Volume 12, Issue 2
Summer and Autumn, 2024
pp. 87-114

Journal of Applied Linguistics and Applied Literature: Dynamics and Advances

Structural and Functional Differences of Lexical Bundles Between Hard Science and Soft Science Researchers: A Comparative Corpus-Driven Study

Ali Alizadeh[1], Davud Kuhi[2*], and Yaser Hadidi[3]

[1] *Ph.D Candidate in ELT, Maragheh Branch, Islamic Azad University, Maragheh, Iran*
Email:alizadehali9228@gmail.com

[2] *Corresponding author: Associate Professor of TEFL, Department of English Language and Literature, Maragheh Branch, Islamic Azad University, Maragheh, Iran*
Email: davudkuhi@iau-maragheh.ac.ir

[3] Associate Professor of TEFL, *Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, University of Tabriz, Tabriz, Iran.*
Email: hadidiy@hotmail.com

Abstract

Lexical bundles, recurrent word combinations serving essential discourse functions, have gained prominence in the realm of academic writing. A novel perspective that bridges their functional significance and formulaic nature is promising for uncovering intricate features within these recurring language patterns. Drawing from the structural and functional taxonomy introduced by Biber et al. (1999) and Hyland (2008), this comprehensive study aims to find the existence of any significant difference in the frequency and use of lexical bundles from both structural and functional perspectives between research articles authored by individuals in soft and hard science disciplines. The corpus, rich in academic content, encompasses a total of 954,615 words, featuring 90 research articles in each sub-corpus. The investigation extends beyond mere structural classification to encompass functional analysis, unveiling insightful findings. The findings indicate while structural distinctions between authors in hard and soft sciences appear negligible, substantial variations emerge in the pragmatic deployment of lexical bundles. Authors in soft sciences exhibit a predilection for noun phrases combined with of-phrase fragments. In stark contrast, authors in hard sciences predominantly employ passive verb + prepositional phrase fragments. Additionally, the divergence in the functional classification of lexical bundles is noteworthy. In the realm of soft sciences, authors heavily emphasize the use of framing signals, underscoring the discursive significance of these elements. In contrast, hard science authors gravitate towards transition signals as the most frequently employed function of lexical bundles. These findings carry substantial implications for researchers, highlighting the importance of embracing lexical bundles as a fundamental aspect of scholarly writing within their specific domains.

Keywords: Disciplinary Variation, Lexical Bundles, Hard Science, Research Articles, Soft Science

ARTICLE INFO

Research Article
Received: Saturday, November 11, 2023
Accepted: Tursday, June 27, 2024
Published: Tuesday, October 1, 2024
Available Online: Thursday, June 27, 2024
DOI: https://doi.org/10.22049/jalda.2024.29174.1630

Online ISSN: 2821-0204; Print ISSN: 28208986

Introduction

The functional significance of recurrent lexical combinations in both spoken and written academic discourses has garnered increasing attention (Simpson-Vlach & Ellis, 2010). These frequently occurring multi-word sequences constitute a vital component of fluent linguistic production (Akbulut, 2020) and play pivotal roles in the realm of academic writing. Moreover, as advocated by Hyland (2008 a or b?), a crucial aspect of mastering a new language is the adept utilization of specific lexical structures, including lexical bundles (LBs). These bundles, pervasive in language, pose a complex challenge for non-native and novice writers in the quest for precise and effective communication, particularly when considering disciplinary variations (Hyland, 2008a).

The adoption of lexical bundles within academic writing exhibits pronounced disciplinary variations, necessitating the compilation of discipline-specific bundles to acquaint students with the idiosyncrasies of specialized discourse communities (Cunningham, 2017). Previous studies examining lexical bundles in History and Biology, Arts, Law and Science, Linguistics, and Education have significantly advanced our comprehension of disciplinary disparities in bundle usage (Byrd & Coxhead, 2010; Candarli & Jones, 2019; Cunningham, 2017). However, a dearth of research delves into the distinctions between hard and soft sciences (e.g., Yin & Li, 2021), and the exploration of lexical bundles in research articles across diverse disciplines remains understudied. Hence, we propose that a comparative analysis of lexical bundles employed in research articles within hard and soft sciences holds immense value for L2 novice academic writers and their instructors, particularly within the vast domain of English for specific purposes.

Recent years have witnessed a burgeoning interest in lexical bundles as multi-word sequences. Research has underscored the significance of specific sets of bundles within academic genres, emphasizing their critical role in achieving academic fluency (Pérez-Llantada, 2014). While existing studies have shed light on the structural and functional aspects of bundles across various registers and academic fields (e.g., Cunningham, 2017; Durant, 2017; Le & Harrington, 2015; Ren, 2021), the influence of disciplinary variations on lexical bundles remains a less explored territory (Nuttall, 2021). These previous studies may inadvertently overlook distinctions owing to the characteristics of distinct groups, such as soft science versus hard science, and the potential confounding effects of register differences—a matter that recent research has vehemently debated. These differences are anticipated to wield substantial influence over the selection and deployment of lexical bundles. Furthermore, the inherent connection between lexical bundles and disciplinary variation suggests that combining functional and structural analyses across diverse fields could unveil overlooked properties of these linguistic constructs.

In the realm of English for specific purposes (ESP), which covers both soft and hard sciences, numerous second language (L2) novice academic writers, especially graduate students, face the challenge of mastering the language conventions within these specialized academic communities. At the same time, many instructors,

often with expertise in linguistics, play an active role in developing and teaching L2 academic writing courses. As a result, these instructors are well-equipped to illuminate the disciplinary differences in lexical bundles to L2 novice academic writers, particularly within the broad domain of English for specific purposes. Given the limited research in this area and the pivotal role of LBs in understanding the nuances of disciplinary variation, this study endeavors to scrutinize on the disciplinary differences and comprehensively investigate, analyze, compare, and contrast the frequency, structure, and utilization of lexical bundles among soft and hard science researchers in published scientific research articles. Consequently, this study aspires to unravel whether authors in soft and hard sciences employ lexical bundles differently in the creation of academic discourse and the conveyance of their scholarly insights.

Literature Review

Achieving idiomaticity, the mastery of conventionalized word combinations like lexical bundles, in academic discourse is a formidable challenge (Shin, 2018). Idiomaticity is not merely an abstract concept; it is tangibly realized through the effective use of recurrent word combinations that are characteristic of specific academic registers and disciplines. Such recurrent word combinations not only contribute to idiomaticity but also serve as a marker of membership within a distinct discourse community, reflecting the values, styles, and affiliations of its members (Wray, 2006). These specific word combinations, which fulfill particular functions and are summoned nearly instinctively by native speakers, are collectively referred to as formulaic language (Schmitt & Carter, 2004). Research in second language acquisition (SLA) indicates that native speakers rely more on formulaic language, especially lexical bundles, compared to non-native users. Furthermore, proficiency levels correlate significantly with the proportion and types of formulaic language employed, underscoring the importance of lexical bundles for academic writers (Akbulut, 2020).

The investigation of high-frequency word combinations known as lexical bundles has shown remarkable promise (Biber et al., 1999). Biber and Barbieri (2007) characterized lexical bundles as "recurrent word sequences." It's vital to emphasize that lexical bundles extend beyond mere sequences of individual words; they serve pragmatic functions in discourse and address recurrent communicative needs (Hyland, 2012). The pedagogical significance of lexical bundles, as commonly used word combinations in academic writing, has been gaining prominence in the field of language learning. Recognizing and distinguishing these word sequences is pivotal to grasping language as a cohesive whole. Additionally, studies indicate that holistic knowledge of word sequences enhances language processing in learners (Siyanova-Chanturia et al., 2011). Moreover, research on phrasal or formulaic language in various English teaching materials underscores the importance of lexical bundles (Grabowski, 2015; Meunier & Gouverneur, 2007). Given the significance and prevalence of lexical bundles, a wealth of studies has been conducted. For example, Chen and Baker (2010) outlined an approach for identifying lexical bundles in academic writings, comparing published academic texts with those of students (both

L1 and L2). The findings revealed less frequent use of lexical bundles in second language students' academic writing compared to their L1 counterparts. Additionally, certain high-frequency words were used less frequently in student academic writing than in published academic texts. The distinction in the use of lexical bundles between native and non-native speakers has been a focal point in other studies, such as Shin's (2018) examination of the structural and functional classifications of lexical bundles in university students' articles. The research revealed that both groups employed lexical bundles in various contexts, including VP-based bundles, stance-expression bundles, idiomatic PP bundles, and informal quantifying bundles.

Disciplinary variations in the use of lexical bundles within academic writing have also been explored in numerous studies (Candarli & Jones, 2019; Durrant, 2017; Lake & Cortes, 2020; Le & Harrington, 2015; Pérez-Llantada, 2014). Hyland (2008b) identified the variance in bundle utilization across different academic disciplines. For instance, bundles like "in the context of" and "it is important to" frequently appear in social science disciplines, while "is shown in figure" and "the presence of the" are likely to characterize writing in the hard sciences. Hyland attributed these differences to argument patterns, with social science writing predominantly connecting aspects of argument, while hard science writing tends to avoid authorial presence, interweaving arguments through data presented in visual formats. Consequently, writers in different academic contexts draw upon distinctive and discipline-specific reservoirs of lexical bundles.

The variation between academic disciplines in terms of writing conventions and language use has been a topic of considerable interest in research. Scholars have extensively explored and emphasized the significant differences in writing practices across different disciplines (Wright, 2019). In the field of English for Academic Purposes (EAP), educators have recognized the challenges of meeting students' academic writing needs, especially when closely related disciplines can employ strikingly different text types. This is particularly relevant as students are increasingly expected to engage with modules from various disciplinary areas (De Chazal et al., 2013).

Given the diverse nature of disciplinary variations, it is crucial for EAP practitioners and researchers to develop a strong understanding of these differences. Corpus linguistics, which involves the systematic analysis of large bodies of text, provides powerful tools for gaining insights into the distinctive characteristics of language use in different academic disciplines. One area of linguistic analysis that has shown promise in this regard is the study of high-frequency word combinations, often referred to as lexical bundles. The concept of lexical bundles, as introduced by Biber and his colleagues (Biber et al., 1999), has become a valuable framework for investigating how specific word sequences are used in different academic contexts.

By examining the lexical bundles prevalent in various disciplines, researchers and educators can gain a more comprehensive understanding of the

language requirements and conventions specific to each academic field. This knowledge is essential for effectively preparing students to engage with and produce texts within their respective disciplines and for enhancing the teaching and learning of academic writing.

Lexical bundles indeed possess unique characteristics that set them apart from traditional linguistic units. They often bridge two or more phrases or clauses, making them a distinct feature of language use, especially in academic writing. These characteristics have made lexical bundles particularly sensitive to differences between text types, and researchers have leveraged these properties to study variations across different domains, including disciplinary variation (Durrant, 2017). The high frequency of lexical bundles in various corpora emphasizes their significance in understanding and analyzing written discourse.

Research on formulaic sequences, particularly in academic writing in English, has been extensive (Esfandiari & Barbary, 2023), but there is a need for extra knowledge about how these sequences have evolved over time. A study by Hyland and Jiang (2018) delved into changes in the use and frequency of formulaic sequences over the past five decades. They conducted this study using a corpus of 2.2 million words sourced from top research journals in four different disciplines. Their findings challenged the notion that formulaic sequences are static and unchanging markers of research writing. Instead, they demonstrated that these sequences adapt to new conditions and contexts, with particularly interesting changes occurring within specific academic disciplines. Furthermore, Omidian et al. (2018) conducted research that compared the use of formulaic sequences in hard science and soft science fields. This study highlighted the differing foci of formulaic sequences within these two broad categories. Writers in soft science fields tended to use formulaic sequences to describe intangible aspects of research processes or procedures. In contrast, their counterparts in hard sciences used these sequences to explain the physical attributes of research objects. These findings underscore the nuanced ways in which formulaic sequences are employed in different academic disciplines and provide insights into the evolution of academic writing practices over time.

A detailed analysis of four-word lexical bundles in quantitative and qualitative research articles on education was conducted by Candarli and Jones (2019). They identified substantial intra-disciplinary disparities in the frequency, grammatical composition, and discourse function of such bundles. For example, quantitative articles employed more verb phrase bundles to convey the writer's perspective compared to qualitative articles. More recently, Xiao et al. (2023) examined the distribution patterns of information content across moves and the variations across disciplines. Their results indicate that information content is distributed unevenly across moves in a research article abstract, with different entropy indices reflecting various linguistic properties. These findings underscore the complex interplay between moves, linguistic meanings, and disciplinary features. Despite previous research hinting at a potential association between language and research paradigms within a given discipline, it remains unclear whether such associations extend beyond two disciplines. Consequently, empirical studies are needed to unveil

the role of disciplinary variations in the structure and function of bundles. Thus, the present paper aims not only to compare and contrast the frequencies and structures of lexical bundles in soft and hard sciences but also to examine the functions of these bundles within these disciplines. While it is acknowledged that previous studies, such as Hyland's, have explored lexical bundles across disciplines, the present research seeks to make a distinctive contribution by specifically focusing on the intersection of the rigidity and adaptability of lexical bundles in the discourse of research articles within both hard and soft sciences. The structural and functional taxonomy introduced by Biber et al. (1999) and Hyland (2008) serves as the foundation for our investigation, allowing us to delve into the intricate features of lexical bundles in a comprehensive manner. Our study not only compares and contrasts the frequencies and structures of lexical bundles but also goes beyond structural classification to conduct a nuanced functional analysis.

The novelty of our approach lies in uncovering pragmatic distinctions in the deployment of lexical bundles between hard and soft sciences, shedding light on the subtle yet crucial differences in how authors from these domains use recurrent language patterns. By examining not only structural variations but also functional classifications, our study extends the existing literature and provides valuable insights into the distinctive preferences and practices within each discipline. Consequently, our research aims to bridge the gap between the formulaic nature and functional significance of lexical bundles, offering a nuanced understanding that goes beyond existing disciplinary boundaries.

We acknowledge the foundation laid by previous studies, and our work builds upon these foundations to contribute a fresh perspective that addresses specific disciplinary variations in the structure and function of lexical bundles. This nuanced examination is crucial for researchers and academics in understanding and adapting scholarly writing conventions within their respective domains.

We anticipate that this research will contribute both theoretically and practically to the understanding and application of lexical bundles in the enhancement of academic writing skills. This study will assist students in improving their writing fluency and precision in academic settings, as suggested by previous researchers (Moynie, 2018; Pang, 2010; Allen, 2010). Additionally, this research will benefit EFL instructors, students, and material authors by highlighting the significance of lexical bundles in teaching materials, curricula, and classroom instruction. Based on the objective of the study, the following research questions were formulated:

RQ1: Is there a significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from a structural perspective between research articles authored by individuals in soft and hard science disciplines?

RQ2: Is there a significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from a functional perspective between research articles authored by individuals in soft and hard science disciplines?

Method

Corpus

This study undertook an in-depth analysis of lexical bundles extracted from two distinct corpora: research articles in the fields of soft and hard sciences. The primary focus of this research was to elucidate the distribution of lexical bundles within each corpus, shedding light on the distinctive patterns characterizing these academic disciplines. The research articles selected for examination were drawn from six diverse academic fields, representing both soft and hard sciences. By the soft science, we mean the fields of psychology, sociology, and applied linguistics, while by the hard sciences, we mean the sciences of computer science, chemistry, and medicine. To ensure the highest quality and relevance, research articles were exclusively sourced from reputable ISI journals published by Sage, Elsevier, Taylor and Francis, Springer, and Wiley Online Library. These journals were chosen for their stringent editorial standards and their status as key contributors to academic discourse.

The selection of research articles was made through a non-random and purposive sampling method. Specifically, articles published between the years 2010 to 2020 in various journals across these disciplines were included in the study. A total of 180 research articles were meticulously examined, with 90 articles representing the soft sciences and an equivalent number for the hard sciences. In classifying disciplines into soft and hard sciences, we employed Becher's (1989) typology of disciplinary grouping, which considers the hard-soft dimension. This dimension gauges the extent to which a discipline adheres to a paradigmatic structure, logical frameworks, and common theoretical models or frameworks (Toulmin, 1972). Disciplines that exhibit well-structured, paradigmatic characteristics are often categorized as "hard" sciences. Conversely, disciplines lacking a universally agreed-upon theoretical foundation are designated as "soft" sciences. As such, the hard disciplines in our study included chemistry, computer science, and medicine, while the soft disciplines encompassed applied linguistics, sociology, and psychology. The selection of disciplines in our study was a deliberate choice aimed at capturing a broad spectrum of academic fields while ensuring a balanced representation of both soft and hard sciences. The inclusion of psychology, sociology, and linguistics as representatives of soft sciences, and computer science, chemistry, and medicine as representatives of hard sciences, was driven by the desire to encompass a diverse range of academic disciplines.

These disciplines were chosen based on their prevalence and significance within their respective domains, reflecting the multidisciplinary nature of academic research. By examining lexical bundles across fields with varying methodologies, theoretical frameworks, and writing conventions, we aimed to provide a

comprehensive analysis that goes beyond the confines of a single discipline. This approach allows us to draw meaningful comparisons and contrasts, elucidating the distinct patterns characterizing soft and hard sciences in the use of lexical bundles.

The data used in this study may be characterized as small and specialized. We justify this choice based on the rationale presented by several scholars, including Flowerdew and Forest (2009), who argue that a specialized corpus comprising texts of the same genre and discipline can yield valuable insights for analysis, irrespective of its size. A small corpus also facilitates certain analyses that necessitate manual coding of lexical bundles according to their functions, which would be unmanageable within a larger dataset (Flowerdew & Forest, 2009).

Materials and Instruments

This study used two data analysis frameworks for lexical bundles' structural and functional analyses. The calculation and analysis of lexical bundles in the study were carried out using two classification models: Biber et al.'s (1999) structural classification model and Hyland's (2008) functional classification model. Biber et al. (1999) are credited with developing the structural taxonomies for lexical bundles. They categorized lexical bundles into three major structural types:

1. Type 1 LBs, which encompass verb phrase fragments (e.g., "is assuming to be").
2. Type 2 LBs, which include dependent clause fragments in addition to simple verb phrase fragments (e.g., "what I require to").
3. Type 3 LBs, which incorporate noun phrase and prepositional fragments (e.g., "of the factors that").

These structural categories are considered to manifest differently depending on the register or context of language use. For instance, in spoken conversation, lexical bundles tend to be more clausal in structure (e.g., "it's going to be"), while in academic prose, they are predominantly phrasal in nature (e.g., "as a result of"). These structural classifications provide insights into how lexical bundles are constructed and used in different types of discourse.

In this study, the functions of lexical bundles were analyzed using Hyland's (2008a) model, which provides a classification system for categorizing lexical bundles based on their functions within academic discourse. This categorization system helps researchers analyze and understand the functions that lexical bundles serve within academic texts, allowing for a more nuanced examination of how language is used in different contexts and for different purposes. Hyland's model organizes these bundles into three main macro-functions: research, text, and participants (See Table 1).

Table 1

Hyland's Macro-Functions Model

Research-oriented bundles
Procedure: These bundles indicate the methodology or purpose of research. For example, "the role of the."
Quantification: These bundles describe the amount or number involved in research. For instance, "the magnitude of the."
Description: These bundles detail the qualities or properties of the material under study. An example is "the structure of the."
Topic: These bundles are related to the field of research itself, such as "the currency board system."
Text-oriented bundles
Transition signals: These bundles establish additive or contrastive links between elements, such as "on the other hand" or "as well as the."
Resultative signals: They mark inferential or causative relations between elements, like "the results of the."
Structuring signals: These are text-reflexive markers that help organize stretches of discourse or direct readers elsewhere in the text, such as "as shown in fig."
Framing signals: These bundles situate arguments by specifying limiting conditions. An example is "with respect to the."
Participant-oriented bundles
Stance features: These convey the writers' attitudes and evaluations, including attitude markers, epistemic-certain, epistemic-uncertain, and intention bundles. An example is "are likely to be."
Engagement features: These address readers directly, as in "as can be seen."

While the terminology may differ between these categories, there are often interlinks and overlaps among them. For instance, stance bundles largely overlap with participant-oriented bundles as they express attitudes. Many discourse organizers can be placed under text-oriented bundles or text connectors, and referential bundles may align with expressions referred to as content presentation bundles or research-oriented

bundles. These categorizations help researchers and educators understand how lexical bundles function in academic texts and how they contribute to discourse cohesion and structure.

Procedure

At the outset of this study, a total of 180 research articles were meticulously curated from reputable ISI databases, including Elsevier (Science Direct), Sage, and Cambridge publications. These articles were then extracted through a combination of non-random and purposive sampling, focusing on the distinction between hard and soft sciences. The selection process adhered to stringent criteria to ensure the highest quality and relevance. Specifically, the chosen articles were sourced from journals published between 2010 to 2020 and were indexed in established and reliable databases. Furthermore, to better reflect the genre aspects of the most recent research articles, only papers with Modified Impact Factors (MIFs) equal to or exceeding 0.5 were included in the corpus.

The categorization of disciplines into soft and hard sciences adhered to Becher's (1989) typology, which emphasizes the hard-soft dimension. In this model, disciplines are classified based on the extent to which they exhibit paradigmatic structures, logical frameworks, and common theoretical models or frameworks (Toulmin, 1972). Notably, the hard sciences, characterized by their rigorous and well-structured nature, included chemistry, computer science, and medicine. Conversely, the soft sciences, represented by applied linguistics, sociology, and psychology, are marked by a relative lack of universally agreed-upon theoretical foundations.

Once the 180 research articles from selected journals were identified, they were downloaded and converted into Word document files for further analysis. To ensure the accuracy and smooth processing of the data, the files were meticulously cleaned of any headers, footers, figures, images, titles, references, irregular capitalizations, and paragraph breaks. The analysis of lexical bundles was carried out using Biber et al.'s (1999) structural classification model and Hyland's (2008 a or b?) functional classification model. Each element was calculated per 10,000 words, providing a basis for comparison between different types of research articles. The decision to calculate lexical bundles per 10,000 words, as opposed to 1,000 words, was made with careful consideration of the nature and scale of academic writing. Academic articles, particularly research articles, often vary significantly in length, with some being substantially longer than others. Calculating lexical bundles per 10,000 words provides a more standardized and representative measure, allowing for a fair comparison across diverse types of research articles.

Using a higher denominator, such as 10,000 words, offers a more comprehensive and meaningful perspective on the frequency of lexical bundles within the context of the entire document. This approach helps mitigate the potential impact

of variations in article length, ensuring that our analysis is not skewed by the inherent differences in the sizes of the research articles from different disciplines.

Additionally, working with a larger unit of words aligns with common practices in corpus linguistics, where researchers often use per 10,000 words as a standard measure to account for the varying lengths of texts under investigation. This choice facilitates a more robust and reliable comparison between different types of research articles, contributing to the overall validity and applicability of our findings.

The study focused on the analysis of 3-to-5-word structures, aiming to encompass the full spectrum of formulaic language used by each group of writers and to ascertain whether lexical bundle length correlated with assessed proficiency. To identify 3-to-5-word structures lexical bundles, the researcher employed Anthony's (2012) concordance tool, which allows for the extraction and analysis of word sequences within the specified length range. The frequency and range of these bundle lengths were determined through the concordance output, providing insights into the most common recurring structures in the corpus.

This approach allowed the researchers to investigate the full spectrum of formulaic language employed by different groups of writers and assess whether there was a correlation between the length of lexical bundles and the writers' assessed proficiency. The utilization of Anthony's (2012) concordance tool facilitated a robust and data-driven exploration of lexical bundles, contributing to the reliability and validity of our findings.

Anthony (2012) was employed to identify the most common lexical bundles. Anthony, developed by Laurence Anthony, is a popular and user-friendly corpus analysis tool widely used in linguistics and language-related research. It allows researchers to analyze and explore patterns within large bodies of text, known as corpora. This freeware concordance program offers features like word lists, n-grams, collocates, and clusters (Anthony, 2012). It systematically scanned the corpus for multi-word bundles, excluding proper nouns such as institution names and mathematical variables or symbols detected as lexical bundles. This meticulous curation aimed to refine the list and ensure its relevance to the study's objectives. To bolster the reliability of the data analysis, a systematic process of data categorization was undertaken. Specifically, 10% of the data was rechecked and independently reanalyzed by a second researcher, a Ph.D. graduate in Teaching English as a Foreign Language (TEFL), with expertise in discourse analysis. This researcher was briefed about the study's objectives. A random subset of the data from the corpus was analyzed by the second researcher, and inter-rater reliability was assessed using Cohen's Kappa formula. The resulting inter-rater agreement was found to be highly reliable with a Kappa value of 0.929 ($p = 0.000$). Subsequently, the gathered data was subjected to Chi-square data analysis to investigate the presence of any significant differences in the use of lexical bundles between authors in soft and hard sciences.

Results

To find the existence of any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from the structural point of view between

research articles written by authors in soft and hard science disciplines, the frequencies and use of lexical bundles were gathered and reported. To this end, a corpus of 45 research articles of approximately 720237 words from recently published Institute for Scientific Education-indexed journals that were authored by the authors of soft science (297468 for applied linguistics; 207267 for sociology; 215502 for psychology) was compiled and analyzed structurally. Furthermore, a corpus of 45 research articles of approximately 354,896 words from recently published journals that were authored by hard science authors (162159 for computer; 141224 for medicine; 51513 for chemistry) was gathered and analyzed. Table 2 shows the structural categorizations of LBs in the research articles written by the researchers of soft and hard science.

Table 2

Structural Classification of Lexical Bundles in Soft and Hard Sciences

STRUCTURE	EXAMPLE	Soft science		Hard science	
		Frequency	%	Frequency	%
Noun phrase with *of*-phrase fragment	the end of the, the beginning of the, the base of the, the point of view of	143	14.3%	88	9.2%
Noun phrase with other post-modifier fragments	the way in which, the relationship between the, such a way as to	60	6.%	80	8.4%
Prepositional phrase with embedded *of*-phrase fragment	about the nature of, as a function of, as a result of the, from the point of view of	31	3.1%	98	10.2%
Other prepositional	as in the case, at the same time as,	40	4.%	55	5.7%

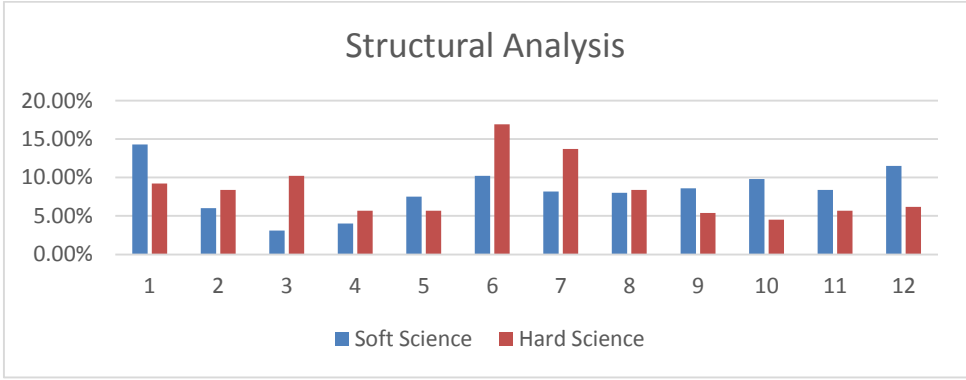
phrase (fragment)	in such a way as to				
Anticipatory *it* + verb phrase/ adjective phrase	it is possible to, it may be necessary to, it can be seen, it should be noted that, it is interesting to note that	74	7.5%	55	5.7%
Passive verb + prepositional phrase fragment	is shown in figure/fig., is based on the, is to be found in	100	10.2%	159	16.9%
Copula *be* + noun phrase/adjectiv e phrase	is one of the, may be due to, is one of the most	81	8.2%	129	13.7%
(Verb phrase +) *that*-clause fragment	has been shown that, that there is a, studies have shown that	80	8%	80	8.4%
(Verb/adjective +) *to*-clause fragment	are likely to be, has been shown to, to be able to	85	8.6%	52	5.4%
Adverbial clause fragment	as shown in figure/fig., as we have seen	98	9.8%	42	4.5%
Pronoun/noun phrase + *be* (+…)	this is not the, there was no significant, this did not mean that, this is not to say that	82	8.4%	55	5.7%
Other expressions	as well as the, may or may not, the presence or absence	116	11.5%	59	6.2%

Total	990	100%	952	100%

Based on what is presented in Table 2, it can be claimed that the authors of soft science used noun phrase with *of*-phrase fragment (e.g., *the beginning of the*) with high frequency and rate (14.3%) and the authors of hard science employed passive verb + prepositional phrase fragment (e.g., *is to be found in*) with high rate (16.9%), while the second place in the soft science devoted to other expressions (e.g., *may or may not*) with 11.5% of the total. The least frequent LBs in terms of structural classification was prepositional phrase with embedded *of*-phrase fragment (e.g., *as a result of the*) with 3.1%, and in hard science, the least frequent LBs from the structural point of view were adverbial clause fragment (e.g., *as shown in figure/fig*) with 4.5% of occurrences. Figure 1 visualizes the results.

Figure 1

Distribution of structural classification of LBs in soft and hard sciences



1) Noun phrase with of-phrase fragment 2) Noun phrase with other post-modifier fragments 3) Prepositional phrase with embedded of-phrase fragment 4) Other prepositional phrase (fragment) 5) Anticipatory it + verb phrase/adjective phrase 6) Passive verb + prepositional phrase fragment 7) Copula be + noun phrase/adjective phrase 8) (Verb phrase +) that-clause fragment 9) (Verb/adjective +) to-clause fragment 10) Adverbial clause fragment 11) Pronoun/noun phrase + be (+…) 12) Other expressions

In order to explore the existence of any significant differences between the authors of soft and hard sciences, a Chi-square test was used. The results of the Chi-

Square analysis, $x^2 = .021$, df = 1; $p \geq 0.05$, revealed that differences in relation to the frequencies of LBs' in terms of structural classification across disciplines were not statistically significant. In other words, although proportionately more LBs are used by the authors of soft science and although in some sub-categories, such authors used more structural sub-categories in comparison to the writers in hard science or vice versa, actually the differences are not found to be statistically different. The second research question posed in the present study aimed to investigate the existence of any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from the functional point of view between research articles written by the authors of soft and hard science (See Table 3).

Table 3

Functional Classification of Lexical Bundles in Soft and Hard Sciences

Major Functions	**Sub-Categories**	Soft science		hard science	
		Frequency	Percentage	Frequency	Percentage
Research-oriented	Location-indicating time and place, e.g. *in the present study*.	106	21%	49	6.5%
	Procedure-indicating methodology or purpose of research, e.g. *the purpose of this*.	60	11.9%	64	8.5%
	Quantification - describing the amount or number, e.g. *is one of the*.	40	7.9%	56	7.4%
	Description-detailing qualities or	32	6.3%	52	6.8%

	properties of material, e.g. *in the control group*.				
	Topic- related to the field of research, e.g. *in the United States.*	35	6.9%	60	8%
Text-oriented	Transition signals- establishing additive or contrastive links between elements, e.g. *on the other hand, as well as the.*	32	6.3%	119	15.6%
	Resultative signals- mark inferential or causative relations between elements, e.g. *the results of the.*	24	4.8%	80	10.4%
	Structuring signals- text-reflexive markers which organize stretches of discourse or direct readers elsewhere in the text, e.g. *as shown in fig*.	46	9.2%	95	12.4%

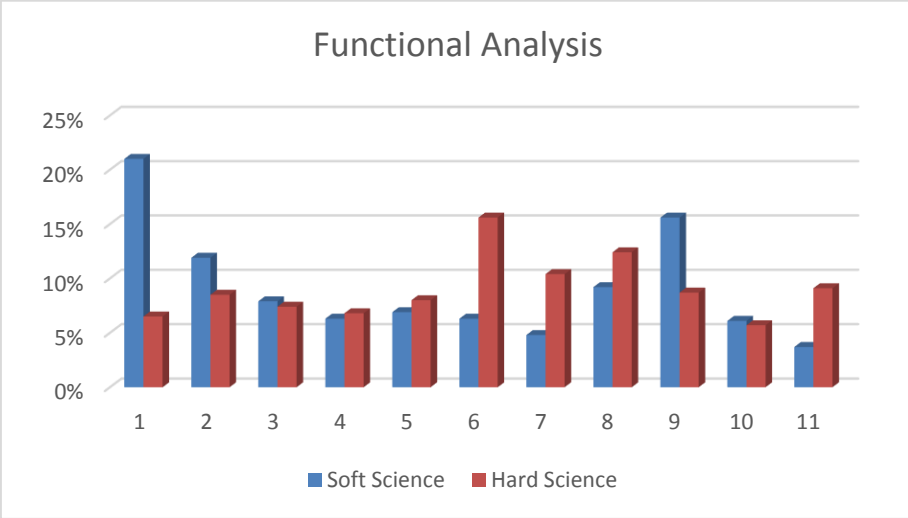
	Framing signals- situate arguments by specifying limiting conditions, e.g. *in the presence of.*	79	15.6%	67	8.7%
Participant -oriented:	Stance features- convey the writers' attitudes and evaluations. According to Cortes (2004), this category includes attitude markers, epistemic-certain, epistemic-uncertain and intention bundles, e.g. *were more likely to*.	31	6.1%	43	5.7%
	Engagement features- address readers directly, e.g. *it should be noted*.	19	3.7%	69	9.1%
Total		504	100%	754	100%

As shown in Table 3, the highest concentration of soft science authors is on framing signals (e.g., *in the presence of*) with 79 times of occurrences and distribution of 15.6%. The high frequent function of LBs used by the hard science is transition signals (e.g., *on the other hand*) with 119 times of occurrences and the same rate (15.6%). Engagement features (3.7%) and stance features (5.7%), the subcategories

of participant-oriented function are the lowest values that the researchers of soft and hard sciences used them in their academic writings. Figure 2 visualizes the results.

Figure 2

Distribution of functional classification of LBs in soft and hard sciences



1) Location 2) Procedure 3) Quantification 4) Description 5) Topic 6) Transition signals 7) Resultative signals 8) Structuring signals 9) Framing signals 10) Stance features 11) Engagement features

As it is crystal clear, the researchers of soft science used more location-related LBs (such as *in the present study*) in the articles and consequently both groups of authors used transition signals and framing signals at the same rate and distribution. To investigate the existence of meaningful differences between two groups of researchers in using functional aspects of LBs, a Chi square data analysis was run. According to the results, there is a significant difference between the authors in the functional classifications of LBs: $\chi^2 = 12.43$; df = 1; $p < 0.05$. Put in short, there is significance difference in terms of frequency of using lexical bundles in terms of functional classification in both soft and hard science corpora. Therefore, the researchers of hard science used more lexical bundles in terms of functional classification compared to those of soft science.

Discussion

The purpose of this study was to explore lexical bundles in soft and hard science research articles in terms of both structures and functions of LBs. The first research question was an attempt to find out the existence of any significant difference in the frequency and use of lexical bundles (3-, 4-, & 5-word) from structural point of view between research articles written by authors in soft and hard science disciplines. To this end, Biber et al.'s (1999) model for structural classification was used. The results indicated that there was not a significance difference in terms of frequency of lexical bundles in soft and hard science in accordance with structural classifications. Based on the information presented in Table 2, it can be observed that the authors of soft science and hard science articles utilize different linguistic structures with varying frequencies. In soft science, noun phrase with of-phrase fragments is used most frequently (14.3%), followed by other expressions (11.5%). On the other hand, in hard science, passive verb + prepositional phrase fragments are employed most frequently (16.9%), while adverbial clause fragments are the least frequent (4.5%). Although the results of Chi-square showed that there are no meaningful differences between hard and soft sciences in terms of LBs' structures, soft science authors tend to rely on noun phrases with of-phrase fragments, which may indicate a preference for descriptive and explanatory language.

According to Biber et al. (1999) and Hyland (2008) noun and prepositional phrases in academic writings shift the focus in the text from the writer to the action being done and the kind of relationship which exists between different elements of the text. Too many instances of noun phrases in this study could be due to the fact that academic writing is informational in nature and informational integration requires using noun phrases (Halliday & Hasan, 1989; Pan et al., 2016). Moreover, its extras use by the researchers of soft science could be because soft sciences often deal with complex social phenomena that require detailed descriptions and explanations.

Additionally, this analysis could provide insights into how language use has evolved within different fields of study. In general, it can be claimed that there are differences in linguistic structures used by authors in soft science versus hard science articles. The justification for the result can be the fact that the authors (no difference in being hard science researchers or soft science) are familiar with the principals of academic writing in general and structures of LBs in particular, hence the difference was not significant between the two groups of authors. This is a hunch and its validity can be measured via interviews with the authors. The relationship between disciplinarity and the use of lexical bundles in academic writing is an area of interest and research. Studies have shown that variations in lexical bundle usage are more closely related to disciplinary differences than to differences in text types. This suggests that the specific terminology, phrasing, and discourse patterns used in different academic fields are reflective of the discipline itself.

For example, studies like Cortes (2004) have highlighted substantial differences in lexical bundle usage between published writing and student writing in specific disciplines like biology and history. Similarly, Hyland (2008b) observed significant variations in lexical bundle usage between different natural sciences, such as engineering and biology, and social sciences, including business studies and applied linguistics. These variations are reflected in the frequency and types of lexical bundles used.

In hard sciences, authors often make extensive use of passive verb + prepositional phrase fragments, which might be indicative of a more objective and concise writing style common in these disciplines. The results of a Chi-Square analysis in your study indicate that there is no statistically significant difference in the frequencies of lexical bundles' structural classification across disciplines. This means that the proportion of lexical bundles used by authors in soft science and hard science disciplines, when viewed from a structural classification perspective, is not significantly different.

The findings are consistent with Abdollahpour and Gholami's (2018) study, which focused on frequently-used four-word general and technical lexical bundles in the abstract sections of research articles in medical sciences. Just like our study, they found that medical abstracts structurally contain a wider range of noun phrase bundles, such as "this study was designed," as compared to clausal phrases. In summary, the relationship between lexical bundle usage and disciplinarity is a significant aspect of academic writing, and these studies illustrate the discipline-specific characteristics of lexical bundles in academic texts. These differences can be attributed to the conventions, writing styles, and content of different academic fields.

Although the results of the current study are in harmony with Abdollahpour and Gholami's study, actually the sections that two studies selected are different. In this study, all of the sections of the research articles of two corpora were analyzed, while in the above-mentioned study, the focus was on the abstract sections of medical articles as a soft discipline. Their justification for the abundant use of phrasal structures in abstracts is that abstracts as an important piece of academic discourse are more compressed rather than elaborated, and therefore, this complexity leads to phrasal embedding than clausal one. Our justification would be that hard sciences often focus on empirical data and experimental results, where passive constructions can be used to emphasize objectivity and remove personal bias. It is important to note that these observations are based solely on the information provided in the current study. To fully understand and justify these results, further analysis would be required,

such as examining a larger sample size or conducting qualitative analysis of the articles themselves.

The second research question aimed to investigate the differences between two corpora from the functional classification outlook. So, Hyland's (2008 a or b?) model for functional classification was used. This investigation yielded a number of key findings in relation to lexical bundles in soft and hard science. Based on the information provided, it can be observed that soft science authors tend to use more location-related linguistic markers (LBs) in their articles compared to hard science authors. This suggests that soft science researchers may place a greater emphasis on providing specific details about the context and setting of their studies. It is interesting to note that our results contrast with Durrant's (2017) findings. Durrant found that writers in the science and technology fields, which are considered "hard sciences," used a significant number of lexical bundles for describing the physical form and presence of objects, particularly in the category of research-description bundles. This is in line with the conventions of hard science writing, where precision and detailed descriptions of physical objects are often essential for conveying research findings. In the current study, however, both soft science and hard science authors use transition signals and framing signals at an equal rate and distribution. This indicates that both groups recognize the importance of guiding readers through their arguments and presenting information in a structured manner. In light of the observed equal rate and distribution of transition signals and framing signals between soft science and hard science authors in the current study, there is a notable inter-disciplinary harmony.

When comparing these results with new articles, it is important to consider the specific research fields and methodologies involved. Different disciplines may have varying conventions and preferences when it comes to language usage. Additionally, changes in academic writing styles over time could also influence the patterns observed in newer articles. The results indicate that there is a significant difference between the authors in the functional classifications of lexical bundles (LBs). In simpler terms, the frequency of using lexical bundles differs significantly between authors when considering their functional classification in both soft and hard science corpora. Specifically, researchers in hard science tend to use more lexical bundles compared to those in soft science. There could be several reasons for this result. One possible explanation is that the nature of hard science research often requires more precise and technical language. As a result, researchers in hard science may rely more heavily on pre-established phrases and expressions (lexical bundles) to convey complex ideas efficiently. On the other hand, soft science research may involve more subjective or qualitative analysis, where there is greater flexibility in language use. This could lead to a lower frequency of using lexical bundles among authors in soft science. The difference in the frequency of using lexical bundles between authors in soft and hard science corpora can be attributed to several factors. One plausible explanation is that the nature of hard science research often demands

more precise and technical language. In hard sciences, research is typically characterized by a need for accuracy, specificity, and precision in conveying complex concepts and findings. The specialized terminology and intricate details inherent in hard science topics may lead authors to rely more heavily on standardized and formulaic language, resulting in a higher frequency of lexical bundles. These linguistic patterns help convey information efficiently and unambiguously, meeting the demands of the rigorous and exacting standards in hard science research communication. It is important to note that these findings are based on statistical analysis and should be interpreted within the context of the specific study. Further research may be needed to explore other factors that could contribute to these differences in the use of lexical bundles between soft and hard science researchers.

The observed variation in writing conventions and language use among academic disciplines has been a focal point of research interest (Wright, 2019). Scholars extensively explore and underscore the substantial differences in writing practices across diverse disciplines. In the realm of English for Academic Purposes (EAP), educators grapple with the challenges of addressing students' academic writing needs, particularly given the stark differences in text types employed by closely related disciplines. This becomes increasingly pertinent as students are now expected to navigate modules from a range of disciplinary areas (De Chazal et al., 2013). In the context of our findings, these challenges underscore the importance of understanding and addressing discipline-specific writing conventions for effective academic communication.

While previous studies, like Jalali et al. (2008), have explored lexical bundles in post-graduate writing, they have typically focused solely on lexical bundles in hard sciences. This research broadens the scope by encompassing both hard and soft sciences, offering a more holistic perspective on how lexical bundles are employed across different academic fields. The results of this study are consistent with the findings of Jalali et al., (2015), which, like the present study, focused on hard sciences. Both studies revealed that, although the most frequent lexical bundles in the corpus fell under the category of research-oriented bundles, medical research articles are distinguished by their substantial use of text-oriented bundles, particularly framing signals. Additionally, they showed a lower use of participant-oriented bundles in these articles. This pattern highlights the unique characteristics of lexical bundles and how their usage can vary within different academic fields and disciplines.

In conclusion, it is crucial for novice researchers, whether in soft or hard sciences, to receive training in effectively establishing the context and background of their research papers. This involves providing background information, referencing

prior research, and highlighting the central theme of the topic. Novice authors should also focus on identifying research gaps in their articles by referencing previous research limitations or identifying areas where further research is needed. Leveraging lexical bundles can be a valuable technique for researchers, as they simplify the process of constructing sentences by providing readily available parts. However, it is essential for researchers to be mindful of the specific uses of lexical bundles in their respective disciplines, as these bundles can be highly discipline-specific.

Conclusion

This study has undertaken an examination of lexical bundles and their comparative analysis in the contexts of two distinct disciplines: soft and hard sciences. While the structural classification of lexical bundles did not reveal a substantial difference between these two corpora, we unveiled a significant variance in the frequency of functional bundles employed across disciplinary boundaries. These findings hold significant implications for the teaching of English for Academic Purposes and teaching of English for Specific Purposes. This approach provided fresh insights into the nuanced landscape of lexical bundles across different scientific fields. Specifically, we highlight the presence of cross-disciplinary bundles, challenging the conventional notion that these linguistic patterns are strictly discipline-specific.

Practically, our findings can help EAP/ESP instructors in developing tailored pedagogical strategies. Recognizing the existence of cross-disciplinary bundles, we propose a hybrid approach. While discipline-specific teaching of lexical bundles remains crucial, we advocate the introduction of 3-5 word structures bundles that cut across various scientific domains to provide students with a broader linguistic toolkit. This approach aligns with data-driven learning and empowers students to explore the functions of these bundles within different disciplines, thereby enhancing their genre awareness.

The implications of this study reverberate across various domains. Researchers, teachers, and syllabus designers can benefit from the nuanced insights into the nature of lexical bundles generated by authors in hard and soft sciences. The nuanced insights into the nature of lexical bundles generated by authors in hard and soft sciences have practical implications for researchers, teachers, and syllabus designers in diverse fields. For researchers, understanding the distinctive patterns identified in this study can inform future investigations into the language features specific to their disciplines. Teachers can use these findings to tailor instructional approaches, focusing on the types of lexical bundles that are most prevalent in academic writing within their subject areas. Syllabus designers, too, can benefit by incorporating this knowledge into curriculum development, ensuring that language proficiency goals align with the actual lexical practices observed in research articles.

For instance, the identification of specific structures favored by authors in soft sciences, such as noun phrases combined with of-phrase fragments, provides concrete information that can guide language instruction in these fields. On the other hand, the prevalence of passive verb + prepositional phrase fragments among authors in hard sciences suggests a distinct pattern that can be integrated into teaching materials for those disciplines.

The findings offer fertile ground for further investigations into the practices of lexical bundles, not only within research articles but also across different genres. Non-native postgraduate students, in particular, can leverage these findings to improve their proficiency in deploying lexical bundles in their research proposals and theses. Beyond the confines of academia, the broader implications of this research stretch into the realm of generic awareness. Instructors involved in English for Academic Purposes, advanced writing, and seminar courses can use these findings to elevate their students' awareness of sound structures and the multifaceted functions of lexical bundles in their writings. Acknowledging the contributions of this study, it's important to recognize certain limitations. The study exclusively examined two types of disciplines and a specific set of academic fields, encompassing psychology, sociology, linguistics, computer science, chemistry, and medicine. A broader analysis involving a more extensive array of disciplines could provide a more comprehensive perspective.

Additionally, this study focused on written medium analysis, overlooking the potential insights that could be gained by exploring lexical bundles within the spoken medium. In future research, extending the scope to spoken discourse could yield valuable insights into how lexical bundles function in oral communication. Furthermore, we must acknowledge the impact of factors such as authors' gender and cultural differences on the use of rhetorical devices, including lexical bundles. Future studies could delve into these influences, offering a more holistic understanding of the dynamics at play. To compare and contrast these findings with new articles, it would be necessary to conduct a similar analysis on a separate dataset of new articles from both soft and hard sciences. This would allow for a direct comparison between the linguistic structures used in older articles (as represented in Table 2) and those found in more recent publications. By comparing the linguistic patterns between older and newer articles, it may be possible to identify any changes or trends over time within each scientific discipline. In conclusion, the evolving landscape of academic publishing, marked by a quest for conciseness within strict word limits, has reshaped scholarly writing, favoring compressed phrasal bundles overelaborated clausal structures. These findings not only provide a valuable contribution to the scholarly community but also assist scholars across different scientific disciplines in navigating

the nuances of written discourse, especially within the evolving contours of academic writing across disciplines.

Acknowledgments

The authors appreciate the contributions of professor Javad Gholami for assisting the authors in confirming rater reliability.

References

Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: Frequency, functions and structures of lexical bundles. *Asian ESP Journal*, *14*(1), 82-110.

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.

Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, *1*, 105-127.

Anthony, L. (2012). AntConc (Windows, Macintosh OS X, and Linux). *Recuperado de: http://www. antlab. sci. waseda. ac. jp/software/README_AntConc3*, *2*.

Becher, T. (1989). Disciplinary discourse. *Studies in Higher Education*, *12*(3), 261-274.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes (New York, N.Y.)*, *26*(3), 263–286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Johnsson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*(5), 31-64.

Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora*, *14*(2), 237-263.

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

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23(4)*, 397-423. https://doi.org/10.1016/j.esp.2003.12.001

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. http://dx.doi.org/10.1016/j.jeap.2016.11.005

De Chazal, E., Rogers, L., & McCarter, S. (2013). *Oxford EAP: A course in English for academic purposes.* Oxford University Press.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193. https://doi.org/10.1093/applin/amv011

Flowerdew, J., & Forest, R. W. (2009). Schematic structure and lexicogrammatical realization in corpus- based genre analysis: The case of research in the PhD literature review. In M. Charles, D. Pecorari and S. Hunston (Eds.), *Academic Writing: At the Interface of Corpus and Discourse* (pp. 15-36). Continuum

Esfandiari, R., & Barbary, F. (2023). A corpus-driven, diachronic analysis of recurrent word combinations across academic disciplines. *Corpora*, *18*(3), 263-295.

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23–33. https://doi.org/10.1016/j.esp.2014.10.004

Halliday, M.A.K & Hasan, R. (1989). *Language, context and text: Aspects of language in a social semiotic perspective*. Oxford University Press.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18(1)*, 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*, 4–21.

https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, *32*, 150-169. https://doi.org/10.1017/S0267190512000037

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, *23*(4), 383-407. http://dx.doi.org/10.1075/ijcl.17080.hyl

Jalali, Z. S., Moini, M. R., & Arani, M. A. (2015). Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. *International Journal of Information Science and Management*, *13*(1), 51-69.

Lake, W. M., & Cortes, V. (2020). Lexical bundles as reflections of disciplinary norms in Spanish and English literary criticism, history, and psychology research. In U. Römer-Barron, V. Cortes and E. Friginal (Eds), *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise* (pp. 183-204). John Benjamins. https://doi.org/10.1075/scl.95.08lak

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the Discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, *39*, 45-61. https://doi.org/10.1016/j.esp.2015.03.003

Meunier, F., & Gouverneur, C. (2007). The treatment of phraseology in ELT textbooks. In E. H. Tenorio, L. Q. Rodríguez-Navarro and J. S. Rodopi, *Corpora in the foreign language classroom* (pp. 119-139). Brill.

Moynie, J. (2018). *Lexical Bundles within English for Academic Purposes written teaching materials: A Canadian context*. [Master's thesis, Carleton University]. https://doi.org/10.22215/etd/2018-13240

Nuttall, C. (2021). Profiling lexical frame use in NSF grant proposal abstracts. *Applied Corpus Linguistics*, *1*(3), Article 100009. https://doi.org/10.1016/j.acorp.2021.100009

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes 36*, 1–14. https://doi.org/10.1016/j.jeap.2018.08.002

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*, *21*, 60-71. http://dx.doi.org/10.1016/j.jeap.2015.11.003

Pang, W. (2010). Lexical bundles and the construction of an academic voice: a pedagogical perspective. *Asian EFL Journal*, *47*, 2-13.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Shared and distinct usage. *Journal of English for Academic Purposes, 14*, 84-94. http://dx.doi.org/10.1016/j.jeap.2014.01.002

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt, Formulaic Sequences: Acquisition, processing and use (pp. 1–22). John Benjamins. https://doi.org/10.1075/lllt.9.02sch

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

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531

Shin, Y. K. (2018). The Construction of English Lexical Bundles in Context by Native and Nonnative Freshman University Students. *English Teaching*, *73*(3), 115-139. https://doi.org/10.15858/ENGTEA.73.3.201809.115

Toulmin, S. (1972). Rationality and Scientific Discovery. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1972*, 387–406. https://doi.org/10.1086/psaprocbienmeetp.1972.3698983

Wray, A. (2006). Formulaic language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 590–597). Elsevier.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, *54(3)*, 1-14. https://doi.org/10.1016/J.ESP.2018.09.001

Yin, X., & Li, S. (2021). Lexical bundles as an intradisciplinary and interdisciplinary mark: A corpus-based study of research articles from business, biology, and applied linguistics. *Applied Corpus Linguistics*, *1*(1), 100-115. http://dx.doi.org/10.22363/2687-0088-27752

Xiao, W., Li, L., & Liu, J. (2023). To move or not to move: an entropy-based approach to the informativeness of research article abstracts across disciplines. *Journal of Quantitative Linguistics*, *30*(1), 1-26. http://dx.doi.org/10.1080/09296174.2022.2037275

Authors' Biographies

Ali Alizadeh received his B.A. and M.A. degrees in English Language Teaching from Farhangiyan University, Tabriz, Iran. In 2003, he joined to the Ministry of Education in Tabriz. He has been teaching in high schools of Tabriz for 20 years. His major area of research and interest are corpus-based studies and discourse analysis. He has had several publications in domestic and international journals.

Davud Kuhi, a full-time member of English Language Department in Islamic Azad University, is mainly interested in analyzing academic/scientific discourses. A distinctive aspect of his work is that the analyses penetrate into deeper sociocultural process behind the academic/scientific discourses.

Yaser Hadidi is Assistant Professor of English Studies in the Department of English, The University of Tabriz, Iran. His areas of Interest include Systemic Functional Linguistics, Discourse Analysis and Literary Linguistics of prosefiction, among others.