Linguistic Forum

Published by
MARS Publishers

ISSN (Online) 2707-5273
Volume 5, Issue 4, 2023
http://doi.org/10.53057/linfo/2023.5.4.3
Pages 57-73

OPEN ACCESS

Effect of Research Design on Distribution of Prototypical Participant-Oriented Lexical Bundles: A Corpus-Based Study of Quantitative Research Articles

Research Article

Correspondence:

Areej Alamghir
<areejalamghir@gmail.com>
M.Phil. Scholar, Department of Applied Linguistics, Government College University, Faisalabad, Punjab, Pakistan.

Dr. Aleem Shakir
<almsha@yahoo.com>
Assistant Professor, Department of Applied Linguistics, Government College University, Faisalabad, Punjab, Pakistan.

Publication Details

Received: November 22, 2023 **Accepted:** December 25, 2023 **Published:** December 30, 2023

Abstract

This study aims to evaluate Prototypical Participant-Oriented Lexical Bundles (PPOLBs) in quantitative research articles to discern potential variations between correlational and experimental research designs. The selection of PPOLBs for this investigation aligns with the taxonomy of functional categories developed by Hyland (2008a) and later refined by Salazar (2014) in terms of prototypicality. The study encompasses five major disciplines, 11 subjects, and a total of 2576 research articles. Data distribution is stratified based on research design, journal impact factors, and article sections. The AntConc program was employed to retrieve the frequency of PPOLBs. Initially, a 3-factor MANOVA was used to explore the effects of research design on the distribution of PPOLBs; however, due to the absence of normal distribution, the analysis transitioned to the Kruskal-Wallis H Test. Among the three categories, only STNC_LBs exhibited a statistically significant difference ($2(1) = 11.489$, $p = 0.01$), while ACK_LBs and the other subcategory displayed no notable variations. The findings of this study aim to equip second language students with insights into the effective use of linguistic elements based on registers and proficient writing.

Keywords: Quantitative research articles, prototypical participant-Oriented Lexical Bundle, Kruskal-Wallis H tests, PPOLBs

1. Introduction

1.1 Scope of Corpus Linguistics in ELT

Corpus linguistics, as a research methodology, employs an empirical approach to study linguistic phenomena by analyzing vast collections of real texts, known as corpora. Examining these corpora involves recognizing patterns, frequency distributions, and linguistic traits to gain insights into language use, diversity, and communication-related issues. Corpus-based research has gained popularity in examining linguistic and stylistic characteristics, offering a more detailed understanding of the real-world application of language. The analysis of massive textual data using corpus linguistics provides verifiable facts and proof, contributing significantly to academic research and language learning.

Students transitioning from classroom language to real-world usage benefit greatly from corpuses, making educational tools more believable and interesting. Major publishing firms use multimillion-word corpora to develop various English language teaching resources, enhancing grammar books, course materials, vocabulary resources, test prep materials, and teacher aids.

Guided exercises and computer-assisted language learning engage students with corpora, enhancing language learning through data-driven learning. Learner corpora, containing authentic writings by English language learners, have also improved English language teaching, providing valuable resources for educators both inside and outside the classroom.

In English for Academic Purposes (EAP), corpus linguistics proves helpful by revealing common language patterns across genres. Corpus analysis has uncovered distinct features of academic discourse, aiding in understanding language nuances specific to academic writing.

1.2 Research Problem

While prior studies gathered quantitative research article (RA) data, they often overlooked their multidisciplinary or interdisciplinary nature. The correlational data with experimental data was rarely explicitly acknowledged, and there was a lack of focus on balanced data statistics. The current study addresses this gap by exploring potential linguistic differences between correlational and experimental investigations.

The study aims to identify any appreciable differences in the frequency of PPOLBs in correlational and experimental research designs. As L2 authors often find it challenging to navigate domain-specific elements while writing, this research seeks to contribute to a better understanding of writing norms in different research paradigms.

While past studies have explored variations in linguistic features related to different research paradigms, there is a dearth of research concentrating on a single paradigm (quantitative) to identify linguistic variations in different research designs, such as correlational vs. experimental.

This research serves the ESP domain by enhancing comprehension of specific topics, aiding in the production of learning materials, and contributing to student and teacher awareness, particularly for those whose L2 is English. The analysis of research articles using different quantitative research designs is still in its early stages, offering insights for researchers to choose the best features for specific article types.

1.3 Hypotheses

H01: The category of research paradigm (experimental and correlational) does not affect the frequency of PPOLBs.

1.4 Research Question

Q 1. Which research paradigm (experimental or correlational) has more prototypical participant-oriented lexical bundles than the other? If so, what are the frequency and statistics of the difference?

1.5 Research Objectives and Significance

This study aims to distinguish itself from past research by presenting different objectives. The first goal is to identify PPOLBs in academic research publications in the humanities and social sciences. The second objective is to evaluate the frequency of these bundles concerning correlational and experimental research designs in quantitative studies.

The comparison will aid researchers in discovering fixed phrases, multiword expressions, or formulaic sentences specific to experimental and correlational designs. The transdisciplinary study also supports EAP lessons, helping researchers and instructors produce materials suited for different disciplines and genres. This research's explanation of how various fields and genres affect multiword expressions benefits teachers and L2 students, contributing to improved writing skills and comprehension of different genres. The corpus-driven analysis of linguistic variation provides opportunities for English research pursuits and educational improvement.

2. Literature Review

2.1 Phraseology and its Scope

In recent years, there has been a growing interest in the study of phraseology; however, its advancement is hindered by a lack of consensus on nomenclature, descriptive techniques, and analytical processes (Granger & Paquot, 2008; Howarth, 1996). Howarth (1996a) attributes this lack of uniformity to academics focusing on specific facets of phraseology, such as idioms, collocations, or speech formulae. The field's almost autonomous evolution across disciplines, including discourse analysis, lexicography, linguistics, second language acquisition, language processing, and artificial intelligence, adds to the complexity (Howarth, 1996).

Granger and Paquot (2008) link semantics, morphology, syntax, and discourse to phraseology, emphasizing its flexible scope with ambiguous boundaries.

2.2 Lexical Bundles and Prototypicality

Disciplinary or academic writing often utilizes linguistically understandable and cohesive multiword units with discernible discourse functions, patterns, and structures (Adel & Erman, 2012; Biber et al., 2004; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a, 2008b; Pan et al., 2016). Cortes (2015) focused on lexical bundles, concurrent groups of three or more words common in a language register, identified through empirical examination of a linguistic corpus. The study revealed that the most common sentences were lexical bundles.

Lexical bundles (LBs) are crucial in formulaic language and have been extensively studied in EAP contexts (Biber et al., 2004; Liu & Chen, 2020), across academic disciplines (Cortes, 2004; Hyland, 2008a, 2008b), in L1 and L2 English (Chen & Baker, 2010; Pan et al., 2016), and other contexts. Salazar's (2011) approach identified PPOLBs, forming the basis for the present study's focus.

2.3 Lexical Bundles in Academic Writing and EAP Pedagogy

Hyland (2008b) argued that readers can recognize the authenticity of expressions in specific situations, illustrated by examples like "as can be seen" in academic writing. Studies examining the use of lexical bundles in various research domains indicate the prioritization of certain bundles by academic fields.

Research has investigated lexical bundles in diverse contexts, such as history courses (Cortes, 2006), European Union registrations in Hungary (Jablonkai, 2010), and studies by Iranian academics (Nasrabady et al., 2020) and Korean students (Lee et al., 2020). These studies contribute to a comprehensive understanding of lexical bundle usage in different academic contexts.

Recent decades have seen extensive research on formulaic lexical bundles, comparing their prevalence in written compositions by native and advanced English speakers. These studies highlight variations in verb-based and non-native speaker preferences, emphasizing functional classifications in scholarly writing.

The sole investigation on LBs in the Algerian corpus (Rezoug & Vincent, 2018) focused on electrical engineering, revealing similarities to previous research on lexical clusters in academic writing.

2.4 Structural and Functional Categories of Lexical Bundles

Structural analysis (Biber et al., 2004; Lu & Deng, 2019) helps determine the organizational properties of lexical bundles. Salazar (2011) observed that most lexical bundles are within phrases or sentences, with the final word frequently initiating the ensuing structure.

Biber et al. (1999) proposed a taxonomy categorizing sentence forms, while Biber et al. (2004) classified patterns into three main structural types. Function-based studies, influenced by Hyland's (2008b) classification, focus on referential, discourse organizer, and attitude bundles.

The present study adopted the participant-oriented functional taxonomy developed by Hyland (2008a), emphasizing three subcategories: Stance, Engagement, and Acknowledgement.

2.5 Relationship between Past and Present Research

The present study builds on Salazar's (2011; 2013) work, creating a unique contribution in corpus linguistics and English for specific purposes (EAP). It focuses on PPOLBs identified in Salazar's (2011) study of LBs in native and non-native scientific writing.

In contrast to previous studies that considered different independent variables, the present research aims to compare PPOLBs across correlational and experimental research articles. This study addresses gaps in existing literature, emphasizing the research design's impact on lexical bundles.

2.6 LBs within Quantitative Research Paradigm

Lu and Deng's (2019) approach to downloading research articles is not fully detailed, but they conducted a comparative analysis of native and non-native academic authors' writing styles. Salazar (2011) investigated lexical bundles in health sciences, emphasizing scientific discourse, and potential pedagogical uses.

2.7 LBs across Quantitative Research Design

Past studies have compared LBs between quantitative and qualitative investigations, but few have explored this within a single paradigm. Cao (2021) examined lexical bundle utilization in developmental and educational psychology research publications, finding notable differences between study paradigms. Varghaei and Khodadadi (2022) focused on abstracts from medical research journals but did not specify the research design.

3. Methodology

3.1 Research Design

This study employed a quantitative research approach and corpus-based analysis to compare and examine the distribution patterns of three subcategories of a dependent variable: PPOLBs. The analysis focused on comparing independent variables between experimental and correlational quantitative research papers.

3.2 Data Collection

The study's corpus included eleven subjects representing five distinct disciplines: social sciences, physical science, biological science, medicine/medical science, and business education. Two subjects were selected from each discipline, resulting in a total of sixty publications per subject. The data was distributed based on distinctive quantitative research sub-design (correlational or experimental), three journal criteria using the HEC Journal Recognition System (HJRS), and the IMRD structure. Thirty articles belonged to correlational design, and thirty articles belonged to experimental design, with an additional categorization under three distinct journal categories (X, Y, and W), ensuring a balanced selection. All research articles were selected from the period between 2019 and 2023.

3.3 Corpus Editing and Preparation

Research articles in PDF format were converted to MS Word format using Acrobat XI Pro. The editing process involved standardizing the text by condensing citations, numerical citations, and parenthetical citations into a singular term, "Ref." Formulas and analytical models were removed to prevent false inflation of noun frequency, ensuring the original order of pointers was maintained. After converting all sections of the Research Article (RA) into Microsoft Word files, these files were gathered and merged into a centralized directory labeled "All Corpus." The corpus was then prepared for AntCon (version 3.5.9) compatibility, converting all Word files into .txt format using AntFile Converter version 2.0.2. A subdirectory named "Txt" within the parent directory "All Corpus" housed the text files.

3.4 Instrument

This study analyzed the occurrence of PPOLBs in research articles (RAs) within two quantitative research sub-designs (experimental and correlational). The primary tool for ascertaining the occurrence rate of PPOLBs was a manual compilation based on prototypical participant-oriented lexical bundles identified by Salazar (2011). AntConc (version 3.5.9) was used for concordance and frequency calculations, AntFile Converter for file conversion, and Adobe Acrobat Pro for PDF to Word format conversion. The HEC Journal Recognition System (HJRS) website (https://hjrs.hec.gov.pk/) assisted in journal selection within the X, Y, and W categories.

3.5 Validity and Reliability

The study ensured reliability and validity through multiple protocols. An exemplary prototypical participant-oriented lexical bundle was double-checked by peers and superiors. Data collection involved determining the CORR or EXP types of research, checking journal categories, and adhering to the IMRD model. If some articles shared results and discussion sections, they were considered a single result section. Data analysis with AntConc underwent a round of peer review, and statistical analysis benefited from various specialists participating in the evaluation process.

3.6 Data Analysis

During the initial stage of data analysis, dependent variables underwent normalization to a standardized scale of 1000 words to address potential influences from disparate file sizes and text lengths. While initially considering a three-factor multivariate analysis of variance (MANOVA), the study reevaluated the appropriateness of MANOVA due to numerous outliers in the dependent variables.

Table 3.1 Outliers in three Subcategories of PPOLBs

STNC_LBs_Normed1000	Highest	1
		2
		3
		4
		5
	Lowest	1
		2
		3

		4
		5
ENG_LBs_Normed1000	Highest	1
		2
		3
		4
		5
	Lowest	1
		2
		3
		4
		5
ACK_LBs_Normed1000	Highest	1
		2
		3
		4
		5
	Lowest	1
		2
		3
		4
		5

a. Only a partial list of cases with the value .00 is shown in the table of lower extremes.

Many examples were assigned a score of 0 for each variable, leading to the classification of these cases as outliers. Furthermore, as shown in Table 2, normality tests for all variables indicated that none of the three dependent variables exhibited a normal distribution ($p > .05$).

Table 3.2 Tests of Normality for three Subcategories of PPOLBs

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
STNC_LBs_Normed1000	.299	2576	.000	.544	2576	.000
ENG_LBs_Normed1000	.440	2576	.000	.261	2576	.000
ACK_LBs_Normed1000	.522	2576	.000	.042	2576	.000

a. Lilliefors Significance Correction

Due to a lack of normal distribution for all variables ($p < .05$), MANOVA was ruled out as a viable option for data analysis.

3.7 Assumptions Checking for 3-Factor ANOVA

Given that each of the three variables represented a subcategory of the single variable (PPOLBs), a composite variable named "TTL_Normed1000" was created by summing the normed scores of the three variables to explore the potential use of 3-factor ANOVA on the data.

Table 3.3 below illustrates that a significant number of cases were identified as outliers through outlier analysis of the dependent variable.

Table 3.3 Outliers in Composite Variable Obtained by Totaling Normed to 1000 Words Scores of three Subcategories of PPOLBs (TTL_Normed1000)

			Case Number	Value
TTL_Normed1000_LBs	Highest	1	1550	57.69
		2	1554	35.02
		3	2509	32.85
		4	1823	25.00
		5	1503	23.53
	Lowest	1	2575	.00
		2	2574	.00

	3	2573	.00
	4	2572	.00
	5	2567	.00[a]

a. Only a partial list of cases with the value .00 is shown in the table of lower extremes.

Upon examination of variables with zero values, it was found that 150 cases had a value of zero on the lower end. These cases were subsequently eliminated. On the higher end, five cases exhibited extreme values, which were also excluded. Therefore, a total of 159 cases were removed.

Following the removal of these values, normality tests were conducted, and the results are presented in Table 3.4 below.

Table 3.4 Test of Normality for Composite Variable Based on three Subcategories of PPOLBs (TTL_Normed1000)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
TTL_Normed1000_LBs	.304	2576	.000	.477	2576	.000

a. Lilliefors Significance Correction

The results of the normality tests indicated that the data did not follow a normal distribution ($p = .000$).

3.8 Selection of Kruskal-Wallis H Test

The Kruskal-Wallis H Test serves as a nonparametric alternative to ANOVA and MANOVA, demonstrating robustness in the presence of outliers and offering flexibility in handling multiple groups and subcategories within a single independent variable. Consequently, the Kruskal-Wallis H Test was chosen because the dependent variable did not adhere to the assumption of normality. The data exhibited a departure from a normal distribution, as indicated by the results of the normality tests ($p < .000$). The study aimed to investigate the potential effects of research design, featuring two subcategories (EXP and CORR), on three subcategories of PPOLBs. To analyze the data, three Kruskal-Wallis H tests were conducted due to the data not meeting the assumptions of MANOVA or ANOVA.

4. Results and Discussion

This chapter is structured into two main sections. The initial part (4.1) presents descriptive statistics results for independent variables (research design) and three dependent variables (PPOLBs). The subsequent part (4.2) details the study's hypotheses' results and provides a corresponding discussion.

4.1 Descriptive Statistics

The descriptive statistics section provides frequencies for both independent and three dependent variables.

4.1.1 Frequency Distribution of Independent Variables: Research Design

This section presents the frequencies of independent variables, where the research design acts as the independent variable. As illustrated in Table 4.1, this variable comprises two levels: correlational and experimental.

Table 4.1 Frequency Distribution of Research Articles by Research Design

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Correlational	1301	50.5	50.5	50.5
	Experimental	1275	49.5	49.5	100.0
	Total	2576	100.0	100.0	

The number of texts in the former is 1301, while it is 1275 in the latter. Among the total number of texts, correlational texts constitute 50.5%, whereas experimental texts constitute 49.5%. The data is well-balanced in terms of the number of texts.

4.2 Frequency Distribution of Dependent Variables: PPOLBs

This section provides descriptive results of the three dependent variables, which are subcategories of PPOLBs. It includes both raw and normed frequencies of different subcategories of lexical bundles for comparison with past studies.

Table 4.2 Raw and Normed Frequency Distribution of Subtypes of PP0LBs in Corpus

Subcategories of POLBs	**Raw (out of 2967056 words)**	**Normed (million words)**	**Normed (thousand words)**
STNC_LBs	2622	883.7	0.88
ENG_LBs	639	215.3	0.21
ACK_LBs	13	4.38	0.00
	Total freq. of POLBs per million words	**1103.38**	

The corpus consists of 2,967,056 tokens, as indicated in Table 12. Normed frequencies reveal that prototypical participant-oriented procedural lexical bundles appear with the highest frequency. In a study conducted by Zare and Valipouri (2022), data on 3,751,006 tokens was collected, but they did not produce normed values for the POLBs. Yang (2022) attempted the same with data on 1,122,690 tokens, but they also failed to do so. However, Xia (2022) collected data from 127,875 tokens,

forming values per million words. The following table presents descriptive results of the three dependent variables.

Table 4.3 Descriptive Statistics for three Subcategories of PPOLBs

		STNC_LBs_Normed1000	ENG_LBs_Normed1000	ACK_LBs_Normed1000
N	Valid	2576	2576	2576
	Missing	0	0	0
Mean		.9328	.2376	.0040
Std. Error of Mean		.03488	.01801	.00113
Median		.0000	.0000	.0000
Mode		.00	.00	.00
Std. Deviation		1.77018	.91387	.05735
Skewness		5.994	13.050	15.433
Std. Error of Skewness		.048	.048	.048
Kurtosis		71.425	287.984	252.163
Std. Error of Kurtosis		.096	.096	.096
Range		32.05	25.64	1.27
Minimum		.00	.00	.00
Maximum		32.05	25.64	1.27
Sum		2402.84	612.06	10.18

The lack of alignment between mean, mode, and median on one hand, and high kurtosis and skewness values on the other hand, in all cases, suggests issues with normal distribution, a concern further supported by the normality tests reported in Chapter 3. Previous researchers, such as Ghorbani et al. (2022), Xia (2022), and Yang (2022), employed the chi-square test for statistical analysis. However, Xia (2022) also utilized the t-test for analysis. In some works, researchers like Akbulut (2020), Hyland & Jiang (2018), and Lu (2019) used Paul Rayson and log-likelihood to analyze participant-oriented lexical bundles.

4.2 Results of Inferential Statistics

This section presents the findings of hypotheses formulated for the study and provides a discussion of the outcomes.

4.2.1 Effect of Research Design on the Frequency of PPOLBs

The first null hypothesis tested the effect of research design on the frequency of three subcategories of PPOLBs (i.e., STNC, ENG, and ACK).

Null Hypothesis: The distribution of subcategories of PPOLBs is the same across research designs (correlational and experimental).

The results of the Kruskal-Walis H Test appear in Table 14 and Table 15 below, representing mean rank scores and test statistics, respectively. The ranks table reveals the rank achieved by correlational and experimental research on each subcategory of PPOLBs.

Table 4.4 Ranks Achieved by Correlational and Experimental Research Articles on Subcategories of PPOLBs

	Research Design	N	Mean Rank
STNC_LBs_Normed1000	Correlational	1301	1333.05
	Experimental	1275	1243.04
	Total	2576	
ENG_LBs_Normed1000	Correlational	1301	1283.44
	Experimental	1275	1293.66
	Total	2576	
ACK_LBs_Normed1000	Correlational	1301	1285.96
	Experimental	1275	1291.10
	Total	2576	

Experimental research articles attain a higher mean rank on ENG_LBs (1293.66 vs. 1283.44) and ACK_LBs (1291.10 vs. 1285.96), whereas correlational research articles achieve a higher mean rank score on ACK_LBs (1333.05 vs. 1243.04). Table 15 presents the results of statistically significant differences between correlational and experimental research in three subcategories of PPOLBs based on three Kruskal-Wallis Tests.

Table 4.5 Test Statistics[a,b]: Effect of Research Design on PPOLBs

	STNC_LBs_Normed1000	ENG_LBs_Normed1000	ACK_LBs_Normed1000
Chi-Square	11.489	.295	2.042
Df	1	1	1
Asymp. Sig.	.001	.587	.153

a. Kruskal Wallis Test
b. Grouping Variable: Research Design

As Table 15 indicates, there was a statistical difference between correlational and experimental research articles in one out of three subcategories of PPOLBs (STNC_LBs: $\chi2(1) = 11.489$, $p = 0.01$; ENG_LBs: $\chi2(1) = .295$, $p = 0.587$; ACK_LBs: $\chi2(1) = 2.042$, $p = .153$). Thus, the first null hypothesis has been generally rejected.

The study investigated three subcategories of common PPOLBs to identify differences between correlational and experimental research articles. Out of the three, only STNC_LBs (2(1) = 11.489, p = 0.01) showed a statistically significant difference. ACK_LBs and the other two subcategories did not exhibit any notable variations. Although the distinction was observed in prior studies, none was able to detect it using PPOLBs.

Hyland (2008a) chose the domains of applied sciences, pure sciences, and social sciences as focal areas for investigating a cross-disciplinary four-word phrase. However, the paper did not specify the exact experimental paradigm in which the RAs were downloaded. It has also been noted that the research design showed differences in linguistic features (Candarli & Jones, 2019; Gray, 2015; Hu & Cao, 2015; Ren, 2021). However, most researchers used quantitative data from multiple disciplines, but none of them mentioned subtypes of quantitative research design (Bao & Liu, 2022; Budiwiyanto & Suhardijanto, 2020; Cortes, 2004; Cortes, 2013; Ren, 2021; Pan & Liu, 2019; Varghaei & Khodadadi, 2022; Wright, 2019). The results of PPOLBs in research designs in this study, however, varied. In the preceding research mentioned in the previous section, which was experimental and correlational, no PPOLBs analysis based on the study paradigm was carried out.

5. Conclusion

5.1 Summary

This study scrutinizes the utilization of PPOLBs in quantitative research papers to discern differences in their usage between correlational and experimental research methods. PPOLBs, specific language constructions selected for examination due to their prototypical nature, fall under the functional categories taxonomy initially defined by Hyland (2008a) and later amended by Salazar (2014). Encompassing a dataset of 2576 research publications across 11 specific topics and spanning five major academic disciplines, the data distribution depends on the study methodology and the journals' impact factors. The frequency of PPOLBs is examined using the AntConc application.

Initially intending to investigate the impact of research design, journal category, and article section on PPOLB distribution through a 3-factor MANOVA, the study pivots to use the Kruskal-Wallis H Test. The findings reveal significant variations in PPOLB usage across diverse research methodologies, particularly in the STNC_LBs subcategory. These results offer practical guidance on effective language use based on registers in writing, holding implications for second language (SL) learners. The study contributes to understanding PPOLBs in academic writing, showcasing variances in different research contexts and offering advice for SL students seeking to enhance their writing abilities.

5.2 Implications

The implications of this study have profound potential to reshape various facets of the academic and scientific publication landscape. Firstly, the emphasis on enhanced research communication underscores the importance of accessibility and clarity in academic discourse, fostering increased understanding and cross-disciplinary collaboration. In an increasingly complex world where diverse solutions are vital for addressing serious issues, the need for multidisciplinary cooperation is highlighted. Improved editorial procedures are essential to ensure published research meets the highest standards of validity and reliability.

The study's findings can inform more effective research writing instruction for educators, empowering the next generation of researchers to present compelling arguments. The connection to corpus linguistics opens avenues for advancements in data-driven research techniques. The prospect of advanced automation and text-mining techniques points to a future where research procedures are expedited, and insights are gained more efficiently. In conclusion, the impact of this study will

resonate across academia, fostering better cooperation, communication, and innovation in the years to come.

5.3 Suggestions for Future Research

Exploring further research avenues in this domain could focus on enhancing teaching and writing support. The creation of educational resources and writing assistance software may prove beneficial, particularly for novice researchers in their early careers, facilitating improved academic writing. Investigating potential variations in the utilization of participant-oriented lexical bundles across languages by broadening the study to include articles in multiple languages is essential, especially in the context of international research collaboration.

Delving into interdisciplinary comparisons, a study could analyze the distribution of lexical bundles in both quantitative and qualitative articles across diverse disciplines. This expanded scope might unveil disciplinary-specific linguistic tendencies. Examining the distribution of lexical bundles in academic and non-academic texts, such as blogs and news stories, could illuminate the unique qualities of academic language in contrast to other forms of discourse. Additionally, exploring how the use of participant-oriented lexical bundles influences reader comprehension in research publications could involve conducting reader surveys or eye-tracking research.

Authorship analysis represents another intriguing avenue. Investigating the consistency with which specific scholars or groups employ particular lexical bundles in their publications could provide insights into authorial preferences and styles. Furthermore, delving into citation research to examine whether the influence of citations from publications is related to the use of specific lexical bundles might facilitate a better understanding of how language choices impact the reception of research.

Funding: This study was not funded in any shape or form by any party.

Conflict of Interest: The author declares that he has no conflict of interest.

Bio-Note:

Areej Alamghir has completed her MPhil in Applied Linguistics in 2022. She is passionate about raising the quality of English language teaching. Her area of interest is ELT, Testing and Evaluation, and Corpus.

Dr. Aleem Shakir is an Assistant Professor in the Department of Applied Linguistics, Government College University, Faisalabad. His areas of interests are ELT, Phonetics and Phonology, ESP, Testing and Evaluation and Corpus Linguistics.

References

Anthony, L. (2014). *AntConc (Version 3.4. 3)[Computer Software]*. Tokyo, Japan: Waseda University.

Anthony, L. (2022). *AntFileConverter (Version 2.0.2) [Computer Software]*. Waseda University. Tokyo, Japan. [Software]. Available from https://www.laurenceanthony.net/software.

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

Biber, D. (2007). University language: A Corpus-based study of spoken and written registers. *TESL-EJ*, *11*(2).

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

Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora, 14*(2), 237e263.

Cao, F. & Hu, G. (2014). Interactive meta-discourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics,66,* 15–31.

Cao, F. (2021). A comparative study of lexical bundles across paradigms and disciplines. *Corpora*, *16*(1), 97-128.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing.

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

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21-42.

Gilquin, G., Granger, S., &Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 319-335.

Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, *37*(3), 538-546.

Granger, S., &Paquot, M. (2008). Disentangling the phraseological web. *S. Granger & Meunier (Eds.)*.

Gray, B. (2015). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial this/these in academic prose. *Journal of English for Academic Purposes*, *9*(3), 167-183

Howarth, P. (1996). *Phraseology in English Academic Writing.* Tübingen: Niemeyer.

Hyland, K. (2000). *Disciplinary discourses*. Harlow: Pearson Education. *International Journal of Lexicography*, *16*(2).

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

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

Johns, T. (2002). Data-driven learning: The perpetual challenge. In *Teaching and learning by doing corpus analysis* (pp. 105-117). Brill.

Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, *39*, 21-36.

Major, M. (2006). Longman exams dictionary.

McCarthy, M. (2008). Felicity O‟ Dell. *Academic Vocabulary in Use. CUP*.

McCarthy, M., McCarten, J., & Sandiford, H. (2005). *Touchstone level 1 workbook*. Cambridge: Cambridge University Press.

Mills, G. E., & Gay, L. R. (2019). *Educational research: Competencies for analysis and applications*. Pearson. One Lake Street, Upper Saddle River, New Jersey 07458.

Neely, E., & Cortes, V. (2011). A little bit about: Analyzing and teaching lexical bundles in academic lectures.

O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Panah, E., Yunus, M. M., &Embi, M. A. (2013). Google-informed patter-hunting and pattern-defining: Implication for language pedagogy. *Asian Social Science, 9*(3), 229.

Paquot, M. (2007). EAP vocabulary in EFL learner writing: from extraction to analysis: A phraseology-oriented approach. *Unpublished PhD thesis. Universitécatholique de Louvain, Centre for English Corpus Linguistics*.

Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes*, *50*, 100968.

Rezoug, F, & Vincent, B. (2018). Exploring Lexical Bundles in the Algerian Corpus of Engineering. *Arab Journal of Applied Linguistics,3*(1), 47-77.

Rundell, M. (2007). *Macmillan English dictionary for advanced learners* (2nd ed.). Macmillan Education.

Salazar, D., Verdaguer, I., Laso, N. J., Comelles, E., Castaño, E., &Hilferty, J. (2013). Formal and functional variation of lexical bundles in biomedical English. *Biomedical English. A Corpus-based Approach. Amsterdam: John Benjamins*, 39-53.

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.

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

Sinclair, J. (2005). *Corpus and Text-Basic Principles in Developing Linguistic Corpora*: A Guide to Good Practice, ed. M. Wynne.

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).

Verdaguer, I., Laso, N. J., Guzmán-González, T., Salazar, D., Comelles, E., Castaño, E., &Hilferty, J. (2013). SciE-Lex: a lexical database. *Biomedical English: A corpus-based approach*, 21-38.

Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, *42*(2), 155-166.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, *54*, 1-14.

Yang, M. (2022). Connecting the functions of lexical bundles and moves in published research articles: The case of developmental and educational psychology. *Nordic journal of English studies*.

Yin, X., & Li, S. (2021). Lexical bundles as an intradisciplinary and interdisciplinary mark: A corpus-based study of research articles from business, biology, and applied linguistics. *Applied Corpus Linguistics*, *1*(1), 100006.