

EFFECT OF JOURNAL CATEGORY ON THE DISTRIBUTION OF PHRASE-FRAMES: A CORPUS-BASED STUDY OF QUANTITATIVE RESEARCH ARTICLES

Mohammad Ishaq

MPhil Scholar, Department of Applied Linguistics, Government College University Faisalabad,
Email: hafizmishaq4441@gmail.com

Dr. Aleem Shakir

Assistant Professor, Department of Applied Linguistics, Government College University Faisalabad, Email: almsha@yahoo.com

ABSTRACT

This study aimed to investigate the effect of journal category on the distribution of phrase-frames (PFs) within quantitative research articles. Employing a corpus-based methodology, the study collected a corpus of approximately 3 million words from research articles that covered eleven different artistic and scientific fields. The study utilized an established phrase-frames list for the analysis (Lu et al., 2018). The distribution of functional phrase-frames; referential phrase-frames (REF_PFs), discourse phrase-frames (DISC_PFs), and stance phrase-frames (STNC_PFs), was examined within experimental and correlational research designs across W, X, and Y journal categories. The statistical findings of the study showed that referential phrase-frames dominated the corpus, followed by discourse phrase-frames and subsequently stance phrase-frames. Further findings revealed significant variations that emerged across journal categories. A Kruskal-Wallis H test and post-hoc analysis were conducted to analyze the differences within journal categories. According to the pairwise analysis, only one pair W-Y category showed differences in the distribution of REF_PFs, while the other two subcategories of phrase-frames (DISC_PFs and STNC_PFs) showed no variations. This study provides insights into academic writing, particularly in the context of research articles, and offers implications for EAP students and teachers. Additionally, it advances the study of phraseology by providing a deeper understanding of academic writing in arts and sciences.

Keywords: Phrase-frames, PFs, P-frames, Journal category, IMRD, Quantitative research, Experimental research articles, Correlational research articles, Academic writing, Corpus Linguistics, Phraseology

1. INTRODUCTION

1.1 Background

Research article (RA) is an academic genre, it includes passing on information within the academic community and presenting research findings. The academics fill in knowledge gaps through research publications and offer fresh perspectives and supporting data that the academic community may review and expand upon (Swales, 2004). Research articles have been referred to be an important way of legitimizing disciplines and results (Hyland, 2000). In the past many studies investigated RA sections, like the introduction section (Samraj, 2008; Swales, 2004) the methodology section (Lim, 2006; Musa et al., 2015), the results section (Bruce, 2009; Suherdi et al., 2020) and discussion section (Golparvar & Barabadi 2020; Kanoksilapatham, 2007). The style and organization of research publications are substantially influenced by the journal selection. This corpus-based study will investigate if the distribution of phrase-frames varies depending on the journals in which the papers are published by classifying them into three categories (W, X, and Y).

In earlier studies, formulaic language was studied subjectively, with academics compiling lists of fixed phrases that they believed were often used in the language (e.g. Pawley & Syder,

1983). Recently, there has occurred a significant change in the study of formulaic expressions as a result of technological developments in computers and their application to the corpus-based analysis of language. Now, a broad number of approaches on how to investigate recurring multi-word structures utilized in a wide range of text genres have been available due to the methodologies provided by corpus linguistics. Corpus linguistics plays an essential role in achieving the objectives of this study. The distribution of PFs in quantitative research articles is examined in the current corpus-based study, and the effects of journal category is investigated. Corpus linguistics enables quantitative analysis by utilizing statistical approaches. The research measures phrase-frame frequencies, discovers patterns, and compares its usage within journal categories (W, X, Y).

1.2 Statement of the Problem

The past researchers explored phrase-frames in various academic texts across different disciplines, and cultures, and few research scholars studied this discontinuous multi-word sequence in major sections (introduction, methodology, results, and discussion) of research articles. Most of the studies selected prestigious texts from diverse fields and disciplines for investigation. However, previous research studies on phrase-frames had some important gaps. These limitations include the omission of research article design considerations, a concentrated focus on one particular section of the research article when generating results, a lack of diversity in disciplinary representation, and a recurrent sample size limitation due to the inclusion of only one discipline in the study. By undertaking a corpus-based analysis, the current research study attempts to address this problem by examining how journal categories (W, X and Y) affects the use and frequency of this linguistic structure.

2. LITERATURE REVIEW

2.1 Phraseology

The study of fixed and frequent patterns of words, phrases, and idiomatic combinations within a language is known as phraseology. These linguistic constructions, referred to as "phrases," are not only produced on a purely grammatical basis, but frequently used together as a result of convention. The use of these recurring and predetermined word combinations and phrases is essential to academic discourse from the perspective of academic writing. These linguistic features provide assistance to maintain genre-specific rules, sustain cohesiveness, and effectively communicate complicated concepts. Formulaic sequences, particularly in academic contexts, are essential for enhancing textual coherence, understanding, and clarity (Wray, 2013).

Phraseology is an essential linguistic element that improves clarity in the context of academic writing. Academic speech represents expert communication, and phraseological study on fixed-word units has mostly concentrated on academic discourse (Fuster-Márquez, 2014). Phraseology in academic writing includes collocations, lexical bundles, word clusters, phrase-frames, and other formulaic language units that support the clear flow and efficiency of written communication in addition to constructing phrases. Language instructors may have been the first to deal with the real-world challenges of formulaic language (Carter & Sinclair, 2004; Herbst, 2011; Stubbs, 2009). The inclusion of assessments of formulaic sequences and their roles in certain approaches, registers, and genres of texts may enhance the English for Academic (EAP) writing curriculum and learning outcome (Coxhead & Byrd, 2007; Paltridge, 2004).

2.2 Formulaic Language

The term “formulaic” refers to lexical units in a language that correspond to common structures or formulae. It describes a collection of familiar and frequent word combinations and phrases that are used in particular contexts. These recurrent interrupted and uninterrupted multi-word sequences in past studies are frequently referred to as n-grams (Stubbs, 2007), formulaic frames (Biber, 2009), formulas (Simpson-Valch & Ellis, 2010), phrasal expressions (Martinez & Schmitt, 2012), lexical-frames (Gray & Biber, 2013) and phrase-frames (Fletcher, 2002a, 2007b). It is now possible for scholars to get updated and recent descriptive data with corpus linguistics that helps investigate phraseological variation or formulaic in text. These studies have mostly concentrated on the function of formulaic sequences in the linguistic production of academic genres, such as academic texts, menu scripts, academic discussions, and published research article part-genres.

There are several studies on discontinuous and continuous multi-word expressions in academic language. These studies have investigated multi-word sequences in a variety of contexts, including academic writings, medical fields (Grabowski, 2015), tourism language (Fuster-Márquez, 2014), engineering field (Nekrasova-Beker, 2019) environmental fields (Bararbadi et al., 2020), promotional writing (Casal & Kessler, 2020), British and American English (Liu, 2022), spoken and written language (Biber et al., 2004; Simpson-Vlach & Ellis, 2010). To better comprehend the language of the corpus, the consequent lists of phrases are then frequently categorized by structure and function. In this study, a discontinuous multi-word sequence (i.e. phrase-frame or p-frame) will be studied in quantitative research articles.

2.3 Phrase-Frames (PFs)

Phrase-frames belong to the linguistic category of phraseology. It is a recently developed theoretical idea created expressly to make it easier to describe phraseological structures in texts. Initially, PFs were described as collections of n-gram variations (of almost any size) that are identical but for one word, such as if you * any or for the purpose of * (Fletcher, 2002a, 2007b). The variable slot is indicated by *. So, a phrase frame "the * of this study" may, for instance, contain n-grams the section of this study, the variables of this study, the findings of this study, and the outcomes of this study. Filler words are those that appear in the empty area denoted by an asterisk (*). As a result, frames may alternatively be described as a frame with a group of fillers. The above-exemplified frame “the * of this study”, for instance, contains the fillers section, variables, findings, and outcomes. PFs might be used to compare patterns of variation between various text kinds as they shed light on how fixed or variable multi-word units are in a particular register (Römer, 2009). It is generally believed that a large number of PF variations refers to a greater degree of phraseological variations in the text.

Phrase-frames are classified both structurally and functionally based on their structures and purposes. Based on their structures, PFs can be divided into three types: verb-based PFs (V-based), PFs containing content words (C-based), and phrase-frames incorporating function words (F-based) (Biber et al., 2004). According to functional categories, phrase-frames are categorized into three categories. For instance, referential phrase-frames (a wide variety of *), discourse organizing phrase-frames (followed by a * of the), and phrase-frames that convey stance expressions (I find that the *). Using the same functional categorization of PFs, this study will examine the distribution of p-frames interdisciplinary across W, X, and Y journal categories.

2.4 Extraction of PFs in Past Researches

Several studies have utilized the *kfNgram* tool (created by Fletcher, 2002a, 2007b), to create a list of the most frequently occurring PFs in their corpora. The tool creates lists of n-grams and p-frames in text and HTML files. This corpus tool was used to extract a list PFs of different word-length according to the methodology of the researchers (Barabadi et al., 2020; Casal & Kessler, 2020; Cunningham, 2017; Fuster-Márquez, 2014; Grabowski, 2015; Garner, 2016; Golparvar & Barabadi, 2020; Lu et al., 2018; Nekrasova-Beker, 2019; Römer, 2010; Win & Masada, 2021). For learners and instructors of English for Academic Purposes (EAP), numerous corpus-based research studies advocated for the requirement of lists of academic formulaic expressions in recent years. It discovered the methodological challenges associated with creating such lists and introduced many such pedagogically useful lists of various types for educational phrases.

This study will use the methodology used by Cunningham (2017) in the investigation of PFs. All of the phrase-frames (PFs with variable slots on any position) irrespective of their word-length will be included in the analysis to explore highly variable patterns in quantitative research articles of arts and science. The rationale behind the increased length is provided by Casal and Kessler (2020), according to them the increased length makes it possible to locate PFs that are more linguistically complete as well as specific to academic writing, particularly research articles. In this study, PFs will be analyzed using an available list of functional p-frames established study (Lu et al., 2018). The analysis will involve investigating the distribution of PFs in RAs concerning journal category.

2.5 Exploration of Phrase-frames in the Past Studies

In previous research, the utilization of PFs in various types of academic texts, manuscripts, and research articles have been examined. They have examined the use and discourse functions of PFs in academic and other particular domains. For instance, English-language academic discourse (Forchini & Murphy, 2008), pharmaceutical discourse, particularly in the field of medical sciences (Grabowski, 2015), research articles in different science and arts disciplines (Barabadi et al., 2020; Cunningham, 2017; Golparvar & Barabadi, 2020; He et al., 2021; Lu et al., 2018). Some studies examined PFs in L2 learners' writings and EFL writing patterns, keeping in view the phraseological patterns (Casal & Kessler, 2020; Garner, 2016; Juknevičienė & Grabowski 2018). Some studies also explored PFs in the language of hotel websites (Fuster-Márquez, 2014; Fuster-Márquez & Pennock-Speck, 2015). Moreover, one of the studies used PFs as an exploratory tool to investigate the cross-linguistic translation patterns in Polish and international communities (Grabowski, 2020).

Research articles have been referred to as an important way to legitimize disciplines and results (Hyland, 2000). The majority of earlier studies that looked at PFs in research articles across numerous disciplines such as Applied linguistics, international business management, environment, medical fields, mathematics, higher education, social sciences, and engineering education (Ang & Tan, 2019; Barabadi et al., 2020; Cao & Wu, 2022; Grabowski, 2015; He et al., 2021; Ishii & Kawamoto, 2022; Nekrasova-Beker, 2019; Win & Masada, 2020; Yoon & Casal, 2020) etc.

Yoon and Casal (2020) explored how formulaic phrase frames are used in the rhetorical construction of Applied Linguistics Conference Abstracts. The results showed that p-frames

were frequently used by conference abstract writers in their rhetorical techniques, both in terms of frequency (approximately 1.7 times per text in the data) and their strong association with certain rhetorical functions. The researchers concluded that the use of p-frames in the abstracts emphasizes the significance of these linguistic tools for enhancing coherence and cohesiveness in academic discourse.

Barabadi et al. (2020) examined publications in the field of environment to draw up a list of key PFs. They collected a corpus of 125 research articles published between 2014-2018 from five highly cited and notable journals. The results of the study showed that content-based PFs predominated in terms of structural types of PFs and verb-based phraseological items were the second most common type. The author of the study advised that L2 learners should build up a strong variety of language forms, including key p-frames, which are regarded as the foundation of academic discourse.

Lu et al. (2018) conducted a corpus-based study to explore particular academic expressions for a specific part of research articles i.e. introduction section of social sciences research articles. The study was aimed to provide a pedagogically valuable list of p-frames and this was done to support previous corpus-based initiatives to compile lists of academic terms.

Golparvar and Barabadi (2020) investigated key phrase frames throughout the discussion portions of research publications in the field of higher education. The results of the research found that non-verbal content word phrases and referential PFs were often used, demonstrating their essential function in communicating specific discourse roles in the discussion sections. The study additionally provided valuable perspectives on the pedagogical implications for improving discipline-specific writing.

Cunningham (2017) examined key PFs in contemporary mathematical research publications. In contrast to earlier researchers who focused primarily on PFs with empty slots only in the middle, Cunningham adopted the original frame definition, permitting variability in all three positions: the beginning, middle, and end. Moreover, the researcher noticed that constructing p-frames from common n-grams failed to account for highly variable p-frames in the analysis. Instead, author recommended that when studying highly variable patterns, all n-grams found in a corpus should be considered during the identification phase. This approach has been used by numerous researchers, as will be discussed below, and is also employed in the current study.

Win and Masada (2015) investigated technical PFs from graduate-level article titles and presented a technique for scholars to study the technical frames using word n-gram extraction. They focused on text- and genre-specific notions of phrases and recurrent wordings. They came to the conclusion that while the top-ranked trigrams obtained by unbalanced PageRank just have an independent meaning, the ones obtained by the method they used are technical PFs, i.e., a word arrangement that makes an entire technical phrase by simply inserting a technical word or words before or after it.

Using a key phrase-frame approach Ishii and Kawamoto (2022) examined the linguistic patterns at the introduction of specific moves in the results sections of experimental medical research papers. This analysis revealed recurring patterns that cause specific movements. The lexical priming theory was supported by this corpus-based move analysis, which also helped to

discover key linguistic markers for the development of coherent discourse, enhancing both research procedures and writing instruction in the field of experimental medical research.

Cao and Wu (2022) carried out an insightful investigation on the language of evaluation in doctoral thesis conclusions and focused on evaluative PFs and phraseological patterns. The PFs were later analyzed based on Appraisal categories evaluated elements and matched into various grammatical patterns. The results showed specific choices for particular appraisal functions and evaluation kinds across various p-frames and patterns. The study shows that evaluative expressions are used in academic writing and the language of assessment in Ph.D. thesis conclusions.

The studies mentioned above have explored PFs in research publications across a range of disciplines, including applied linguistics, international business management, environmental science, higher education, mathematics fields, social science, and medical science. These studies have not only looked at the entirety of research articles but have also specifically examined certain sections, often aligning with the common IMRD (Introduction, Methodology, Results, Discussion) structure (Cao & Wu, 2022; Golparvar & Bararbadi, 2020; Ishii & Kawamoto, 2022; Lu et al., 2018; Win & Masada, 2015; Yoon & Casal, 2020) focused on PFs in the titles, abstracts, introductions, results, discussion and conclusions sections of research articles respectively.

2.6 Research Gaps in Past Studies

The past researchers have explored PFs in various academic texts across different disciplines, cultures, and different major sections (introduction, methodology, results and discussion) of research articles. Most of the studies have selected prestigious texts for investigation. However, they have not defined the specific research article types and the category of journals used in their corpora. Some of the studies worked on a very small sample and their data was not balanced. Then this study will investigate the distribution of PFs in RAs with reference to journal category. The investigation will mainly focus on phrase-frames in research articles from about eleven disciplines of arts and science. These quantitative research articles will be of two distinct research designs (correlational and experimental) and published in three journal categories (W, X, and Y as categorized by the Higher Education Commission of Pakistan). The sample size will be approximately 600 research articles. For analysis, an available list of functional PFs proposed by Lu et al. (2018) will be taken into consideration.

2.7 Research Questions

To fulfill the research objectives, the following research questions will be addressed in this study:

1. Which of the functional phrase-frame subcategory is dominated in the quantitative research articles?
2. To what extent does the distribution of phrase-frames vary across different research journal categories in quantitative research articles?

2.8 Significance of the Study

This study is distinguished from past corpus-based studies as it will represent two distinct quantitative research designs. Thus, this research will examine the effect of journal category (W, X, and Y, as determined by HEC's HJRS) on the distribution of p-frames in quantitative research

articles. Furthermore, this study holds significant implications for several domains of study as well. In the first place, it contributes to the linguistic field by offering a thorough examination of phrase-frame use in academic papers. Second, the results could make it easier to comprehend the features of disciplinary discourses and their most popular linguistic methods, enhancing interdisciplinary interaction and teaching of languages. Thirdly, by using phrase frames appropriately, the study can help researchers and writers improve the organization and readability of their research papers. Lastly, this study will also add useful insights into the field of research on phraseology and linguistic variations studies across disciplines and genres.

2.9 Null Hypotheses

The distribution of subcategories of phrase-frames (PFs or p-frames) is the same across journal categories (W, X, and Y).

3. METHODOLOGY

3.1 Research Design

The current corpus-based study adopts a quantitative approach to investigate quantitative research articles from different educational fields. Research studies from the arts and sciences fields are included in the present corpus-based investigation. The investigation involves one independent variables (journal category) and three dependent variables, for instance, three subcategories of phrase-frames (i.e., discourse phrase-frames, referential phrase-frames, and stance phrase-frames).

3.2 Data Collection and Procedure

3.2.1 Selection of Disciplines for Data Collection

The subjects chosen for this study were specifically limited to those that were academic and educational in nature. Following this criteria, the study included five distinct faculties within the arts and sciences, each comprised of a minimum of two subjects. The main disciplines examined were Social Sciences (Applied Linguistics, Psychology, Geography, Economics), Physical Sciences (Physics, Chemistry, Environmental Science), Biological Sciences (Biochemistry, Biotechnology, Biophysics), and Medical Science (Pathology, Pharmacology). Research articles were subsequently gathered from these disciplines.

3.2.2 Description of the Corpus

The corpus used in this study was especially targeted at academic research papers written in English across several different fields of arts and science. The present study collected a corpus of 660 research articles of two specific designs (i.e. correlational and experimental). The study focused only on those studies that were published in W, X, and Y category journals (HEC's HJRS). The corpus was compiled with the following criteria in mind. Only research articles with IMRD structure published during the previous 2.5 years (June 2019 to Dec 2022) have been included in the corpus to ensure their relevance. This guarantees that the conclusions of the research and analysis are supported by methods and data that are largely recent.

3.2.3 Validation of the Corpus

The data underwent a two-phase validation procedure after collecting all the quantitative research articles that met the above-discussed criteria. It was initially examined by PhD researchers, and then it was approved by faculty members regarded as experts in the relevant subjects. Their evaluation verified the IMRD structure of research articles and journal categories (W, X, and Y) in which these research papers had been published. Additionally, an essential phase in the validation process was the confirmation of the studies' designs. The evaluators

carefully reviewed the study designs (experimental and correlational) used in the downloaded papers.

3.2.4 Cleaning of the Corpus

Research articles were originally downloaded in PDF format from different online websites and virtual libraries, and then they were converted into Word format by using pdf2go.com. After that, each file was further edited by excluding all those elements that were not of interest to this study. The elements that were removed included headers and footers, graphs, pictures, page numbers, tables, headings and captions of the tables, visuals, formulas, models, titles, and captions of the graphs and pictures, models, references and appendices, etc. Both integral and non-integral citations, as well as numerical and parenthetical citations, were replaced uniformly with a single word, “Ref.” This modification was made in order to process the corpus in the software without any obstructs. This replacement was also necessary to lessen any possible increase in the frequency of nouns caused by varied citation styles across journals. Furthermore, the initials and abbreviation in scientific papers were replaced by their full forms.

The whole data-set was then uploaded onto the AntFile Converter software (version 2.0.2) and converted into Text files. This modification was required to process the files for analysis. The file names were carefully checked for correctness and the files with improper names were renamed. After that, the corpus was properly selected and prepared for in-depth analysis using Laurence Anthony’s AntConc software (3.5.9) by following this systematic process for organizing, transforming, and preparing files. This systematic methodology helped us effectively explore PFs inside the selected research papers and assured the accuracy of the data.

3.3 Data Analysis

The data consisted of one independent variables (i.e. journal category) and three dependent variables (three sub-types of phrase-frames i.e., REF_PFs, DISC_PFs, STNC_PFs). the independent variable *journal category* had three levels (W Category Journals, X Category Journals, Y category Journals). The first step in data analysis involved normalization of dependent variables (to 1000 words). Normalization is mandatory in corpus analysis to control the effect of text length in different files.

The data was analyzed using a Kruskal-Wallis H test for the independent variable and three dependent variables because the data did not meet the requirement for MANOVA nor did the data fulfill the requirements for ANOVA. However, before reaching this decision the possibility of conducting MANOVA and ANOVA was checked.

3.3.1 Checking the Assumptions of MANOVA

The data, on face value (because of having involved one independent variable (i.e. journal category) and three dependent variables (i.e. three subcategories of PFs) appeared to qualify for a MANOVA, however, after checking the assumptions of MANOVA it was discovered that it was not appropriate due to a large number of outliers (extreme values) in each dependent variable.

A large number of cases achieved zero scores in each variable (REF, DISC, and STNC), all of which were identified as outliers. In addition, normality tests for all variables showed a lack of normal distribution for all three dependent variables ($p > .05$). Due to the lack of normal distribution for all variables ($p < .05$), MANOVA was eliminated as an option for data analysis.

3.3.2 Checking the Assumptions of ANOVA

Since all three variables were subcategories of a single variable (phrase-frames), a composite variable (named “TTL_Normed1000”) was created by totalling the normed score of the three variables for exploring the possibility of applying ANOVA on the data. Outlier examination identified a large number of cases as outliers from the dependent variable.

On inspection of variables having zero value, it was discovered that 154 cases had a value of zero on the lowest end. These cases were removed. On the highest end, five cases had extreme values, which were also removed. Thus, in total 159 cases were removed. After removing these values, normality tests were conducted, the results of which reveals that the p-value is equal to .000. So, the tests of normality revealed that data was not normally distributed (p=.000).

3.3.3 Choice of Kruskal-Wallis H Test

Since the data failed to satisfy MANOVA and ANOVA assumptions, a Kruskal-Wallis H test was conducted to study the effect of the independent variable (journal category) on three dependent variables (subcategories of phrase-frames; DISC_PFs, REF_PFs, and STNC_PFs). As the independent variable involved more than 2 categories (W, X and Y), the differences were further investigated through post hoc test.

4. RESULTS AND DISCUSSION

4.1 Descriptive Statistics

Without undergoing detailed statistical analysis, descriptive statistics offers a brief overview of significant features, facilitating the discovery of trends, patterns, and findings. This descriptive statistics section offers frequencies of one independent and three dependent variables.

4.1.1 Frequency Distribution of Independent Variable: Research Journal Category

Frequencies of the independent variable, journal category (W, X, and Y) are given in this section. The following table 1 presents the frequency distribution of research articles by journal category (W, X, and Y category).

Table 1

Frequency Distribution of Research Articles by Journal Category

Categories of Journals		Freq.	Percent	Valid Percent	Cumulative Percent
Valid	W Category Journals	866	33.6	33.6	33.6
	X Category Journals	868	33.7	33.7	67.4
	Y Category Journals	840	32.6	32.6	100.0
	Total	2574	99.9	100.0	
Missing	System	2	.1		
Total		2576	100.0		

As Table 1 shows, the variable consists of three levels (W, X, and Y). Almost similar number of texts come from W category journals (866 texts, representing 33.6%), X category journals (868 texts, making up 33.7%), and Y category journals (840 texts, accounting for 32.6% of the total). Thus, the data concerning the independent variable is also balanced. This balance makes sure that no particular category is overrepresented, which would make it difficult to explore the relationship between the independent and dependent variables.

4.1.2 Frequency Distribution of Dependent Variables: Phrase-frames

This section offers descriptive results of the three dependent variables (three subcategories of PFs) preceded by the raw and normed frequencies of different subcategories of PFs for comparison with past studies. The following table 2 shows the raw and normed frequency distribution of subcategories of PFs in the corpus.

Table 2

Raw and Normed Frequency Distribution of PFs Sub-types in Corpus

Subcategories of Phrase-frames	Raw (out of 2967056 words)	Normed (million words)	Normed (thousand words)
DISC_PFs	3593	1210.67	1.21
REF_PFs	8430	2838.68	2.84
STNC_PFs	2717	915.36	0.92
The total frequency of PFs per million word		4964.71	

A total of 2,967,056 tokens have been found in the corpus. Notably, PFs appear with the highest frequency when normed frequencies are taken into account. The total raw frequency of all PFs found in the corpus amounts to 14,740 instances. The distribution of PFs by major function, as provided by Simpson-Valch and Ellis (2010), is also included in the table. According to these frequencies given in Table 2, referential phrase-frames (REF_PFs) are the most prevalent followed by discourse phrase-frames (DISC_PFs) and stance phrase-frames (STNC_PFs). The results find that REF_PFs make up the biggest portion of these phrase-frames, accounting for 57.19% of the total, followed by DISC_PFs at 24.37% and STNC_PFs at 18.43%.

The quantitative results of the study reveal that referential PFs constitute the largest majority in the corpus (57.19%). Results by Barabadi et al. (2020), Golparvar and Barabadi (2020), Garner (2016), Lu et al. (2018), and Walcott (2021) found that referential phraseological frames predominated over discourse PFs and stance PFs functional-based frames in respective corpora, provide support to the findings of this study. However, Cunningham's (2017) and Grabowski's (2015) studies, in which the majority of PFs found in his corpus of mathematics RAs and pharmaceuticals were discourse PFs, disprove the findings of this study. This discrepancy between the findings of Cunningham's (2017) and other previous studies from the current study, can be explained by the fact that mathematics and pharmaceuticals seem to be diverse even within the highly restricted genre of published research studies.

Several elements are consistent with the nature and goals of quantitative research that contribute to the frequency of referential PFs in quantitative research papers. According to Simpson-Valch and Ellis (2010), referential PFs express attributes (i.e. an overview of the *, a number of studies *), deictic and locative (i.e. at * points in time, over the past * decades), identification and focus (i.e. the body of research has *, focuses on the * of) and comparison and contrast (i.e. is related to the *, related to the literature on *) expressions. Since our corpus consists of research papers in a wide range of disciplines, it is very reasonable to predict that these expressions will constitute a vast majority of the corpus. Simpson-Vlach and Ellis (2010)

claim that this functional category is common in academic discourse owing to the essential role it plays in academic speech and writing.

The second dominant functional category under investigation in this study is discourse-organizing PFs (constituting 24.37%). The results of this categorization broadly agree with those of Lu et al.'s (2018) study, which revealed a 24% frequency rate. In contrast, it shows discrepancies with the studies by Barabadi et al. (2020) and Golparvar and Barabadi (2020), where the occurrence percentages were 0% and 10%, respectively. These p-frames frequently consist of predetermined word or phrase sequences that are used to accomplish a variety of communication goals. Lu et al. (2018) further categorized these PFs into discourse markers (i.e. in addition to the *, as we discuss in section *), textual and meta-discourse references (i.e. a brief * of the, in the following * we), topic introduction and focus (i.e. aim of * article is to, our * in this paper is) and topic elaboration (i.e. as a result of *, in order to * the). This subcategory consists of expressions that look into clarifying and expanding on a previously presented subject, either in a cause-and-effect relationship or a non-causal relationship.

The least prominent functional category investigated in this study is stance PFs (constituting 18.43%). The percentage of occurrence of this functional category is almost consistent with Lu et al.'s (2020) findings (i.e. 17.4%). This functional category focuses on the reader's attitude and point of view towards a notion or proposition of text. A stance expression (i.e. attributed to the * of) having an epistemic stance sub-function that deals with claims of knowledge and demonstration. While stance PFs are the second-biggest category in Golparvar and Barabadi's (2020) and Walcott's (2021) corpora, they are the smallest category in our analysis. The differences in the content of corpora that exist between our findings and those of Golparvar and Barabadi (2020) and Walcott (2021) can be related to this disparity.

According to the results of our analysis, referential PFs maintained the most dominant position, followed by discourse PFs, and subsequently by stance PFs. The same order of subcategories of PFs was observed in the studies conducted by Barabadi et al. (2020) and Lu et al. (2018). On the other hand, research by Golparvar and Barabadi (2020), Garner (2016) and Walcott (2021) showed deviation from this pattern. According to their results, a new hierarchy formed, with referential PFs assuming the first rank followed by stance PFs and then discourse PFs. The differences in the occurrence of the phrase-frame category highlight the variation throughout this research.

The following table 3 offers descriptive results of three dependent variables (subcategories of phrase-frames; DISC_PFs, REF_PFs, and STNC_PFs).

Table 3

Descriptive Statistics for Phrase-frames Subcategories

Descriptive Statistics		DISC_PFs Normed1000	REF_PFs Normed1000	STNC_PFs Normed1000
N	Valid	2576	2576	2576
	Missing	0	0	0
Mean		1.4249	3.0544	1.0295
Std. Error of Mean		.03944	.06538	.03321

Median	.9658	2.4510	.0000
Mode	.00	.00	.00
Std. Deviation	2.00194	3.31823	1.68548
Skewness	5.103	5.531	4.655
Std. Error of Skewness	.048	.048	.048
Kurtosis	58.688	75.078	47.129
Std. Error of Kurtosis	.096	.096	.096
Minimum	.00	.00	.00
Maximum	36.59	67.57	28.57
Sum	3670.64	7868.24	2651.97

The lack of resemblance between mean, mode, and median on one hand, and high kurtosis and skewness values on the other hand in all cases allude to issues of normal distribution, which is further strengthened through normality tests already reported in methodology section.

4.2 Results of Inferential Statistics

This section contains the findings corresponding to the hypothesis that was formulated for the present study, followed by an in-depth examination and explanation of these findings.

4.2.1 Effect of Journal Category on Frequency of PFs

The null hypothesis verified the effect of journal category (W, X, and Y) on the frequency of three subcategories of PFs (i.e. discourse phrase-frames, referential phrase-frames, and stance phrase-frames).

A Kruskal-Wallis H test was used to investigate the hypothesis, which referred to the difference in phrase-frame (PF) score utilization across different research articles within research journals of different categories (W, X, and Y). To examine possible differences in PFs scores among the articles falling under the defined research journal categories, the above statistical method was used. The following table 4 shows the mean ranks obtained by W, X, and Y category on three subcategories of PFs.

Table 4

Effect of Journal Category on the Frequency of Phrase-frames

Subcategories of PFs	Journal Category	N	Mean Rank
DISC_PFs_Normed1000	W Category Journals	866	1259.08
	X Category Journals	868	1295.13
	Y Category Journals	840	1308.92
	Total	2574	
REF_PFs_Normed1000	W Category Journals	866	1256.85
	X Category Journals	868	1263.61
	Y Category Journals	840	1343.79

	Total	2574	
STNC_PFs_Normed1000	W Category Journals	866	1288.58
	X Category Journals	868	1299.59
	Y Category Journals	840	1273.89
	Total	2574	

Research articles from the Y journal category demonstrate higher average ranks in the DISC_PFs and REF_PFs categories with scores of 1308.92 and 1343.79 respectively. Subsequently, the X category follows with mean ranks of 1295.13 for DISC_PFs and 1263.61 for REF_PFs, while the W category shows slightly lower average ranks (1259.08 for DISC_PFs and 1256.85 for REF_PFs). While in the case of STNC_PFs X category journals receive a higher mean rank (1299.59) followed by W category journals (1288.58) and Y category journals (1273.89). The following Table 5 shows the statistical differences in the distribution of three subcategories of PFs (i.e. DISC_PFs, REF_PFs, and STNC_PFs) across journal categories.

Table 5

Test Statistics^b: Effect of Journal Category on Phrase-frames

Test Statistics	DISC_PFs Normed1000	REF_PFs Normed1000	STNC_PFs Normed1000
Chi-Square	2.162	7.226	.590
Df	2	2	2
Asymp. Sig.	.339	.027	.745

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

According to the test results as shown in table 5, the effects of journal categories on various phrase-frame types differ. The low p-value of 0.027 specifically shows that there is a statistically significant difference between the journal categories shown in the REF_PFs. The remaining two categories, DISC_PFs, and STNC_PFs with higher p-values of 0.339 and 0.745 respectively, do not demonstrate statistically significant differences across the journal categories (W, X, and Y).

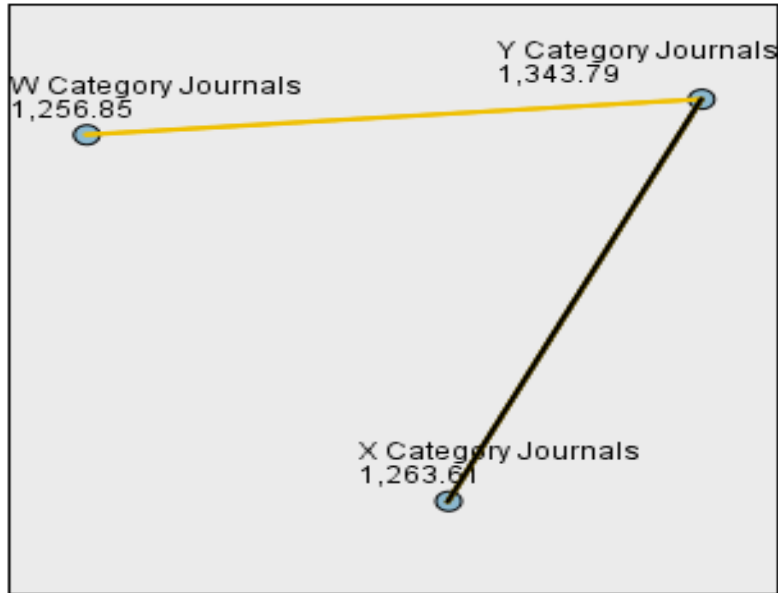
A Post hoc analysis was performed to determine whether types of journal categories had statistically significant differences in the frequency of REF_PFs. The results of the analysis are shown in the following section. It summarizes any notable differences found in the frequency of REF_PFs among the different types of journals.

4.2.2 Results of Post Hoc Test

The results of the post hoc analysis appear both in the figure and tabulated form below. Each node in the visual presents mean rank scores achieved by each category of journal (W, X, and Y) on referential PFs. Whereas, in the table, a detailed breakdown of pairwise results is given, highlighting statistically significant differences.

Figure 1

Pairwise Comparison of Journal Categories on the Frequency of REF_PFs



The pairwise analysis of journal categories is presented in the above Figure 1. The average rank scores for each journal category on referential PFs (REF_PFs) are expressed by each node in the picture. The graphic shows the average rank scores for the W category (1256.85, or 32.52%), the X category (1263.64, or 32.7%), and the Y category (1343.79, or 34.8%). The following table 6 shows the results of pairwise analysis.

Table 6

Pairwise Comparison of Journal Category on the Frequency of REF_PFs

Sample 1-Sample 2	Test Statistics	Std. Error	Std. Test Statistics	Sig.	Adj. Sig.
W Category Journals-X Category Journals	-6.766	35.600	-.190	.849	1.000
W Category Journals-Y Category Journals	-86.940	35.895	-2.422	.015	.046
X Category Journals-Y Category Journals	-80.174	35.875	-2.235	.025	.076

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

Table 6 provides an in-depth presentation of the pairwise results that reveal statistically significant differences. The table offers details about the test statistics, standard errors, standard test statistics, significance value, and adjusted significance values of three of the pairs mentioned in column one. The highlighted p-value shows that W-category journals and Y-category journals exhibit differences in the distribution of referential phrase-frames (REF_PFs).

According to the post-hoc analysis, the results showed that the distribution of linguistic features, particularly referential PFs varies across journal categories. So, the null hypothesis is

rejected. Academic journal articles in the modern day have demonstrated a more diversified and creative style that addresses the particular needs and criteria of each academic subject due to the increasing competence (Kuo, 1999). Furthermore, a journal article is frequently organized in terms of content structure from general to specific (organizing a researcher's study from a wider research context) and then specific to general (connecting the researcher's study outcomes to the wider study context).

Various academic institutions have classified journals from top-tier to low based on impact citations and impact ratio. The current study used the journal categorization suggested by the HEC's Journal Recognition System (HJRS) for the construction of the corpus. It divides journals into three distinct categories W, X, and Y (highly prestigious to least prestigious) within the relevant academic fields, based on a variety of worldwide validated indicators that assess the standards of a journal. In contrast to earlier studies (Barabadi et al., 2020; Cunningham, 2017; Golparvar & Barabadi, 2020; He et al., 2021; Lu et al., 2018; Win & Masada, 2015), that only mentioned the qualities or impact ratio of the journals incorporated in their corpora, this study collected corpus based on this categorization. Furthermore, previous studies did not analyze REF_PFs across journal categories. The findings of this study are significant in this respect.

According to the pairwise comparison given in the above figure, the findings reveal that referential PFs constitute significant differences in the W-Y category combination. While the W-X and X-Y category pairs do not exhibit any difference in the distribution of REF_PFs, as their p-values are greater than .05 (1.00 and 0.76 respectively) and their mean rank scores are in equal percentages (32.5% for W and 32.7% for X). The highest-ranking journals in HJRS are included in the W category followed by the X category and subsequently, the Y category (Rehman, 2021). W-category journals are regarded as impact factors and those of the highest prestige, and their approaches to writing are highly persuasive, pattern variable, and effective. On the contrary, the Y category, which constitutes the least prestigious journals with mediocre convincing styles, includes articles from national-level researchers (Bashira & Siddique, 2022). Consequently, after post-hoc analysis, the p-value of .046 on W-Y category journals confirms this obvious difference in impact factor between the W and Y categories.

5. SUMMARY

In this study, the distribution of phase-frames from a corpus of research articles in eleven distinct fields of arts and science was explored. Recent studies (Barabadi et al., 2020; Cunningham, 2017; Golparvar & Barabadi, 2020; He et al., 2021; Lu et al., 2018; Yoon & Casal, 2020) that examined the diversity in p-frame use across a range of disciplines, genres, and registers served as the inspiration for this research. According to the quantitative findings of the study, among three functional subcategories of PFs, referential phrase-frames (REF_PFs) made up the vast majority of the corpus (57.19%). Discourse phrase-frames (DISC_PFs) represented the second largest majority (24.37%) and stance phrase-frames (STNC_PFs) made up 18.43%, the results of this category were similar to the findings of Lu et al. (2018). This difference within journal categories was further identified through post-hoc analysis. Only referential phrase-frames showed significant differences across journal categories. The pairwise analysis across journal categories revealed that REF_PFs showed variation within the W-Y category pair.

5.1 Recommendations

This study examined the distribution of phrase-frames (PFs) in research articles. Since this study solely focused on quantitative research articles for corpus construction, it may be challenging to extend and generalize the findings from this study to other types of academic

publications. The corpus for this study was taken from eleven different disciplines which include social science, physical science, biological science, and medical science. This study, which focuses on four broad disciplines, falls approximately in the middle segment of the range of discipline specialization. The distribution of PFs that are specifically valuable in particular fields will undoubtedly provide a fruitful path for future research, even though we did not focus on cross-disciplinary variation. Furthermore, this study incorporated two specific quantitative research designs (i.e., experimental and correlational) and examined phrase-frame use, however, the inclusion of other quantitative research designs would result in a wider range of findings and insights.

5.2 Pedagogical Implications

This study has educational ramifications for many educational fields, but it has particular relevance for ESP (English for specific purposes) and EAP (English for academic purposes) writings. It may be worthwhile for EAP learners and teachers to look into the educational values of the p-frames discovered in this study. The results of this study can be a helpful resource for students in EAP courses that teach academic writing using a specific genre approach as they analyze the language characteristics of RAs. The distribution of p-frames analyzed in this study can assist students in identifying trends and formulas that are frequently employed in research articles. When writing research articles, novice scholars in the arts and sciences fields might utilize this study as one of many helpful reference materials. The findings from this study paved the way for our ongoing investigation into how to align the distribution of p-frames in research publications with IMRD structures.

References

- Barabadi, E., Robotjazi, M. A., & Bayat, M. (2020). A phraseological examination of research articles in the field of environment using key phrase frame. *Eurasian Journal of Applied Linguistics*, 6(1), 81-100.
- Bashira, A., & Siddiqueb, Z. Assessing the research productivity of funded and non-funded sources: Implications of enabling environment on research quality and quantity in higher education.
- 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., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.
- Bruce, I. (2009). Results sections in sociology and organic chemistry articles: A genre analysis. *English for specific Purposes*, 28(2), 105-124.
- Cao, F., & Wu, S. M. (2022). Evaluative Phrase-Frames and Patterns in Doctoral Thesis Conclusions. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 12(1), 1-17.
- Carter, R., & Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Casal, J. E., & Kessler, M. (2020). Form and rhetorical function of phrase-frames in promotional writing: A corpus- and genre-based analysis. *System*, 95, 102370.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of second language writing*, 16(3), 129-147.

- Cunningham, K. J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, 25, 71-83.
- Fletcher, W. (2011). H. 2002–2007. *KfNgram*. Annapolis: USNA. [Online] Available from: <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>. [Accessed: 20th November 2011].
- Forchini, P., & Murphy, A. (2008). N-grams in comparable specialized corpora: Perspectives on phraseology, translation, and pedagogy. *International Journal of Corpus Linguistics*, 13(3), 351-367.
- Fuster-Márquez, M. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84-121.
- Fuster-Márquez, M., & Pennock-Speck, B. (2015). Target frames in British hotel websites. *International Journal of English Studies*, 15(1), 51-69.
- Garner, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, 2(1), 31-67.
- Golparvar, S. E., & Barabadi, E. (2020). Key phrase frames in the discussion section of research articles of higher education. *Lingua*, 236, 102804.
- Grabowski, Ł. (2015). Phrase frames in English pharmaceutical discourse: a corpus-driven study of intra disciplinary register variation. *Research in Language*, 13(3), 266-291.
- Grabowski, Ł. (2020). Phrase frames as an exploratory tool for studying English-to-Polish translation patterns: a descriptive corpus-based study. *Across Languages and Cultures*, 21(2), 217-240.
- Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International journal of corpus linguistics*, 18(1), 109-136.
- He, M., Ang, L. H., & Tan, K. H. (2021). A corpus-driven analysis of phrase frames in research articles on business management. *Southern African Linguistics and Applied Language Studies*, 39(2), 139-151.
- Herbst, T. (2011). Choosing sandy beaches—collocations, probabemes and the idiom principle. *The phraseological view of language. A tribute to John Sinclair*, 27-57.
- Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Longman.
- Ishii, T., & Kawamoto, T. (2022). Typical Patterns in the Initial Position of the Moves in the Results Section of Experimental Medical Research Articles: A Key Phrase-Frame Approach. Available at SSRN 4195433.
- Juknevičienė, R., & Grabowski, Ł. (2018). Comparing formulaicity of learner writing through phrase frames: A corpus-driven study of Lithuanian and Polish EFL student writing. *Research in language*, 16(3), 303-323.
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. *Discourse on the move: Using corpus analysis to describe discourse structure*, 73-119.
- Kuo, C. H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for specific purposes*, 18(2), 121-138.
- Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, 25(3), 282-309.
- Liu, C. Y., & Chen, H. J. H. (2022). A phraseological exploration of university lectures through phrase frames. *Journal of English for Academic Purposes*, 58, 101135.
- Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, 36, 76-85.

- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied linguistics*, 33(3), 299-320.
- Musa, N. F., Khamis, N., & Zanariah, J. (2015). The structure of method section in Engineering research articles. *Asian Social Science*, 11(17), 74.
- Nekrasova-Beker, T. M. (2019). Discipline-specific use of language patterns in engineering: A comparison of published pedagogical materials. *Journal of English for Academic Purposes*, 41, 100774.
- Paltridge, B. (2004). Academic writing. *Language teaching*, 37(2), 87-105.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and native like fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-230).
- Rehman, S. U. (2021). Role of Central Induction Policy (CIP) in getting General Surgery Residency in Punjab: Is it Working Fairly for the Residents?. *Archives of Surgical Research*, 2(1), 43-44.
- Römer, U., & Arbor, A. (2009). English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies*, 20(2), 89-100.
- Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for academic purposes*, 7(1), 55-67.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, 31(4), 487-512.
- Stubbs, M. (2007). Quantitative data on multi-word sequences in English: The case of the word world. *Text, discourse and corpora: Theory and analysis*, 163-189.
- Stubbs, M. (2009). Technology and phraseology. *Exploring the Lexis-Grammar Interface*. Amsterdam: John Benjamins, 15-31.
- Suherdi, D., Kurniawan, E., & Lubis, A. H. (2020). A genre analysis of research article's findings and discussion sections written by Indonesian undergraduate EFL students. *Indonesian Journal of Applied Linguistics*, 10(1), 59-72.
- Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.
- Walcott, K. (2021). Informing academic writing pedagogy through the study of phrase-frames. *Journal of Language Teaching and Research*, 12(1), 158-171.
- Win, Y., & Masada, T. (2015, March). Exploring technical phrase frames from research paper titles. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops* (pp. 558-563). IEEE.
- Wray, A. (2013). Formulaic language. *Language teaching*, 46(3), 316-334.
- Yoon, J., Casal, J. E., Römer, U., Cortes, V., & Friginal, E. (2020). P-frames and rhetorical moves in applied linguistics conference abstracts. *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise*, 95, 281-305.