. The teachers may use this list to teach the syntactic structure of academic writing, significantly to help the learners create their stance and want to see the time and place of the academic writing texts. This list is also likely to help the learners with semantic interpretations and may help them understand the discourse functions of academic writing.

Table 2

The Types of Referential Expressions in Pakistani Research Articles

Referential Expressions	Texts	Occurrences
In terms of the	Text-1+12	1+2=3
As a function of	Text-1 & 4	3
How	Text-3, 6, 7, & 10	1+2+2+4+9
in the case of	Text-1, 11 & 3	2+1+2=5
Identification and Focus		
Nill		
Comparison and contrast		
more than half	Text-1, 6, & 4	1+1+1=3
Deitics and locatives		
This study shows that	Texts-1, 4, & 8	2+3+1=6
This study aims to	Texts-5-7-12	1+1+1=3
that are common in	Texts-4, 7, & 9	2+1+1=4
Vagueness markers		
and so on the	Texts-1-3	1+3+4

Based on our findings, it seems reasonable to suggest that genre-specific instructions may be helpful as these lexical bundles differ in different academic genres, such as books, research articles, theses, and academic lectures. Tale 2 precisely consists of referential expressions only. There are different categories of referential expressions in the original list by Simpson-Vlach and Ellis. However, for the sake of ease and precision, we have separated the lexical bundles into different tables. Table two presents the various types of referential expressions based on the original list of the lexical bundles in Simpson-Vlach and Ellis. Twenty referential expressions form a significant part of these expressions. The other four types of expressions are less than the collective referential expressions. The lexical bundles such as in terms of, as a function of, how, and in the case of, refer to something present in the text. The second type of referential expression is identification and focus. There were zero occurrences of identification and focus of the four-word lexical bundles. This may imply that Pakistani

authors are more objective and focus less on identifying as authors of the research articles. This also implies that the traditional concepts in research article writing compel the authors to avoid showing their identification and focus on their work. Only one lexical bundle for the referential expressions of comparison and contrast was found in the corpus. The expression of more than half occurred 3 times in texts 1 and 4 of the corpus. This finding suggests that the corpus may lack papers based on the comparison and contrast studies. The two lexical bundles were based on the deictic expression and locations. These two bundles occurred ten times in the corpus. This may suggest that Pakistani authors use more location expressions and deictic expressions to refer to ideas, concepts, or objects in the world. The vagueness markers occurred only once in the corpus. This finding suggests that the authors want to sound clear and robust in their claims.

As indicated in Table 2, referential expressions are the most frequent lexical bundles. Simpson-Vlach and Ellis categorise referential bundles into five subcategories the specification of attributes (quantity specification, framing attributes, tangible and intangible attributes), identification and focus, comparison and contrast, deictics and locatives, and vagueness markers. Table 3, given below, summarises the bundle types.

Table 3

Types of the Bundles of Each Subcategory for Referential Bundles

Category	# of Types	# of Tokens
Specification attributes	20	90201
Identification & focus	Nill	90201
Comparison & contrast	3	90201
Deictics & locatives	10	90201
Vagueness markers	4	90201
Total	37	90201

Table three summarises the number of lexical bundles taken from the whole corpus. However, the original subcategories were five in number, and specification attributes were further divided into other subcategories. However, we have included only the main category of specification attributes, not its subcategories, for ease for the instructors, teachers, and students. This result suggests that teachers may use these different referential expressions to teach the academic language by comparing, specifying, or elaborating on the texts, genres, ideas and concepts. In contrast to the Chen-Yu Liu (2020) and Simpson-Vlach and Ellis' (2010 lists of lexical bundles, our findings revealed that Pakistani authors use fewer identification and focus markers. Our findings show that these markers have zero occurrences. However, the above-cited studies have 15 identification and focus lexical bundles occurrences. This study shows that the comparison and contrast lexical bundles are second last as the frequency of the occurrences in the research articles by Pakistani authors. The last occurrences in the above-cited studies are of vagueness markers. However, this study shows that there are 4 occurrences of vagueness markers.

Table 4

Types of Stance Markers in Research Articles in Pakistani Research Articles

Stance markers		
Hedges	Texts	Occurrences
A small number of	Texts-1-7, & 10	2+2+15
Epistemic stance		
It seems that the	Texts-1-4 & 5	2+3+1=6
I don't know whether	Texts-1-3-4	1+2+2+5
Obligation and directive		
To make sure that	Texts-3	4
Expressions of ability and possibility		
must be done in	Texts-1-3 & 8	3+2+3=7
Would be possible to do	Texts-1-3	4+4+8
Can be done that	Texts-3-4	2+4=6
Intention/volition, predicon		
Nil		

Table 4 consists of stance markers extracted from the corpus of the research articles by Pakistani authors. Under this category, the first subheading is hedging. There is only one lexical bundle that occurred five times in the corpus. Two lexical bundles for epistemic markers occurred eleven times in the corpus. The corpus has only one lexical bundle for obligation and directive stance markers. The ability and possibility subcategory expressions have three lexical bundles that occurred 21 times in the whole corpus. There are zero occurrences of intention/volition and prediction stance markers in the corpus of Pakistani researchers' papers. This finding contrasts the Chen-Yu Liu (2020) study in which the highest frequency bundles are intension, volition, and prediction markers. Our study found the highest number of lexical bundles under the subcategory of expression of ability and possibility. The function of this subcategory is to indicate that the authors are capable of doing something. The model verbs found in the study contrast Biber's (2006) proposition, which states that model verbs are primarily found in spoken language (Chen-Yu Liu, 2020). The scholars use these model verbs' stance markers in their writing. The function of these models may also give direction for further research. The lexical bundle must be done in a way that suggests some direction by the researchers for the readers or future researchers. The least number of stance markers in our study are obligation, directives, and hedges. These support the findings presented by Chen-Yu Liu. The finding suggests that Pakistani researchers also use hedges and epistemic markers, as the Chen-Yu Liu and Simpson-Vlach and Ellis studies have found. Our findings support the above-cited results.

The Simpson-Vlach and Ellis' (2010) list of lexical bundles is very elaborate and extensive and includes one other type of category: discourse organising functions. However, we have

included this category in our study. We have delimited our study to the first two categories only.

The following figure presents the results of research question number 3 about the distribution of the lexical bundles in the Pakistani researchers' corpus.

Figure 2

Distribution of Referential Expressions in Research Articles

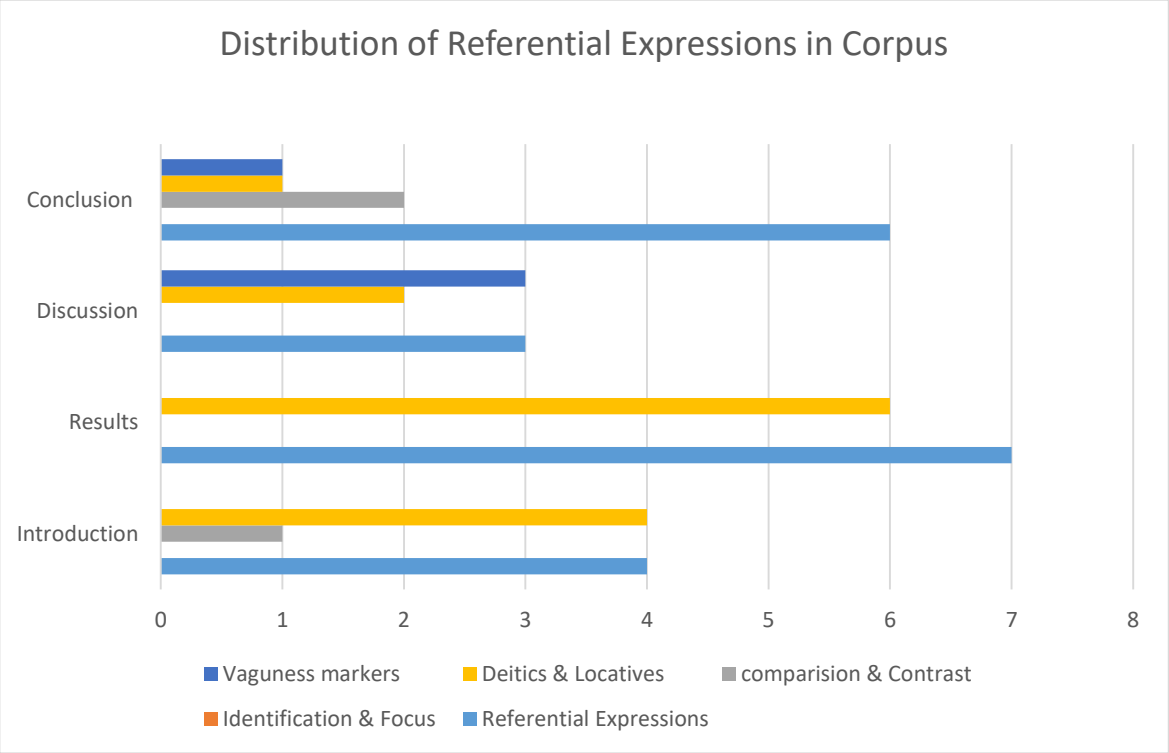

Figure 1 shows the distribution of the lexical bundles of Referential Expressions according to Simpson-Vlach and Ellis' list of the lexical bundles. Pakistani authors used the highest number of referential expressions in the results sections. However, the lowest number of referential expressions have been used in the discussion section. The high frequency of the referential expressions in the results sections illustrates that authors are more concerned with referring to the ideas and abstractions of the concepts presented in the results sections. The second highest number in the conclusion sections also illustrates the referential indication of the concepts discussed in these sections. Unlike Simpson-Vlach and Ellis's (2010) list of identify and focus lexical bundles, which has 14 occurrences, this study shows zero identity and focus expressions. This finding suggests that Pakistani authors are less concerned about their voice or identity in their academic writing. Only three occurrences of the lexical bundles of comparison and contrast have been found in this study. One has been found in the introduction and 2 in the conclusion sections. This finding is important as it suggests that Pakistani authors may rarely compare and contrast their results with previous similar studies. This finding also suggests that the author might not know how to compare and contrast their results with other studies or are least concerned about what others have said in the field.

Figure 1 shows that the highest number of deictic and locative lexical bundles are used in the results sections of the corpus. The expressions in this study suggest that the authors want to relate to their results through these referential expressions. These bundles function as logical connectors or cohesive devices, which may add clarity and precision to the study. These lexical bundles also refer to the particular aspects of the argument or research of the Pakistani authors.

Figure 3

The Distribution of the Stance Marker in Research Articles

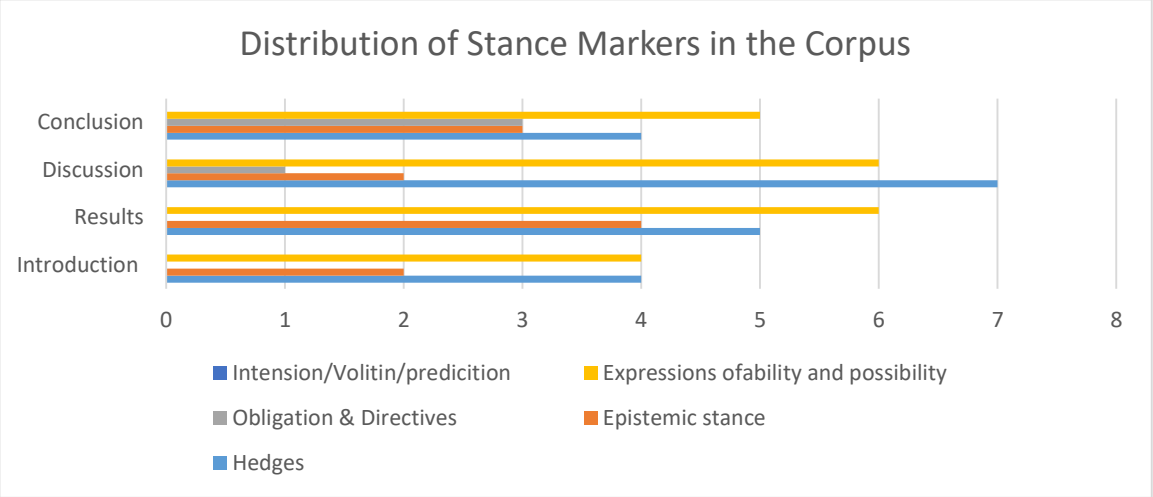

Figure 2 shows that the highest number of the stance markers in discussion sections of the Pakistani authors are hedging. This highest number of hedges in the discussion section suggests that Pakistan authors are less direct and avoid more substantial claims in their research articles. The use of such hedging, as indicated in Figure 2, suggests that Pakistani authors fulfil the discourse function of modesty and avoid overgeneralising their claims. Their claims are modest. This study supported the findings concerning hedges and their stance functions. They argued that hedges in Pakistani history research articles are used to make more careful claims in their research articles (Abbas & Zahra, 2019).

The obligation and directive stance markers are the least used markers and have zero occurrences in the introduction and results sections. However, the conclusion section includes the instances of the obligation and direction. These stance markers perform the discourse function of showing the attitude and desire of the authors for some action in response to the results of their study. These stance markers usually convey commands, recommendations, and instructions. The Pakistani authors also show their authority over their claims based on their findings. Overall, the hedges form the second-highest number of stance markers in research articles published by Pakistani authors. The highest number of the stance markers in Figure 2 are expressions of ability and possibility in research articles of the Pakistani authors. These expressions of ability and possibility show the likelihood of an action or event. These stance markers reflect the speaker's belief in any specific situation's possibility and expected potential occurrences. Figure 2 shows the equal use of the expression of ability

and possibility in the results and discussion sections of the research articles. These stance markers perform the discourse function of showing the possibility, feasibility and capacity of the research outcomes. The three stance markers also indicate the authors' evaluation and assessment of their research findings. These markers also shape the authors' perspective and show their stance on their study.

The highest number of epistemic stances was found in the results sections followed by the discussion sections of the research articles. These epistemic markers indicate the degree of certainty and uncertainty towards a specific claim in the research articles. The epistemic stance markers found in this study suggest that the Pakistani authors/ researcher authors express their doubt, reflecting the uncertainty of their claim in the section. The expression, "I do not know whether", illustrates the authors' epistemic stance as lacking knowledge or uncertainty of the findings. This expression also suggests that the authors cannot provide a precise answer to a specific result of the question. The results of this study are in contrast to the results of the study on lexical bundles in standalone literature (Write, 2019). The epistemic markers found in his study show the certainty of the claim. The difference between the uncertainty of the claim of Pakistani authors and Write's claim suggests that Pakistani authors are more careful and less confident in forming a strong claim in the research.

There were zero intention, volition and prediction expressions in Pakistani authors' research articles. These markers convey the authors' attitude toward events in the future. The intension marker, e.g., I am going to indicates the authors' intention to finish or perform some action in the future. The volition markers illustrate the desire of the authors to perform some action for the sake of better results in the future. The prediction markers predict the future course of action for the research. One of the possible reasons for the lack of these stance markers is that the corpus developed for this study was mainly based on quantitative research articles. Suppose we had included the qualitative research articles. In that case, the chances are that we might have found the instances of intention, volition, and prediction stance markers in the corpus of the Pakistani authors' research articles. The second possibility for the lack of occurrence of the stance markers is that Pakistani authors assume that they have avoided using first-person pronouns in their academic or research writing. This strong adherence to the concept might have led the Pakistani researchers to avoid using subjective language to maintain the objectivity of their claims.

Concerning question number three, we saw the concordance plot through the AntConc 3.5.7 corpus tool of every lexical bundle that we extracted from the corpus. It was found that most of the referential expressions were part of the results, discussion, and conclusion, except for very few referential expressions that were part of the introduction section. This distribution suggests that Pakistani authors are more concerned with their positionality by elaboration or comparison to the ideas, concepts, and objects in the discussion and conclusion sections. The stance markers were also used in the corpus's results, discussion, and conclusion and recommendation sections. Using these markers suggests that Pakistani authors are cautious while making claims. The expressions of ability and possibility were mainly present in the recommendation part of the papers. The instructors may use these bundles for pedagogical

purposes and teach the students to use a proper proportion of such expressions in the recommendation section. A lexical bundle under the caption hedges was found in the literature review section of the research articles. This finding may suggest that Pakistani authors have cited a small number of studies, and they claim clearly that the studies reviewed were small in number rather than extensive studies. The lexical bundle under the subcategory obligation and directives was found in the introduction section. These findings suggest that the authors wanted to fulfil the ethical considerations. However, this was found only in text 4 of the corpus. The other two texts of the corpus had zero occurrences. The teachers and instructors may suggest that the students use such bundles to meet the required criteria for ethical considerations when conducting research. This can be one of the pedagogical hints for the teachers/instructors to focus on this critical aspect of the research, which is ethical considerations.

Conclusion

To provide insights into teaching lexical bundles, we have analysed high-frequency lexical bundles based on their high-frequency occurrence in academic writing. The results showed that expressions of ability and possibility were the bundles with the highest frequency that Pakistani authors used more frequently than the referential bundles in their research articles. In Figure 1, the highest number of occurrences were of the referential expressions, with 20 occurrences. In Figure 1, the second highest number of occurrences were of the deictics and locatives, with 13 occurrences. This finding suggests that Pakistani authors are more concerned with locating and referring to the concepts and ideas mentioned in their writing. This strategy helps the writers to show logical connection and coherence to their academic writing. In Figure 2, the most frequently occurring expressions are of ability and possibility, with 21 occurrences. These expressions even form the overall highest number of occurrences in this study. The third highest frequency bundles were hedges with 19 occurrences. These two findings illustrate that Pakistani authors are more concerned with expressions of ability and possibility in their academic writing. Secondly, they are also concerned about the soft claims and hedges in their writing. The Pakistani authors avoid providing strong claims about their research.

However, both categories of bundles were found in the research, which suggests that Pakistani authors are well aware of the use of these lexical bundles. However, the authors have mainly focused on using hedges in the results, discussion, conclusion, and recommendation sections. The literature section may use the hedge bundles to avoid direct responsibility for the claim.

Pedagogical Implications

To back up teachers' efforts to teach lexical bundles, we have generated a limited yet helpful list of the academic lexical bundles that teachers can use in the classroom to teach academic writing. Teaching students these lexical bundles can increase their understanding of the organisational patterns of academic texts. For example, the students may learn the

comparison and contrast pattern of the argument, spatial pattern, and so on. The teachers can also show students the lists generated by Hyland Simpson-Vlach and Ellis (2010) and the list generated in this study and use all three lists as comprehensive data to teach undergraduate students to use high-frequency lexical bundles in academic writing. For the functions of the lexical bundles, the teacher can also direct students towards a data-driven approach where they ask them to look for the functions of lexical bundles through the concordance line in the corpus.

Every research may have limitations, so this study also has limitations. This study included data only from a single discipline of English Linguistics. Therefore, the recommendations made in this study are for English Linguistics only. The data can be obtained from various disciplines to collect the lexical bundles and teach academic writing in all disciplines. The corpus was also very limited and was created from the twelve research articles by Pakistani authors. Our study also included only two categories from Simpson-Vlach and Ellis' list of lexical bundles. The third category was excluded from this study because of the limited period it took to conduct this study. The third category of discourse organisation patterns may be studied in another project to explore high-frequency lexical bundles and their functions in academic discourse.

References

Abbas, A., & Zahra, S. T. (2019). Exploring metadiscursive culture of writing in history research articles published in Pakistan. *Pakistan Journal of History and Culture, XL*(2).

Arnaudet, M. L., & Barrett, M. E. (1984). *Approaches to academic reading and writing*. Englewood Cliffs. Retrieved from https://www.worldcat.org

Bal-Gezegin, B. (2019). Lexical bundles in published research articles: A corpus-based study. *Journal of Language and Linguistic Studies, 15*(2), 520–534.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Biber, D., Conrad, S., & Cortes, V. (2004b). If you look at . . .: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics, 25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

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

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

Gray, B. (2016). Lexical bundles. In P. Baker, & J. Egbert (Eds.), *Triangulating methodological approaches in corpus-linguistic research.* Routledge.

Hyland, K. (2000). *Disciplinary discourses: Social interaction in academic writing*. Longman.

Hyland, K. (2002). Options of identity in academic writing. *ELT Journal, 4*, pp. 351–358. doi:10.1093/elt/56.4.351

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

Lester, J. D., & James D. Lester, J. (1993). *Writing research papers*. Pearson.

Liu, C. Y., & Chen, H. H. (2020). Analysing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes, 58*, 122–137. https://doi.org/10.1016/j.esp.2019.12.003

Neely, E., & Cortes, V. (2009). A little bit about: Analyzing and teaching lexical bundles in academic lectures. *Language Value, 1*(1), 17–38. Retrieved from http://www.e-revistes.uji.es/languagevalue

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics, 11*(3), 283–304. doi:https://doi.org/10.1075/ijcl.11.3.04nes

Nikolopoulou, K. (2023, June 22). *What is purposive sampling?* | *Definition & examples*. Retrieved from https://www.scribbr.com/methodology/purposive-sampling/

Pérez-Paredes, P. (2020). Corpus linguistics approaches to understanding language use. In *Routledge ebooks* (pp. 35–63). https://doi.org/10.4324/9780429243615-3

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Wray, A. (2003). Formulaic language and the lexicon. *Journal of Pragmatics, 35*(10–11). doi:http://dx.doi.org/10.1016/S0378-2166(03)00079-1

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes, 54,* 1–14. https://doi.org/10.1016/j.esp.2018.09.001