Vol. 6 No.4 2023

ISSN Online: 2709-7625

ISSN Print: 2709-7615

# EFFECT OF JOURNAL CATEGORY ON DISTRIBUTION OF PROTOTYPICAL TEXT-ORIENTED LEXICAL BUNDLES: A CORPUS-BASED STUDY OF QUANTITATIVE RESEARCH ARTICLES

**Hurmet Fatima**
MPhil Scholar, Government College University, Faisalabad, Punjab, Pakistan. Email Id: hurmetfatima@gmail.com
**Dr. Aleem Shakir**
Assistant Professor, Department of Applied Linguistics, Government College University, Faisalabad, Punjab, Pakistan. Email Id: almsha@yahoo.com

## Abstract

*This study intends to assess the Prototypical Text-Oriented Lexical Bundles (henceforth, PTOLBs) in a quantitative research article to determine whether there are any changes in the journal categories research articles of multiple disciplines. The taxonomy of functional categories developed by Hyland (2008b) includes the categories of Text-Oriented Lexical Bundles (as known as, TOLBs) which were later modified by Salazar (2011) in terms of prototypicality. The study examined how these PTOLBs are distributed concerning journal ratings by HEC's HJRS research articles. Based on these goals, the study chose five major disciplines, 11 subjects, and 2576 Research Articles' sections. There are approximately 3 million words data is used in this corpus-based study. The frequency of PTOLBs was retrieved using the AntConc software. The distribution of PTOLBs in RAs was investigated initially using a 3-factor MANOVA but later on, due to the lack of normal distribution, shifted to the Kruskal-Walis H Test and post hoc test. This study showed that in journal categories, Y with W and X showed a significant difference on the base of the Objective Lexical Bundles' subcategory. The findings will provide SL students with an understanding of how to use linguistic elements based on registers and effective writing. The examination of the distribution of PTOLBs conducted in this research might aid students in recognizing patterns and formulae that are often used in scholarly papers.*

***Keywords:*** Prototypical Text-Oriented Lexical Bundles, Journal Category, Quantitative Research Articles, HJRS, Corpus-Based, IMRD, Phraseology, L2 Learners, Academic Writing.

# 1. Introduction

## 1.1 Background of the Study

In the last few years, multiple investigations have been dedicated to examining the characteristics and significance of formulaic language. Formulaic language is a broad term including combinations of multiple words that function as cohesive units, such as idiomatic terms, static statements, collocations, and irregular phrases, among other examples (Segalowitz, 2010; Titone & Connie, 1999)

Multi-word combinations, cannot be considered idiomatic, have excessively prominent meanings, and are fundamentally deficient fall under the category of formulaic language. We call these sets "lexical bundles" (LBs). According to Biber and Barbieri (2007), a large proportion of formulaic language patterns are LBs. LBs are not only linguistically explicit but structurally unfinished. In the realm of interdisciplinary or academic literature, it is common to employ a

Vol. 6 No.4 2023
ISSN Online: 2709-7625
ISSN Print: 2709-7615

variety of rhetorically accessible and highly coherent multiword units that include observable language processes, analytical sequences, and hierarchies (Adel & Erman, 2012; Biber et al., 2004; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a, 2008b; Pan et al., 2016).

For a more comprehensive look at language statistics, as well as quantitative analyses, Corpus Linguistics (henceforth, CL) is invaluable (Gries 2010). The analytical instruments from the field of CL comprise word databases and frequency counts, lexicon variety, coherence diagrams, collocations, and LBs (Biber et al. 1999). The use of corpus-based research has seen an increase in its prevalence for the examination of linguistic and stylistic attributes of a given text (Panah et al., 2013).

Quantitative counting of language features is merely a component of Corpus-based (further used as CB) research; qualitative assessments of statistical Information are equally essential. The CB investigation, as defined by Biber et al. (1998), aims to do more than just report quantifiable linguistics facts; rather, it seeks to draw attention to trends in language usage by analyzing that information. Because of this, the study analyses data based on journal quality by examining the use of PTOLBs (Salazar, 2014), which is divided into three categories: Y, X, and Y, making it even more distinctive and deserving in the field of linguistic analysis (HEC recognized criteria).

## 1.2 Problem Statement

In previous studies, the researchers investigated Lexical bundles in multiply types of texts including academic texts across disciplines, journals etc. However, while choosing across disciplinary input data in previous studies, researchers never emphasized balanced data statistics. But in the past researches, the domain of journal categories had been neglected. The researchers used to choose prestigious journals to check the difference in disciplines, however the need for awareness in the ESP domain for material development as well as for student and teacher comprehension of a particular discipline, especially for those whose L2 is English, is still unmet. Additionally, it aids researchers in selecting the appropriate feature for a particular article type.

## 1.3 Research Question

1. Does on the base of the ranking of the research journal (W, X, and Y), any effect count on the PTOLBs, if the difference what are its frequency and statistical differences?

## 1.4 Research Objectives and its Significance

Since it offers two distinct objectives that together make up the study's goals, it will set itself apart from earlier studies. Determining the PTOLBs in academic research publications of the arts and sciences is thus the first objective. To comprehend the PTOLBs from a larger viewpoint in terms of research article journals, i.e., W, X, and Y, this study analyses additional elements or variables that may be associated with Text-Oriented Lexical Bundles as well

With the alteration of the journal categories, the comparison will assist researchers in identifying fixed phrases, multiword expressions, or formulaic sentences. Additionally, the multidisciplinary analysis helped with EAP teachings. Researchers can discern between disciplinary standards and generic language aspects by understanding linguistic distinctions in journal category. Teachers and EAP instructors can create materials that are appropriate for

ISSN Online: 2709-7625
ISSN Print: 2709-7615

students' skill levels and genres with the support of these multidisciplinary analyses and research design comparisons.

Additionally, this research will help teachers and L2 students understand how different fields and genres influence multiword expression (also known as formulaic sentences). Understanding those register-specific elements aids L2 students in understanding various genres and enhances their writing ability as a result, corpus-driven examination of linguistic variance will open a variety of chances in English for scholarly endeavors and educational advancement.

# 2. Literature Review

## 2.1 Phraseology and its Scope

According to Cowie (1994), phraseology may be defined as the scholarly examination of the arrangement, significance, and use of combinations of words. The area of phraseology has garnered significant attention and research efforts in recent times. However, its progress is hindered by the lack of agreement on nomenclature, descriptive methodologies, and analytical processes (Granger & Paquot, 2008).

Biber et al. (2004) argue that the inclusion of diverse research methods and perspectives is necessary for gaining a comprehensive understanding of a complex phenomenon like phraseology Various academics have assigned various labels to phraseological components, including lexical phrases, formulae, routines, fixed expressions, preset patterns, and lexical bundles. However, it is also acknowledged by Granger and Paquot (2008) that this diversity can impede effective communication among linguists and contribute to a sense of ambiguity within the field. Granger and Paquot (2008) establish a connection between the changeable scope of phraseology and its indeterminate bounds with four interconnected disciplines, namely semantics, morphology, syntax, and discourse.

## 2.2 Phraseology in Academic Writing

The study of phraseology using corpora has recently provided valuable insights into the significance of frequent multi-word combinations in defining registers, genres, and disciplines. Various research studies have emphasized the significance of fixed phrases in specific discourse communities. According to Hyland (2008a), there is a tendency for words to occur nearby more often than would be predicted by random chance.

Previous studies have examined the characteristics of formulaic language in academic speech and writing across various research fields. These investigations have shed light on this topic and have been conducted by researchers such as (DeCarrico & Nattinger, 1988; Oakey, 2002). Some of these studies have specifically focused on scientific genres, including the works of (Gledhill, 1995; Williams, 1998). These studies provide evidence for the functional importance of frequently recurring sequences of words in academic discourse. According to Williams (2002), a thorough examination of texts is necessary to identify the specific lexico-grammatical tactics used to facilitate communication within a specialized group.

Given the aforementioned study results, it is becoming widely recognized that phraseological units play a crucial role in facilitating coherent communication within academic contexts. In the following section, the present study will now shift our focus towards a particular

Vol. 6 No.4 2023
ISSN Online: 2709-7625
ISSN Print: 2709-7615

category of multi-word units that has garnered significant attention in several research endeavors, including the academic domain. This category is often referred to as lexical bundles, and it has been the topic of numerous pioneering investigations conducted in diverse contexts.

## 2.3 Lexical Bundles

Lexical bundles (also known as, LBs) refer to clusters of three or more words that often appear across the whole of a given text (Biber et al., 1999). Sequences in phraseology may consist of entire phrases whose meaning can be deduced from their constituent parts, or they may be unfinished units. The classification of words into groups is based only on their separation and frequency of occurrence in texts. In contrast, idioms are whole phrases that possess a meaning that is unrelated to the individual components of the sentence (Wei & Lei, 2011).

Biber and Barbieri (2007) argue that LBs have significant importance as fundamental constituents of speech. The present study has shown variations in the frequency of bundles between different registers as well as between spoken and written forms. They also examined bundles in particular university registers, they discovered that writing in course management and instructional registers exhibited a greater frequency of bundles compared to the spoken registers they analyzed. This finding contradicts the prevailing notion that spoken language generally contains a higher number of LBs than written language, as suggested by Biber et al. (1999).

Meaningful circumstances provide an additional classification for the grouping of words. In academic writing, LBs fulfill three primary roles: text-oriented, participant-oriented, and research-oriented. The text-oriented function, also known as discourse organizers, establishes connections between various sections of the text. Participant-oriented bundles also referred to as stance expressions, convey the author's attitudes or evaluations of another proposition. Lastly, research-oriented bundles, or referential expressions, make explicit or implicit references to other texts (Biber et al., 2002).

The use of LBs might be influenced by the specific field in which the author is writing, due to the many purposes that these bundles serve (Hyland, 2008a). Therefore, although LBs are crucial components of discourse construction, different types of speech and writing need significantly diverse lexical bundles. A more comprehensive comprehension of the inherent distinctions among these bundles may be achieved by categorizing the forms, structures, and functions of bundles across various registers, topics of study, and levels of competence.

The identification and categorization of lexical units can only be achieved via the use of corpus linguistics approaches, which include the use of computer-assisted functions that analyze bundles within their respective speech contexts. The identification of LBs is contingent upon two frequency elements, namely the frequency of occurrence of the bundles and the types of texts in which they are seen.

It is essential to identify and record all conceivable combinations of items, as well as document their occurrences inside each paragraph. Achieving this task for a substantial volume of texts has challenges that cannot be overcome without the use of computer-assisted technologies. The identification and quantification of probable LBs may be achieved by the use of computer methods known as concordances (Cortes, 2004).

Vol. 6 No.4 2023
ISSN Online: 2709-7625
ISSN Print: 2709-7615

The determination of statistical significance is contingent upon the specific kind of corpus being analyzed. Bundles are more prevalent in spoken language compared to written language. Consequently, when analyzing spoken data within extensive corpora exceeding one million words, researchers often focus on identifying bundles that manifest at a frequency of at least 40 occurrences per million words and are present throughout a minimum of five texts.

The need for variety arises from the fact that the presence of LBs in several authors' writings is a stronger indication of a formulaic sequence rather than the idiosyncrasies of the individual author (Biber, 2009). The frequency of recurrence for written content is often lower, ranging from 20 to 25 instances per million words, and is distributed throughout a minimum of five texts (Biber et al., 1999).

## 2.4 Lexical Bundles Within the Articles' Sections

Research papers get significant attention in the realm of academic writing, alongside general academic writing and writing produced by individuals who are non-native speakers of the language. Several scholars have provided examples of relevant research in the field. (Callies, 2013; Hu & Huang, 2017; Liu & Chen, 2020; Nekrasova, 2009; Qi & Pan, 2020) are among the researchers who have contributed to the literature

Research articles, often known as RAs, are a kind of scholarly writing that seeks to disseminate original ideas and findings among the academic community (Flowerdew, 2005; Kanoksilapatham, 2003; Kim, 2014,). According to Swales (1990), RAs can be described as concise accounts of the research endeavors undertaken by the author or authors. Proficiency in various academic writing skills and methodologies related to the production of RAs is essential for individuals in academia, including students, educators, researchers, and professors, to establish themselves as credible experts in their respective fields. Hence, it may be argued that RAs have paramount significance within the global academic discourse community. The establishment of an academic reputation via the dissemination of research results is of utmost importance (Swales, 1990).

In addition, research assistants provide a wide range of models, each including several disciplines. For example, the majority of RAs use the IMRD framework as proposed by Swales (1990), including four main sections: introduction, methods, results, and discussion/conclusion. However, the IMRD structure incorporates the Literature Review (LR) as a crucial element in the domain of applied linguistics, as noted by Jian (2008). Additional research is required to investigate the validity, application, and effectiveness of the model proposed by Ruiying and Allison (2004) for analyzing secondary RAs, specifically those of a theoretical nature. This need arises despite the widespread use and evaluation of the conventional Introduction, Methods, Results, and Discussion (IMRD) model in many studies spanning several fields.

Based on Hyland's (2008b) classification system, a functional investigation was carried out by Geoffrey et al. (2022). This study built a corpus of academic discourse by using discussion sections from 150 RAs. These publications included various research designs, including quantitative, qualitative, and mixed methods, and were published in high-impact journals between the years 2015 and 2018. The researchers have conducted individual analyses on TOLBs, which refer to the structural components of a text that convey a certain message.

Vol. 6 No.4 2023
ISSN Online: 2709-7625
ISSN Print: 2709-7615

Once the data for each bundle was collected individually, they proceeded to categorize them into functional groups.

In the previous studies, Text-oriented lexical bundles as the functional categories had been explored widely in range of genre and disciplines of academic writing. Most of the studies had selected and observed the academic genre only in high ranked journals. Some studies used insignificant and unbalanced corpora for their studies. In the present study, the corpus is collected balanced data from all range of journals high to low ranked (almost 600 RAs) which are scrutinized by HEC' HJRS. This study examined the normal distribution of PTOLBs among these RAs.

# 3. Methodology

## 3.1 Research Design

In the present corpus-based study, quantitative research design is used to investigate quantitative RAs from five disciples and 11 subjects. This study has one independent variable, i.e. Journal category and one dependent variable, PTOLBs. The journal category of the RAs into three distinct classifications in given criteria provided by HEC Pakistan through the online system known as the HEC Journal Recognition System (HJRS): W (representing the highest level of prestige), X (indicating a lower level of prestige compared to W), and Y (representing publications that are relatively unknown or obscure). The dependent variable is further divided into nine subcategories, named as Additive, Causative, Comparative, Inferential, Framing, Structing, Framing, Objective, and Generalization.

## 3.2 Data Collection

The categorization of journals in this study was conducted using the HEC Journal Recognition System (HJRS), a tool that evaluates the quality of articles across different academic fields. The classification of journals by the HJRS involves categorizing them into three unique groups, namely W, X, and Y, which are determined by the quality and influence of their particular fields.

Each subject in this study was defined by a set of 60 Research Articles, 20 of each category The articles chosen for each journal category were selected according to the list supplied on the HEC HJRS website. Research papers were downloaded from HJRS-categorized journals of W, X, and Y. Additionally, the procedure for picking research papers followed a set of specified criteria that comprised three fundamental components. Initially, the selected publications must adhere to the usual pattern of Introduction, Method, Results, and Discussion (IMRD). In addition, it was necessary for the selected works to be relevant to the fields of 11 subjects (i.e., Applied Linguistics, Biophysics, Biochemistry, Chemistry, Environmental Science, Ecology, Geography, Pharmacology, Psychology, Physics, and Pathology) and have been published between the timeframe of 2019 to 2023.

## 3.3 Cleaning of the Corpus

The preliminary stage of the investigation included the acquisition of Research Articles in PDF format from several journal websites and internet sources. Following that, all RAs were transformed into Microsoft Word format to streamline the process of data cleansing. The primary objective of article alteration is to standardize the material by using four main characteristics.

ISSN Online: 2709-7625
ISSN Print: 2709-7615
Vol. 6 No.4 2023

Initially, it is noteworthy that various forms of citations, including integral or non-integral citations, numerical citations, and parenthetical citations, were uniformly substituted with a singular term, namely "Ref.". Additionally, the material underwent the removal of formulae and analytical models, resulting in a reduction in word counts. Furthermore, the study omitted certain features, such as figures, tables, and headers, that had a descriptive character

## 3.4 Data Analysis

The normalization of dependent variables (to 1000 words) was the initial stage in data analysis. In corpus analysis, normalization is required to mitigate the impact of varying text lengths across files. Due to insufficient sample size, we ran nine separate Kruskal-Wallis H tests (1x9) on the data to see whether there was any significant relationship between the independent factor i.e. journal category (classified into W, X, and Y) and the nine dependent variables. Before settling on this option, however, we double-checked that we could indeed do MANOVA and ANOVA.

### *3.4.1 Checking the Assumptions of MANOVA*

At first glance, the data seemed suitable for a 3-factor MANOVA due to the presence of one independent variable i.e., journal category and nine dependent variables i.e., nine subcategories of PTOLBs. However, after verifying the MANOVA's assumptions, it became clear that MANOVA was not appropriate due to a large number of outliers in each dependent variable. In a high proportion of cases, all variables were zero, marking them as outliers. In addition, normality tests revealed that none of the nine dependent variables had a normal distribution ($p > .05$). We could not use MANOVA since not all of our variables followed a normal distribution ($p > .05$).

### *3.4.2 Checking the Assumption of ANOVA*

To investigate the viability of using 3-factor ANOVA to the data, we aggregated the normed scores of the nine variables into a single composite variable (called "TTL_Normed1000"), each of whose components was a subcategory of the parent variable (PTOLBs). The results of an outlier analysis that found multiple cases to be outliers within the dependent variable. Inspecting the range of variables with zero values, we found that in 150 cases, the minimum value was zero. There is no longer any evidence of these occurrences. Also eliminated were nine other cases with very high values. The final tally of successful removals is 159. The results of the normality tests performed after these values were subtracted. There was significant evidence that the data did not follow a normal distribution ($p = .000$).

### *3.4.3 Choice of Kruskal- Wallis H test*

To examine the impact of journal category on nine categories of prototype Text-oriented bundles (1X9), researchers used nine separate Kruskal-Wallis H tests since the data did not meet the assumptions of a one-way or two-way analysis of variance (MANOVA) or analysis of covariance (ANOVA). The independent variables (journal category) had three categories, necessitating the use of Post Hoc tests to further examine the differences.

# 4. Result and Discussion

The following are two main divisions in this chapter. The first subsection (4.1) provides descriptive data for the nine PTOLBs and three independent variables (study design, journal

Vol. 6 No.4 2023
ISSN Online: 2709-7625
ISSN Print: 2709-7615

category, and RAs section). In Section 4.2, we provide the results of the study and give a comprehensive analysis of the three main hypotheses.

## 4.1 Descriptive Statistics

The part about descriptive statistics presents the frequencies of three independent variables and nine dependent variables.

### *4.1.1 Frequency Distribution of Independent Variables: Research Journal Category*

This section provides an analysis of the frequency distribution of variable that is considered to be independent variable. The independent variable being examined pertains to the Journal categories (i.e., W, X, and Y) as shown in table 1.

***Table 1: Frequency Distribution of Research Articles by Journal Category***

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | W Category Journals | 868 | 33.7 | 33.7 | 33.7 |
| | X Category Journals | 868 | 33.7 | 33.7 | 67.4 |
| | Y Category Journals | 840 | 32.6 | 32.6 | 100.0 |
| | Total | 2576 | 100.0 | 100.0 | |

According to the table 1 provided, the variable is comprised of three distinct levels, namely W, X, and Y. In a manner akin to the concept of frequency distribution within the realm of study design, it is seen that nearly the same quantity of texts is sourced from journals falling under the W category (867, 33.7%), X category (869, 33.7%), and Y category (840, 32.6%).

### *4.1.2 Frequency Distribution of Dependent Variables: Prototypical Text-Oriented Lexical Bundles*

This part presents the descriptive findings of the nine dependent variables, which correspond to nine subcategories of PTOLBs. These results are preceded by the raw and normed frequencies of various subcategories of lexical bundles. This comparison allows for a contextual understanding of the findings of previous research.

***Table 2: Raw and Normed Frequency Distribution of Subtypes of T0LBs in Corpus***

| Subcategories of Text-oriented Lexical Bundles | Raw (out of 2967056 words) | Normed (million words) | Normed (thousand words) |
|---|---|---|---|
| ADD_LBs | 2271 | 765.41 | 7.6 |
| CAUSE_LBs | 291 | 98.1 | 0.098 |
| COMP_LBs | 932 | 314.12 | 3.1 |
| INF_LBs | 671 | 226.15 | 2.2 |
| STRUCT_LBs | 2281 | 768.77 | 7.6 |
| FRAME_LBs | 1806 | 608.68 | 6.1 |
| CITE_LBs | 662 | 223.12 | 2.2 |

ISSN Online: 2709-7625
ISSN Print: 2709-7615
Vol. 6 No.4 2023

| | | | |
|---|---|---|---|
| GENERAL_LBs | 21 | 7.0777 | 0.007 |
| OBJECT_LBs | 64 | 21.57 | 0.021 |
| | **Total freq. of** Text-oriented Lexical Bundles **per million words** | **6460.9** | |

The corpus is comprised of a total of 2,967,056 tokens, as in Table 2. The normed frequencies indicate that prototypical text-oriented lexical bundles have the highest frequency. Zare and Valipouri (2022) conducted a study in which they gathered a corpus consisting of 3,751,006 tokens. However, the authors did not provide any standardized values for TOLBs. Similarly, Yang (2022) collected a corpus including 1,122,690 tokens, but, like the previous study, did not supply normed values. Xia (2022) conducted a data collection process that included gathering 127,875 tokens, which were then normalized based on a per million words metric. Table 3 presents descriptive findings for a total of nine dependent variables. The corpus consisting of 409,210 units was gathered and then standardized based on a per million words basis (Varghaei & Khodadadi, 2022)

***Table 3: Descriptive Statistics of Nine Subcategories of PTOLBs***

| | | ADD_LBS_Normed1000 | CAUSE_LBs_Normed1000 | COMP_LBs_Normed1000 | INF_LBS_Normed1000 | STRUCT_LBs_Normed1000 | FRAME_LBs_Normed1000 | CITE_LBs_Normed1000 | GENERAL_LBs_Normed1000 | OBJECT_LBs_Normed1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| N | Valid | 2576 | 2576 | 2576 | 2576 | 2576 | 2576 | 2576 | 2576 | 2576 |
| | Missing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | | .8285 | .1169 | .3082 | .2253 | .9169 | .6111 | .2356 | .0083 | .0205 |
| Median | | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |
| Std. Deviation | | 1.55594 | .43982 | .95825 | .62245 | 1.68834 | 1.27362 | .65689 | .11296 | .21296 |
| Skewness | | 6.357 | 5.203 | 10.693 | 4.993 | 4.721 | 6.982 | 4.143 | 15.644 | 18.060 |
| Std. Error of Skewness | | .048 | .048 | .048 | .048 | .048 | .048 | .048 | .048 | .048 |
| Kurtosis | | 82.827 | 36.065 | 187.756 | 47.160 | 45.347 | 103.425 | 22.819 | 262.089 | 416.244 |
| Std. Error of Kurtosis | | .096 | .096 | .096 | .096 | .096 | .096 | .096 | .096 | .096 |
| Minimum | | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| Maximum | | 30.69 | 6.25 | 21.82 | 10.95 | 29.20 | 27.31 | 6.76 | 2.19 | 6.10 |

The absence of similarity between the mean, mode, and median, and the presence of significant kurtosis and skewness values in all instances indicate deviations from a normal distribution, as shown in Table 3. This observation is reinforced by the results of normality tests previously discussed in Chapter 3. Previous research in statistical analysis has used the chi-

ISSN Online: 2709-7625
ISSN Print: 2709-7615
Vol. 6 No.4 2023

square test as a method of analysis (Xia, 2022; Yang, 2022). Xia (2022) used a t-test for the analyses as well. Several studies have used Paul Rayson and log Likelihood as analytical tools to investigate Text-oriented lexical bundles (Akbulut, 2020). Previous studies have shown problems in the validity of their findings.

## 4.2 Results of Inferential Statistics

This part presents the findings of the hypotheses made for the research and provides a comprehensive analysis and interpretation of the data.

### *4.2.1 Effect of Journal Category on Frequency of Prototypical Text-Oriented Lexical Bundles*

The study examined the impact of journal category on the occurrence rate of nine subcategories of PTOLBs, namely Additive LBs, Causative LBs, Comparative LBs, Inferential LBs, Structuring LBs, Framing LBs, Citation LBs, Generalization LBs, and Objective LBs.

Null Hypothesis: The distribution of subcategories of PTOLBs is the same across journal categories (W, X, and Y).

Table 4 provide the results of a Kruskal-Wallis H test that was performed to examine the results about the variation in PTOLBs score among articles categorized under distinct research journals (W, X, and Y).

***Table 4: Test Statistics[a,b] Effect of Journal categories (W, X and Y) of Research Articles on Subcategories of Prototypical Text-Oriented Lexical Bundles***

| | ADD_LBS_Normed1000 | CAUSE_LBs_Normed1000 | COMP_LBs_Normed1000 | INF_LBS_Normed1000 | STRUCT_LBs_Normed1000 | FRAME_LBs_Normed1000 | CITE_LBs_Normed1000 | GENERAL_LBs_Normed1000 | OBJECT_LBs_Normed1000 |
|---|---|---|---|---|---|---|---|---|---|
| Chi-Square | 1.633 | 2.038 | .686 | 1.899 | 1.684 | 2.198 | 3.153 | .412 | 10.412 |
| Df | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .442 | .361 | .710 | .387 | .431 | .333 | .207 | .814 | .005 |

a. Kruskal Wallis Test
b. Grouping Variable: Journal Category

As Table 4 reveals, there was no statistically significant difference among journal category i.e., W, X and Y in one out of 9 subcategories of PTOLBs (ADD_LBs: $\chi2(1) = 1.633$, $p = 0.442$; CAUSE_LBs: $\chi2(1) = 2.038$, $p = 0.361$; COMP_LBs: $\chi2(1) = 0.686$, $p = 0.710$; INF_LBs: $\chi2(1) = 1.899$, $p = 0.387$; STRUCT_LBs: $\chi2(1) = 1.684$, $p = 0.431$; FRAME_LBs: $\chi2(1) = 2.198$, $p = 0.333$; CITE_LBs: $\chi2(1) = 3.153$, $p = 0.207$; GENERAL_LBs: $\chi2(1) = 0.412$, $p = 0.814$; OBJECT_LBs: $\chi2(1) = 10.412$, $p = 0.005$). The difference in OBJECT_LBs in which the value of $P< 0.05$, the value is 0.005. Thus, the null hypothesis has been in general rejected in the OBJECT_LBs subcategory of Prototypical Text-Oriented Lexical Bundles.

Research journals play a crucial role in aiding the distribution of academics' discoveries, insights, and developments to a worldwide audience. Sharing information facilitates the opportunity for others to expand upon pre-existing research, enriching the common comprehension across many disciplines. In previous studies, scholars used esteemed academic publications as a means to identify lexical bundles. However, it should be noted that these studies acknowledged the potential for variations in the incidence of lexical bundles if data were

collected from both prestigious and obscure journals (Al-Shujairi et al., 2020; Geoffrey et al., 2022; Ren, 2021; Wright, 2019; Zare & Valipouri,2022). Researchers also used the peer-reviewed criterion for assessing the credibility of publications (Budiwiyanto & Suhardijanto, 2020; Ren, 2021).

The research is not concluded at this point; more examination is required to determine the impact of differences between national and international corpora on the frequency of lexical bundles. The study on lexical bundles was undertaken by Ghorbani et al. (2022) utilizing a selection of five foreign journals and five Iranian journals. In their study, Belghalem and Melgani (2022) examined the differences between Algerian native journals (L2 corpus) and two specific sources, namely JSTOR and the Journal of Sociolinguistics, which constituted the first sub-corpus (L1 corpus). In Salazar's study conducted in 2011, the researcher used two sources for their LBs' analysis: the Phillippe English language journal, which is native to the English-speaking community, and the British medical magazine, which has an international readership.

Similarly, in 2020, Güngör and Uysal replicated Salazar's (2014) approach by examining journals written by native English academics as well as Turkish researchers, with English being the first language of the former group. This replication study of native and non-native also referenced Salazar's work from 2014. In another study conducted, Kurniawan and Haerunisa (2023) used both approved and rejected research publications from the Indonesian Journal of Applied Linguistics. The acceptance or rejection of research publications is contingent upon the writing style used by researchers. In the present research, variations were seen across the different journal categories. However, previous studies did not investigate the categorization of journals according to the criteria set by HJRS. Furthermore, the linguistic elements of both high-ranked and low-ranked journals were not studied by any of the researchers to identify potential writing trends.

However, previous studies have mostly focused on conducting analyses using research publications authored by both native and non-native speakers. It is well recognized that W-category journals primarily focus on the use of native language, whereas X-category journals partly address the topic of native-like language usage. However, the Y-category journals exhibit a greater degree of non-native-like characteristics. Previous studies have shown that there is a higher prevalence of additive LBs in research articles on RAs published in non-native journals (Akbulut, 2020; Belghalem and Melgani, 2020; Gungor and Uysal, 2020). In research conducted by Belghalem and Melgani (2020), it was shown that non-native journals had greater levels of causal LBs. This finding aligns with a previous study by Gungor and Uysal (2020), which reported a lower prevalence of causal LBs in non-native contexts.

According to a previous study conducted by Belghalem and Melgani (2022), it was shown that non-native publications tend to have a lower frequency of comparative LBs compared to native journals. According to Esfandiari and Barbary (2017), there was a higher frequency of comparison signals seen in Persian research articles (by non-native speakers of English) compared to English research articles (by native speakers of English) (18.5% vs. 27.6%). The use of these bundles has significance in achieving a level of fluency similar to that of native speakers in the context of academic writing (Belghalem and Melgani, 2022). According

ISSN Online: 2709-7625
ISSN Print: 2709-7615
Vol. 6 No.4 2023

to Güngor and Uysal (2020), the frequency of comparison bundles in MCRA was found to be lower compared to the data obtained from Turkish authors.

According to Belghalem and Melgani (2022), the number of inferential bundles possessed by foreign RAs was found to be lower compared to Algerian RAs. The use of these bundles has significant importance in achieving a writing style that closely resembles that of a native speaker within academic contexts. According to Esfandiari and Barbary (2017), it was shown that Persian authors had a higher propensity for using inferential bundles (35%), compared to English writers (10%). This disparity was mostly seen via the use of "there" patterns as a means to indicate the existence of something.

The patterns exhibited distinct roles in the two languages, indicating nuanced syntactic variations. According to Güngor and Uysal (2020), the number of inferential bundles seen in MCRA was found to be comparatively lower than that observed in the data derived from Turkish authors. In a study conducted in 2020, Akbulut examined the use of lexical bundles (LBs) by native speakers (NSs) and non-native speakers (NNSs) across several language categories, revealing the emergence of certain patterns. One notable disparity is that non-native speakers (NNSs) exhibit a much higher frequency of using inferential packages compared to native speakers (NSs), with the former employing such packages approximately twice as often as the latter.

According to Güngor and Uysal (2020), the number of structural bundles seen in MCRA was comparatively lower when compared to the data obtained from Turkish authors. In their study, Budiwiyanto and Suhardijanto (2020) identified a total of 49 occurrences of structuring bundles among the Indonesian research articles they examined. Reflexive textual markers were used to structure the text and direct the reader's attention towards certain sections. In their recent study, Belghalem and Melgani (2022) analyzed Algerian and international RAs to examine the disparity in the use of text-oriented LBs. Their findings revealed that framing LBs were more prevalent in Algerian RAs compared to international RAs.

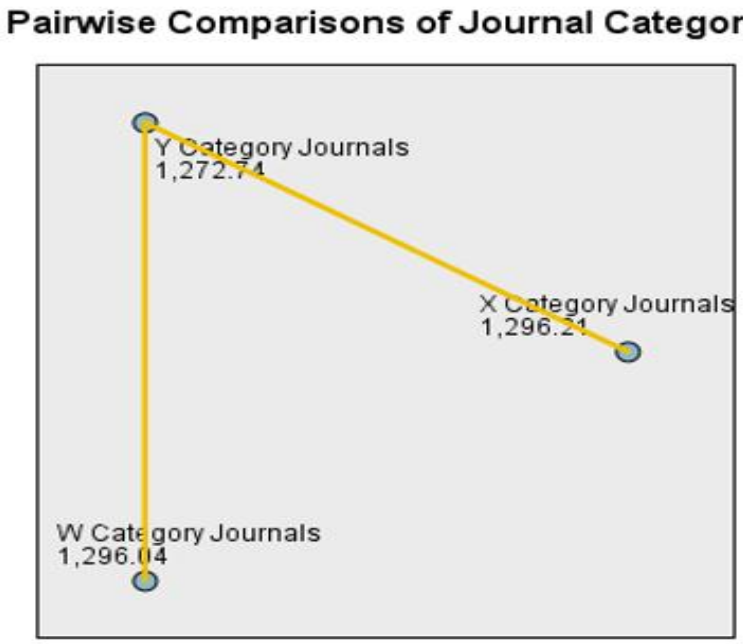
According to Akbulut (2020), there is variation in the framing bundles used by native speakers (NSs) and non-native speakers (NNSs). NSs tend to depend on prepositional phrase-based LBs, while NNSs show a preference for verb phrase-based LBs. Nevertheless, the chosen research papers for analysis did not divide the articles into parts. To verify which categories of journals, have statistically significant differences in the occurrence of OBJECT_LBs, a Post Hoc analysis was conducted, the results of which appear below.

### *4.2.2 Results of Post Hoc Test Among W, X, and Y Category Journals on OBJECT_LBs.*

The findings of the Post Hoc analysis are presented in both visual and tabular formats, as seen below. The graphic representation displays the average rank scores attained by different categories of journals concerning OBJECT_LBs. The average rank scores attained by journals in the W, X, and Y categories are around 1296, 1226, and 1272, respectively. The figure 1 provides pairwise data indicating statistically significant differences.

***Figure 1: Post Hoc analysis of statistical significance Test Statistics[a,b]: Effect of Journal category on Frequency of OBJECT_LB***

ISSN Online: 2709-7625
ISSN Print: 2709-7615
Vol. 6 No.4 2023




The findings of the Post Hoc analysis indicate a statistically significant distinction between journals categorized as W and Y ($p$ = .016), as well as between journals categorized as X and Y ($p$ = .015) in table 6 given bellow. However, no statistically significant difference was observed between W and X category journals in relation to this aspect.

**Table 6:** ***Pairwise Comparison of Journal Category on the Frequency of Object _LBs***

Each node shows the sample average rank of Journal Category.

| Sample1-Sample2 | Test Statistic | Std. Error | Std. Test Statistic | Sig. | Adj.Sig. |
|---|---|---|---|---|---|
| Y Category Journals-W Category Journals | 23.297 | 8.345 | 2.792 | .005 | .016 |
| Y Category Journals-X Category Journals | 23.471 | 8.345 | 2.813 | .005 | .015 |
| W Category Journals-X Category Journals | -.175 | 8.277 | -.021 | .983 | 1.000 |

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Budiwiyanto and Suhardijanto (2020) underscored the significance of lexical bundles in the context of Indonesian academic writing which is non-native. A collection of 47 objective bundles was identified, each serving as a means of conveying the author's intention behind the

creation of the work. In another study, Taieb and Toumi (2021) discovered a significant discrepancy in the prevalence of text-oriented bundles between the PC data collection and the EWC across several functional subcategories. The use of the "to" tri-gram with verb components was seen as significant among student authors, particularly those hailing from Tunisia. This observation aligns with previous studies conducted by Chen and Baker (2010) as well as Lu and Deng (2019), who also identified the prevalent usage of the tri-gram "to" in L2 student writing. Figure 1 displays a visual representation of the distinctions between journal categories W, X, and Y in terms of objective lexical bundles.

In their study conducted in 2018, Jalilifar and Ghoreishi examined formulaic sequences from a cross-disciplinary standpoint while also considering an overall viewpoint. Textual operations were prevalent in both general and specialized sequences, including 58.51 percent and 55.18 percent, respectively. In particular, objective sequences, consisting of 5.05% general sequences and 5.32% specialized sequences, exhibit similar patterns. In the year 2020, Akbulut conducted a comparative analysis of the use of lexical bundles (LBs) in different linguistic settings by native speakers (NSs) and non-native speakers (NNSs). Both groups use objective bundles rarely, but non-native speakers (NNSs) tend to depend more on prepositional phrase (PP)--based bundles such as "to be/improve" compared to speakers of other languages. In a study conducted by Kurniawa and Haerunisa (2023), it was observed that the objective bundles had the most significant proportional variance. These objective bundles are used to introduce the authors or outline the aims of the research. In terms of raw frequency, it was seen that the occurrences of RRAIs (20 instances) exceeded the occurrences of ARAIs (12 instances).

# 5. Conclusion

In this research, the investigation of the distribution of PTOLBs in quantitative research articles. As we know this research is completely focused on journal categories which is given by HEC's HJRS. The objective lexical bundles showed the statistical difference in journal categories that was $p = 0.005$. The difference among the journal categories was further elaborated through the post hoc analysis. This visual representation (Figure 1) showed the significant difference in-between W and Y categories of journals is $p = 0.016$ and value in between Y and X categories of journals is $p = 0.015$. However, other than OBJECT_LBs, none of the subcategories showed up any results. The corpus used in this research was derived from a diverse range of academic fields, including social science, physical science, biological science, and medical science. The present research, which centers its attention on five overarching disciplines, is situated within the intermediate range of specialization within these fields. The examination of PTOLBs that have unique significance within various areas presents a promising avenue for further scholarly investigation, notwithstanding our omission of analysis of inter-disciplinary differences.

The present research has effects on education for several academic disciplines, with a special focus on its pertinence to ESP (English for specified purposes) and EAP (English for academic purposes) compositions. It is advisable for English for Academic Purposes (EAP) learners and instructors to explore the educational merits of the PTOLBs identified in this research. The findings of this research may serve as a valuable tool for students enrolled in English for Academic Purposes (EAP) classes that use a genre-based approach to teach academic writing. The results obtained from this study have provided a foundation for our continuing research,

which aims to examine the alignment between the distribution of PTOLBs in research papers and the IMRD structure.

## 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.

Akbulut, F. D. (2020). A bibliometric analysis of lexical bundles usage in native and non-native academic writing. *Journal of Language and Linguistic Studies*, *16*(3), 1146-1166.

Al-Shujairi, Y. B. J., Tan, H., Abdullah, A. N., Nimehchisalem, V., & Imm, L. G. (2020). Lexical bundles in the discussion section moves of high-impact medical research articles. *Social Sciences & Humanities*, *28*(3), 2043-2061.

Belghalem, B., & Melgani, H. (2022). A Corpus-based study of lexical bundles in the research papers of Algerian and native english speaking researchers.

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.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL quarterly, 36*(1), 9-48.

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

Biber, D., S. Conrad, and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. New York: Cambridge University Press.

Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, *10*(2).

Callies, M. (2013). Agentivity as a determinant of lexico-grammatical variation in L2 academic writing. *International Journal of Corpus Linguistics*, *18*(3), 357-390.

Chen, Y., & 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.

DeCarrico, J., & Nattinger, J. R. (1988). Lexical phrases for the comprehension of academic lectures. *English for specific purposes*, *7*(2), 91-102.

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.

Flowerdew, J. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, *24*(3), 321–332.

ISSN Online: 2709-7625

ISSN Print: 2709-7615

Geoffrey, R. K., Behruz, L. G., & Milad, M. (2022). A functional analysis of lexical bundles in the discussion sections of applied linguistics research articles: A cross-paradigm study. *Russian Journal of Linguistics*, *26*(3), 625-644.

Ghorbani, A., Okati, F., & Lotfi Gaskaree, B. (2022). Lexical Bundles in the Conclusion Section of Research Articles in Applied Linguistics: A Comparative Study of International and Iranian Journals. *Critical Applied Linguistics Studies*, *1*(1), 67-84.

Gledhill, C. (1995). Collocation and genre analysis. The discourse function of collocation in cancer research abstracts and articles. *Zeitschrift für Anglistik und Amerikanistik*, *1*, 1-26.

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

Gries, S. T. (2010). Useful Statistics for Corpus Linguistics. *A Mosaic of Corpus Linguistics*: *Selected Approaches, 66*: 269–291.

Güngör, F., & Uysal, H. H. (2020). Lexical bundle use and crosslinguistic influence in academic texts. *Lingua*, *242*, 102859.

Hu, X., & Huang, Y. (2017). A contrastive study on the use of common high-frequency lexical bundles in different moves of English abstracts in Chinese and English engineering journals. *For. Lang. Learn. Theory Pract*, *4*, 31-36.

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.

Jalilifar, A., & Ghoreishi, S. M. (2018). From the perspective of: Functional analysis of lexical bundles in Applied Linguistics research articles. *International Journal of English Studies*, *18*(2), 161-186.

Jian, H. U. (2008). The macrostructure and rhetorical structures of research articles in linguistics. *Shandong Foreign Language Teaching Journal, 5,* 37–42.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, *24*(3), 269–292.

Kim, E. (2014). An Analysis of Move Patterns in Abstracts of Social Sciences Research Articles. *Journal of Korean Library and Information Science Society*, *45*(2), pp.283-309.

Kurniawan, E., & Haerunisa, Z. F. (2023). A comparative study of lexical bundles in accepted and rejected applied linguistic research article introductions. *Studies in English Language and Education*, *10*(2), 628-648.

Liu, C. Y., & Chen, H. J. H. (2020). Functional variation of lexical bundles in academic lectures and TED talks. *Register Studies*, *2*(2), 176-208.

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.

Nekrasova, T. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Lang. Learn., 59,* 674–68.

Oakey, D. (2002). A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. *Using corpora to explore linguistic variation*, 111-129.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 100*(21), 60-71.

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.

Qi, H., & Pan, F. (2020). Lexical bundle variation across moves in abstracts of medical research articles. *Southern Afri. Ling. Appl. Lang. Stud., 38*, 109–128.

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.

Ruiying, Y., & Allison, D. (2004). Research articles in applied linguistics: Structures from a functional perspective. *English for specific Purposes*, *23*(3), 264-279.

Salazar, D. (2011). Lexical bundles in scientific English: *A corpus-based study of native and nonnative writing* [Unpublished doctoral dissertation]. University of Barcelona.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). John Benjamins Publishing Company

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. Routledge.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Taieb, A. B., & Toumi, N. (2021). Exploring Conventionalized Phraseology in Advanced Academic Writing of Tunisian Linguistics Students: A Lexical Bundles Analysis. *Arab Journal of Applied Linguistics*, *6*(2), n2.

Titone, D. A., and Connine, C. M. (1999). On the compositional and non-compositional nature of idiomatic expressions. *Journal of Pragmatics, 31*(12), 1655–1674.

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).

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

Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, *3*(1), 151-171.

Williams, G. (2002). In search of representativity in specialised corpora: Categorisation through collocation. *International Journal of Corpus Linguistics*, *7*(1), 43-64.

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

Xia, D. (2022). Formulaic Sequences in Business and Academic Writing of English Learners (Doctoral dissertation, University of Cincinnati).

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*.

Zare, J., & Valipouri, L. (2022). Lexical Bundles in Chemistry Research Articles. *Frontiers in Psychology*, *13*, 906641.