ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

LEXICAL BUNDLES IN PAKISTANI SOCIAL SCIENCES RESEARCH ARTICLES: A CORPUS-BASED ANALYSIS OF FREQUENCY, STRUCTURE, AND FUNCTION

Umair Ashraf
PhD Scholar, Center for Languages and Translation Studies, Allama Iqbal Open University, Islamabad, Pakistan (umairashraf30@gmail.com)
Muhammad Ali* (Corresponding Author)
Visiting Lecturer, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan (muhammadali.jav@gmail.com)
Irfan Rasool
BS (H) English Literature & Linguistics, GCUF
irfanrasool2811@gmail.com

Abstract
Lexical bundles are recurrent multiword sequences that contribute to the organization, fluency, and rhetorical patterning of academic discourse. This corpus-based study examines the frequency, structural forms, and discourse functions of three- and four-word lexical bundles in Pakistani research articles in the social sciences. The corpus comprised 500 research articles published in Higher Education Commission-recognized Pakistani journals between 2013 and 2017, representing linguistics, literature, political science, psychology, and sociology. The corpus contained 2,079,944 word tokens and 46,951 word types. Lexical bundles were extracted through AntConc 3.5.2 and then manually filtered to remove overlapping or incomplete bundle forms. Structural analysis was guided by bundle-pattern categories derived from Biber et al. (1999) and Biber, Conrad, and Cortes (2004), whereas functional analysis followed Hyland's (2008) research-oriented, text-oriented, and participant-oriented framework. The results show 1,931 lexical bundles across the corpus, with sociology producing the largest number and political science the smallest. Noun phrase-based bundles dominated the structural profile across disciplines, indicating the nominal and informational density of Pakistani academic prose. Functionally, research-oriented and text-oriented bundles were more frequent than participant-oriented bundles, suggesting that Pakistani social science writers prioritize description of research procedures, textual organization, and reporting of results over explicit writer-reader interaction. The findings have implications for English for Academic Purposes instruction and for corpus-informed academic writing pedagogy in Pakistan.

Keywords: *lexical bundles, academic writing, Pakistani English, corpus linguistics, research articles, social sciences*

Introduction

Academic writing relies not only on individual lexical choices but also on recurrent phraseological patterns that help writers structure arguments, report findings, signal relationships, and position themselves within disciplinary communities. In corpus linguistics, such recurrent strings are commonly discussed as lexical bundles: statistically frequent sequences of words that occur repeatedly within a register or corpus (Biber et al., 1999; Hyland, 2008a). Unlike idioms, lexical bundles are not necessarily semantically opaque or grammatically complete; rather, they function as conventionalized building blocks of discourse.

The importance of lexical bundles is especially visible in academic prose. Formulaic patterns such as on the other hand, in the present study, the results show that, and the purpose of this help writers establish coherence, create disciplinary identity, and perform recognizable

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

rhetorical moves. For second-language and outer-circle English users, the ability to use such patterns appropriately is linked to fluency, idiomaticity, and participation in academic discourse communities (Cortes, 2004; Hyland, 2008a).

Pakistani English has been studied from several linguistic perspectives, including lexico-grammatical and multidimensional analyses. However, comparatively less attention has been given to the phraseological features of Pakistani academic writing, particularly the structural and functional behavior of lexical bundles in social science research articles. This gap is pedagogically significant because Pakistani novice researchers often write in English in institutional and publication contexts where academic phraseology plays an important role in clarity and acceptability.

The present article, developed from a larger thesis, investigates lexical bundles in Pakistani research articles in five social science disciplines. It addresses four questions:

- What are the frequencies of lexical bundles in Pakistani academic writing?
- What types of lexical bundles are used by Pakistani academic writers?
- What structures do these lexical bundles exhibit?
- What functions do they perform in academic discourse?

Literature Review

Lexical Bundles and Formulaic Academic Discourse

The term lexical bundle was popularized by Biber et al. (1999) to describe recurrent word strings identified through frequency-based corpus analysis. Lexical bundles form part of the broader domain of formulaic language, which also includes collocations, idioms, fixed expressions, and recurrent phrase frames (Wray, 2002). Their identification usually depends on frequency, range, and length thresholds, making corpus tools central to their analysis.

Research has shown that lexical bundles vary by register, discipline, and writer expertise. In conversational discourse, bundles often contain verb phrase and clause fragments, whereas academic prose frequently relies on noun phrase and prepositional phrase fragments (Biber et al., 2004; Biber & Barbieri, 2007). This structural difference reflects the informational density and compressed style of written academic registers.

For academic writers, lexical bundles act as rhetorical resources. They help writers introduce topics, report procedures, frame claims, quantify phenomena, and guide readers through argumentation. Hyland (2008a) argues that academic bundles are not merely repeated strings but discipline-sensitive resources through which writers enact research practices and interact with readers. Therefore, the study of lexical bundles contributes both to discourse analysis and to English for Academic Purposes (EAP) pedagogy.

Structural and Functional Approaches

Structural analyses of lexical bundles often classify bundles according to their grammatical patterning. Biber et al. (1999, 2004) identified major structural groups such as noun phrase-based bundles, prepositional phrase-based bundles, verb phrase-based bundles, dependent clause fragments, and other expressions. In academic writing, noun phrase and prepositional phrase bundles frequently dominate because they condense information and support nominal style.

Functional analysis focuses on what bundles do in discourse. Biber and colleagues proposed categories such as stance expressions, discourse organizers, and referential expressions. Hyland (2008a), working specifically with research-focused academic genres, proposed a model consisting of research-oriented bundles, text-oriented bundles, and participant-oriented bundles.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Research-oriented bundles help writers describe research activities, procedures, locations, topics, and quantities. Text-oriented bundles organize argument flow through transition, resultative, structuring, and framing signals. Participant-oriented bundles express stance or engage the reader. Because the present study examines written research articles, Hyland's model is particularly suitable.

Methodology

Corpus

The study used a specialized corpus of Pakistani social science research articles. The corpus comprised 500 articles published in Pakistani Higher Education Commission-recognized journals between 2013 and 2017. Five disciplines were represented: linguistics, literature, political science, psychology, and sociology. Each discipline contributed 100 articles. The articles were further divided into five rhetorical sections: abstract, introduction, literature review, methodology, and results and discussion.

The full corpus consisted of 2,079,944 word tokens and 46,951 word types. Table 1 presents the corpus composition by discipline. The disciplinary subcorpora were broadly comparable in size, although sociology was smaller than the other subcorpora.

Table 1

Corpus Composition by Discipline

Discipline	Word types	Word tokens	Type-token ratio
Linguistics	19,490	473,219	0.041186
Literature	22,871	449,078	0.050920
Political science	18,685	411,912	0.045360
Psychology	17,978	446,105	0.040290
Sociology	13,625	299,630	0.045470
Total corpus	46,951	2,079,944	0.022573

Extraction and Filtering of Lexical Bundles

Lexical bundles were extracted with AntConc 3.5.2 using the Clusters/N-Grams function. The analysis focused on three- and four-word bundles. Because the corpus was divided into discipline, year, and section files, extracted bundles were sorted by range as well as frequency to ensure that selected bundles were not restricted to isolated texts.

The extraction was followed by manual filtering. Overlapping sequences were examined so that a shorter bundle embedded in a longer, more meaningful bundle was not counted redundantly. For example, when a three-word sequence occurred as part of a more complete four-word expression, the more complete bundle was retained where appropriate. Conversely, when a four-word sequence contained an unnecessary additional word and a three-word expression represented a complete discourse unit, the shorter form was retained. This filtering step was necessary to improve functional interpretability.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Analytical Framework

The analysis proceeded in three stages. First, a frequency analysis counted the number of lexical bundles in each discipline, year, and article section. Second, a structural analysis classified bundles into noun phrase-based, prepositional phrase-based, verb phrase-based, clause-based, and other categories. Third, a functional analysis classified bundles using Hyland's (2008a) framework: research-oriented, text-oriented, and participant-oriented bundles. The functional categories were further interpreted through subcategories such as procedure, quantification, transition signal, resultative signal, structuring signal, framing signal, stance, and engagement.

Results

Frequency of Lexical Bundles Across Disciplines

The corpus yielded 1,931 lexical bundles. The distribution differed across disciplines. Sociology contained the highest number of bundles (n = 544), followed by linguistics (n = 409), psychology (n = 402), literature (n = 293), and political science (n = 283). Table 2 summarizes the distribution by discipline and identifies the article section(s) in which each disciplinary subcorpus showed the highest bundle concentration.

Table 2

Summary Distribution of Lexical Bundles by Discipline

Discipline	Total bundles	Corpus share	Section(s) with highest bundle concentration
Linguistics	409	21.2%	Literature review (115)
Literature	293	15.2%	Literature review (118)
Political science	283	14.7%	Literature review (94)
Psychology	402	20.8%	Literature review, methodology, and results/discussion (100 each)
Sociology	544	28.2%	Methodology and results/discussion (120 each)

Note. Percentages are calculated from the total number of bundles in the corpus (N = 1,931).

The pattern indicates that lexical bundles are not evenly distributed across disciplinary writing. Sociology's high bundle count suggests a strong reliance on recurrent phraseological patterns, particularly in sections that report procedures and interpret findings. Political science, in contrast, produced the lowest number of bundles, indicating either greater lexical variation or a smaller set of recurrent expressions within the extracted thresholds.

Structural Patterns

Structural analysis showed that noun phrase-based bundles were the dominant type across all five disciplines. In the abstract sections, noun phrase-based bundles accounted for 54.28% of linguistics bundles, 63.10% of literature bundles, 50.00% of political science bundles, 52.27% of psychology bundles, and 43.75% of sociology bundles. Examples include the role of, the importance of, the aim of this, this study attempts to, and the present study was. Prepositional phrase-based bundles were also common, especially expressions such as on the basis of, in the field of, in the context of, and in the light of.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table 3

Dominant Structural Profiles in Abstract Sections

Discipline	Most frequent structural type	Percentage	Examples
Linguistics	Noun phrase-based bundles	54.28%	the use of; the importance of
Literature	Noun phrase-based bundles	63.10%	the aim of this; a vast set of
Political science	Noun phrase-based bundles	50.00%	this study attempts to; the aim of this paper
Psychology	Noun phrase-based bundles	52.27%	the role of; the importance of
Sociology	Noun phrase-based bundles	43.75%	the present study was; sampling technique was used

Verb phrase-based, clause-based, and other bundles occurred less frequently. However, psychology showed a comparatively stronger presence of clause-based bundles than several other disciplines. Overall, the structural profile supports earlier findings that written academic prose relies heavily on nominal and prepositional packaging of information.

Functional Patterns

Functional analysis revealed a disciplinary contrast between research-oriented and text-oriented bundles. In linguistics abstracts, text-oriented bundles were the most frequent category (59.26%), with resultative signals such as the study reveals a and the findings of the. Research-oriented bundles accounted for 33.33% in linguistics. Literature also showed a strong text-oriented tendency (52.63%), while participant-oriented bundles were absent in the literature abstract corpus.

In political science, psychology, and sociology, research-oriented bundles were more prominent. Political science and psychology each showed research-oriented bundles at 65.85%, whereas sociology showed the strongest research-oriented pattern at 73.07%. These bundles often performed procedural and topic-related functions, as in the aim of this, the purpose of this, the role of, to investigate the, and for the purpose of. Participant-oriented bundles were generally rare; for example, political science showed only 7.69% participant-oriented bundles, and several disciplines showed none in the analyzed abstract sections.

Table 4

Functional Distribution in Abstract Sections

Discipline	Dominant functional category	Percentage	Representative bundles
Linguistics	Text-oriented bundles	59.26%	the study reveals a; the findings of the
Literature	Text-oriented bundles	52.63%	the aim of this; is an attempt to
Political science	Research-oriented bundles	65.85%	the aim of this; the purpose of this
Psychology	Research-oriented bundles	65.85%	the role of; to investigate the

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Sociology	Research-oriented bundles	73.07%	for the purpose of; present study aims at

Discussion

The findings demonstrate that lexical bundles are a salient feature of Pakistani social science research writing. The high overall number of bundles suggests that Pakistani academic writers rely substantially on formulaic phraseology to construct research discourse. This reliance is not inherently negative. In academic writing, recurrent bundles provide recognizable rhetorical scaffolding and help writers perform conventional moves such as identifying aims, describing procedures, reporting findings, and connecting arguments.

The dominance of noun phrase-based bundles confirms that Pakistani research articles share an important structural feature with academic prose more generally: dense nominal expression. Bundles such as the role of, the importance of, and the aim of this condense abstract meanings and help writers organize information economically. This pattern aligns with Biber et al. (2004) observation that noun phrase and prepositional phrase fragments are central to written academic registers.

The functional results are particularly important for EAP pedagogy. Research-oriented bundles were highly prominent in political science, psychology, and sociology, indicating that writers frequently use bundles to describe research aims, procedures, quantities, and topics. Text-oriented bundles were especially visible in linguistics and literature, where writers used bundles to manage transitions and report outcomes. However, participant-oriented bundles were rare. This suggests that Pakistani social science writers may be less inclined to use explicit stance and engagement markers in abstracts. Such avoidance may reflect disciplinary convention, writer caution, or limited confidence in managing evaluative and interactive academic stance.

A pedagogical implication is that lexical bundles should be taught not as isolated memorized phrases but as discourse-functional resources. Novice researchers need to know when a bundle reports a procedure, when it frames a claim, when it signals a result, and when it engages the reader. Corpus-informed teaching materials can therefore help students notice discipline-specific patterns and avoid overusing vague or mechanically repeated expressions. At the same time, teachers should emphasize variation, appropriacy, and rhetorical purpose so that bundle use contributes to effective academic style rather than formulaic redundancy.

The study also contributes to the description of Pakistani English academic discourse. By documenting how Pakistani social science writers use recurrent phraseological units, it provides a basis for further comparative work with other varieties of English, other disciplines, and different levels of writer expertise.

Conclusion

This corpus-based study examined three- and four-word lexical bundles in Pakistani social science research articles. The corpus of 500 articles produced 1,931 lexical bundles across five disciplines. Sociology showed the highest frequency of bundles, whereas political science showed the lowest. Structurally, noun phrase-based bundles dominated across disciplines, reflecting the nominal and information-dense character of academic prose. Functionally, research-oriented and text-oriented bundles were more frequent than participant-oriented bundles, showing that Pakistani writers mainly use bundles to describe research activities, organize texts, and report results rather than to explicitly engage readers or express stance.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

The findings support the pedagogical value of lexical bundles in academic writing instruction. For Pakistani researchers and postgraduate students, explicit awareness of bundle structure and function can improve rhetorical control, fluency, and discipline-sensitive expression. Future research should compare Pakistani social science writing with international published research, examine full article sections in greater statistical detail, include other disciplines such as natural sciences and engineering, and investigate whether novice and expert Pakistani writers differ in their use of stance and engagement bundles.

References

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. https://doi.org/10.1016/j.esp.2011.08.004

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

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer software]. Waseda University.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. International Journal of Corpus Linguistics, 14(3), 275–311.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. English for Specific Purposes, 26(3), 263–286.

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

Biber, 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. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), Out of corpora: Studies in honour of Stig Johansson (pp. 181–190). Rodopi.

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

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, 5, 31–64.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. Language Learning & 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, 23(4), 397–423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing-intensive history class. Linguistics and Education, 17(4), 391–406.

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

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

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. International Journal of Applied Linguistics, 18(1), 41–62.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Nattinger, J. R., & DeCarrico, J. S. (1992). Lexical phrases and language teaching. Oxford University Press.

Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), Language and communication (pp. 191–226). Longman.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. System, 42, 220–231. https://doi.org/10.1016/j.system.2013.12.003

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press.

Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press.