A CORPUS-BASED STUDY OF LEXICAL BUNDLES IN THE ABSTRACTS OF HIGH-IMPACT RESEARCH ARTICLES ACROSS DISCIPLINES

Umi Khoirunnisa[1]; *Eri Kurniawan[2]
[1-2]English Language and Literature Study Program, Universitas Pendidikan Indonesia, Bandung, West Java, Indonesia
*umi.khoirunnisa@upi.edu; *eri_kurniawan@upi.edu*
(*) Corresponding Author

Abstract: Lexical bundles are essential for creating coherent academic writing and for forming high-quality research articles. Researchers have previously examined lexical bundles in a number of academic fields and sections. However, there is a lack of investigation on cross-disciplinary comparisons in research article abstracts. Thus, this study seeks to examine the utilization of lexical bundles (LBs) in abstracts across disciplines. The analysis used a corpus-based study design to investigate the frequency, structural patterns, and function distributions of lexical bundles in abstracts across disciplines categorized as soft sciences (Linguistics, ELT, Psychology) and hard sciences (Electronic Engineering, Medicine, Biochemistry). To guarantee the representativeness of high-impact research articles, 180 abstracts of research articles, 30 from each discipline were chosen from high-citation Scopus-

indexed journals. A descriptive quantitative approach was employed within Biber's (2004) structural taxonomy and Hyland's (2008b) functional classification, along with other categories to provide comprehensive understanding of LBs manifestations. The results indicate significant disciplinary commonalities in LB usage albeit numerous noticeable variations. A bias for phrasal over clausal constructs was observed in the structural use of bundles in both the Hard and Soft Sciences, with noun phrase-based bundles being the most prevalent. In terms of functionality, research-oriented bundles predominated in both domains, with Soft Sciences preferring description bundles and Hard Sciences stressing topic bundles. This study sheds light on disciplinary conventions in abstract writing and emphasizes the significance of understanding the structural and functional differences in LBs in cross-disciplinary engagement and effective academic communication.

Keywords: *functions, high-impact, lexical bundles, research article abstract, structures*

INTRODUCTION

In the academic realm, research articles have become a significant work. As a written document which outlines the author's investigations, research articles present new knowledge through exploring theoretical and/or methodological concerns and typically compare the findings to those of others (Swales, 1990). Furthermore, research articles also facilitate scientists strengthen their credibility and convince others of their discoveries by taking into account the audience and social implications (Hyland, 1998). Ultimately, research articles are vital for expanding knowledge and fostering productive scholarly interactions. Thus, a research study should feature the best multidisciplinary content and be the most recent (Kurniawan et al., 2019).

Once the significance of research articles has been recognized, particular focus must be placed on the kind of publications that

significantly impact the academic community, i.e., high-impact research papers. One critical factor contributing to a study's high impact is the output and the citation frequency associated with the researchers' work (Hirsch, 2005). High citation counts indicate that a piece of article has significantly contributed to the field's academic community (Kurniawan et al., 2019). Additionally, journal indexing also takes into account other aspects. The journal indexation is also regarded as a portal to excellent research and publications (Kurniawan et al., 2019) since it is seen as a general indicator that the journal is of high quality. Moreover, the primary determinant of a study's high-impact status is its publishing in a high impact factor journal, followed by reader interest, which is affected by the study's topic and design (Verma & Yuvaraj, 2022).

As a first point of contact for readers with journal research articles, abstract is considered important for several reasons. First, abstracts are an essential component that provide the basis for assessing a paper's merit for more reading (Ghasempour & Farnia, 2017; Hyland, 2000). By providing a brief summary of a research's objectives, methodology, results, and discussion, the abstracts provide readers a general understanding of the nature and scope of the research piece; they highlight the topic and key conclusions of the work without providing a thorough, step-by-step breakdown (Lorés, 2004). By then, the abstracts seek to draw readers in and facilitate their rapid evaluation of the article's relevance to their own research (Fauzan et al., 2020; Kafes, 2012). Second, a global academic community has the ability to read online research articles due to the expanding use of online scholarly web indexes and the public access to abstracts (Tocalo, 2021). Given the significance of abstracts, now they have been incorporated into the introduction, methods, results, and discussion (IMRAD) structure, which is the main benchmark for research articles (Wu, 2011). As a result, analyzing abstracts becomes critical for understanding their role in academic writing as well as their contribution to excellent scientific communication.

Building on the fundamental role of abstracts, it is necessary to examine more closely the linguistic elements that support their

efficacy, such as lexical bundles, which are vital resources for creating trustworthy and persuasive scholarly discourse. Lexical bundles, known as multi-word expressions, were initially created by Biber et al. (1999). Lexical bundles are clusters of three or more words that statistically occur together in a register, or extended collocations (Biber et al., 1999). Some examples of lexical bundles include: "as a result of," "on the other hand," "in the case of the," "the context of the," and "it is likely to" (Cortes, 2004).

In academic writing, lexical bundles serve as discourse building elements (Biber et al., 1999). To identify them, frequency, fixedness, idiomaticity, and structural status are important characteristics that distinguish lexical bundles from others (Biber et al., 1999; Cortes, 2004). Frequency refers to the high occurrence rate of these expressions within a particular discourse, with LBs being more frequent than pure idioms, which tend to appear less often or not at all. Most bundles are formally regular and semantically transparent, which makes them the foundation of coherent conversation, in contrast to colloquial phrases. In other words, because they usually span structural units, their frequency—rather than their structure—is the only empirical basis for their identification (Hyland, 2008b).

This frequent occurrence is directly linked to fixedness, where LBs are defined by their stable word combinations that meet specific frequency criteria, regardless of alternative forms (Biber et al., 1999). However, LBs are not rigid; they exhibit a certain degree of flexibility, allowing them to adapt across different discourses. For example, some LBs, such as "I don't know what", are commonly found in spoken discourse, while others, like "as a result of", are more prevalent in written academic texts (Biber, 2009). On the other hand, the majority of LBs are non-idiomatic in terms of idiomaticity, which means that their meanings are obtained from their constituent parts rather than from a collective, figurative interpretation (Cortes, 2004). Moreover, the majority of LBs bridge two structural units rather than forming entire structures; they usually start at the boundary of a phrase or clause and continue into the next unit. Because of their structural adaptability, LBs

can be used in a variety of syntactic configurations, which enhances their usefulness in academic writing (Biber et al., 1999).

In order to better understand how LBs function in academic texts, several later research have used the frameworks for classifying LBs based on their structural and functional roles provided by Biber's (2009) work, which is widely regarded as basic in LB studies.

Table 1. Structural classification of LBs (Biber et al., 2004)

Structures (LBs incorporating ...)	Substructures	Example
Verb phrase (VP) fragments	(connector +) 1st/2nd person pronoun + VP fragment (connector +) 3rd person pronoun + VP fragment Discourse marker + VP fragment Verb phrase (with non-passive verb) Verb phrase with passive verb *yes-no* question fragment WH-question fragment	'well I don't know' 'it's going to be' 'I mean I don't' 'take a look at' 'can be used to' 'do you want to' 'what do you think'
Dependent clause (DC) fragments	1st/2nd person pronoun + dependent clause fragment WH-clause fragment If-clause fragment (verb/adjective +) to-clause fragment That-clause fragment	'you might want to' 'what I want to' 'if we look at' 'to be able to' 'that there is a'
Noun phrase (NP) and prepositional phrase (PP) fragments	(connector +) Noun phrase with of-phrase fragment Noun phrase with other post-modifier fragments Other noun phrase expressions Prepositional phrase expressions Comparative expressions	'the end of the' 'the way in which' 'a little bit more' 'of the things that' 'as well as the'

In terms of functionality, Biber et al. (2004) state that the three primary functional categories of lexical bundles are referential

expressions, discourse organizers, and stance expressions. Hyland (2008a) expanded on this paradigm by modifying the classification to better fit the needs of academic writing, namely research papers, master's theses, and doctoral dissertations. Since then, this revised taxonomy has gained widespread acceptance and been used in later research on scholarly speech. Table 2 provides an overview of these functional categories.

Table 2. Functional classification of LBs (Biber et al., 2004; Hyland, 2008b)

Functions	(Biber et al., 2004)	(Hyland, 2008b)
Concerning the research topic	**Referential expressions** Identification/focus 'one of the most' Imprecision 'or something like that' Specification of attributes 'a lot of the', 'the size of the' Time/place/text reference 'in the United States', 'at the same time', 'as shown in figure'	**Research-oriented** Location 'at the beginning of' Procedure 'the use of the' Quantification 'one of the most' Description 'the structure of the' Topic 'the currency board system'
Concerning the text organization	**Discourse organizers** Topic introduction/focus 'in this chapter we' Topic elaboration/clarification 'on the other hand'	**Text-oriented** Transition signal 'on the other hand' Resultative signal 'as a result of' Structuring signal 'in the next section' Framing signal 'in the case of'
Concerning the author/audience	**Stance expressions** Epistemic stance 'I don't know if' Attitudinal/modality stance 'can be used to'	**Participant-oriented** Stance features 'may be due to' Engagement features 'it should be noted that'

Diving deeply inside LBs, lexical bundles are crucial elements, particularly when the author is attempting to create coherence and successfully communicate information. These expressions, which assist readers recognize a specific context or variety, such academic or legal, are made up of phrases or word sequences that typically appear repeatedly in a text (Biber et al., 1999; Hyland, 2008b). Previous research has shown that lexical bundles serve as a pivotal role in scholarly communication, where these patterns suggest subject-matter competence in addition to influencing the structure of academic narratives (Hyland, 2008b).

Up to this point, LBs have been investigated from various perspectives. In order to find patterns of variance and similarity, LBs have been investigated both inside and across disciplines (e.g., Farhang-Ju et al., 2024; Kashiha, 2023; Samraj, 2024). The way that LBs satisfy the communication requirements of several sectors has been made clear by this cross-disciplinary approach. Moreover, numerous studies have examined LBs in particular research article sections, like introductions, discussions, and conclusions (e.g., Deng & Liu, 2023; Goodarzi et al., 2024; Kurniawan & Haerunisa, 2023), or more than one sections (Shahmoradi et al., 2021; Shahriari, 2017), demonstrating their tactical value in bolstering rhetorical devices and accomplishing communication objectives.

Additionally, from the existing LBS research, only a few attention to the abstract. For instance, Shahmoradi et al. (2021) examined LBs in the information technology and applied linguistics articles, focused on the abstract and conclusion sections, whereas Varghaei and Khodadadi (2022) compared LBs in medical abstracts from Iranian and international journals. Abdollahpour and Gholami (2018) examined the manifestation of LBs in medical sciences abstracts, whereas Qi and Pan (2020) investigated LB variance across movements in medical abstracts. These investigations demonstrate how discipline norms and abstract-specific limitations are reflected in LBs. Moreover, there is still a gap in research providing a comprehensive, cross-disciplinary analysis of LBs in abstracts, addressing their structure and function.

Following previous studies in disciplinary discourse (Biglan, 1973a; Becher & Trowler, 2001), this research adopts the distinction between hard and soft sciences. While the dichotomy is sometimes viewed as reductive, it remains a widely recognized framework for examining cross-disciplinary variation in academic writing. Accordingly, the present study focuses on how lexical bundles (LBs) are manifested within abstracts from both hard sciences (e.g., Medicine, Biochemistry, Electrical Engineering) and soft sciences (e.g., English Language Teaching, Psychology, Linguistics) structurally and functionally. These inquiries serve as a guide to address these research questions: (1) What are the most frequent LBs in the abstract sections of hard sciences and soft sciences? (2) Do structural and functional differences exist between lexical bundles in hard science and soft science research article abstracts? If so, how are they expressed? By addressing these research questions, the study seeks to learn more about how LBs are used in article abstracts and how they increase the chances of acceptance and visibility in the journals that are being pursued.

METHOD
Research Design

This study applied a corpus-based study design. Researchers are able to find hidden meanings in lexical elements, such as through the study of collocations, and detect recurring linguistic patterns in language usage by using quantifiable data from a corpus-based analysis to validate the presence of discourses (Baker, 2006). Furthermore, a descriptive quantitative approach is used since it is one of the corpus-based study characteristics. That is, quantitative analysis uncovers key patterns by assessing the frequency and relative usage of language features, whereas descriptive analysis interprets the communicative functions associated with those observed quantitative patterns (Conrad, 1999). It aims to get a better comprehension of the LBs manifestation in six disciplines (Electrical Engineering, English Language Teaching, Linguistics, Biochemistry, Psychology, and Medicine). Besides, the Z-test is also applied to calculate the differences

in the proportion of LBs between two corpora, Hard Science and Soft Science (Kurniawan & Haerunisa, 2023), testing the following hypotheses:

- H_0= There is no significant difference in the proportion of LBs occurrences in the two corpora.
- H1= There is a significant difference in the proportion of LBs occurrences in the two corpora.

Additionally, the Z-test's alpha threshold was set at 0.05, meaning that a p-value higher than 0.05 does not provide sufficient evidence to reject H_0. To give a better understanding of the results, the study's results are presented using figures, tables, explanations, and extracts.

Corpora

The data used in this study was articles from various disciplines include: Electrical Engineering (Maswana et al., 2015), Biochemistry (Kanoksilapatham, 2005), Medicine (Fryer, 2012), English Language Teaching (Rochma et al., 2020), Psychology (Yang, 2022), and Linguistics (Alamri, 2020). All RAs were sourced from Scopus indexed databases which accessed from the Scopus.com website. The data comprised 180 high-impact RAs. The 180 RAs consist of 30 abstracts of each discipline (Alamri, 2020; Wannaruk & Amnuai, 2016), which are grouped into two corpora, as seen in Table 3.

Table 3. Description of the Corpora

	Corpus	Number of RAAs	Word counts
Hard Science	Electrical Engineering	30	4325
	Biochemistry	30	4523
	Medicine	30	9511
Soft Science	English Language Teaching (ELT)	30	5810
	Psychology	30	4310
	Linguistics	30	5351
Total		180	33.83

With a total number of words of 33,380, this corpus is considered small because it is in the middle of the corpus size studied based on Chen and Baker (2016), namely 26,000 to 88,000 words.

Data Collection and Analysis

For identifying the LBs, this study used a computer software program named AntConc 4.3.1 (Anthony, 2024). AntConc is the common computer software program which is used to analyze language corpus, especially lexical bundles (Farhang-Ju et al., 2024; Goodarzi et al., 2024; Jasim, 2023; Kurniawan & Haerunisa, 2023; Richter et al., 2022). Many features in Antcont are useful for in-depth textual analysis, particularly the N-Gram feature, which is used to identify lexical bundles' occurrence.

To achieve successful analysis with the program, the abstract section of each RA is separated and converted into a plain text (txt format). The txt file of abstracts was imported into the software which separated into two corpora named Hard Science and Soft Science. Then, the imported corpus was analyzed with the N-Gram feature. This study employed a cut off criteria based on the guideline established by Biber and Barbieri (2007) which state that in order to prevent the influence of individual author styles, word combinations must appear at least three times and be dispersed across at least three different texts. Moreover, this study only generates the three- and four-word bundles as it is more common than the other longer lexical bundle (Biber et al., 1999).

FINDINGS

In order to address the RQs, this section reports the lexical bundles of three- and four-words that were found in both hard science and soft science corpora include the discussion of their structural and functional characteristics using taxonomies.

After the application of preset criteria about the identification of LBs, the N-gram feature of AntConc presents the different number of LBs in each corpus. The program shows a list of 166 three-word and 47 four-word LBs used in Hard Science RAAs. On the other hand, the

program generated the list of 70 three-word and 14 four-word LBs in Soft Science RAAs. However, a significant number of initial lexical bundles in both corpora exhibited repetitive patterns. Consequently, a manual elimination process was conducted based on the exclusion criteria outlined by Salazar (2014). The applied criteria for exclusion included:

1. Fragments of other bundles. For example, the three-word bundle "on the other" and the four-word bundle "on the other hand" appeared eight times. Bundles "on the other" always appeared as a fragment of "on the other hand." "On the other" was thus not included.
2. Bundles ending in articles. For example, "in accordance with the" was a four-word bundle that was an extension of "in accordance with," and both words occurred with the same frequency. "In accordance with the" was ignored.
3. Bundles that lack semantic meaning or textual support such as "years and were."
4. Bundles that contain arbitrary numbers, like "two or more."
5. Pointless groupings, such "et al in."

However, this study included topic-specific bundles such as "patients with Covid" following Kurniawan and Haerunisa (2023). After following the elimination procedure, the final lists of LBs were 115 three-word and 38 four-word LBs in Hard Science, also 55 three-word and 9 four-word LBs Soft Science, as shown in Table 4.

Table 4. LBs occurrences information

Lexical bundles	Hard Science		Soft Science	
	No. of bundles	Overall freq.	No. of bundles	Overall freq.
Three-word	115	568	55	233
Four-word	38	160	9	30
Total	153	728	64	263

There appears to be a difference in the amount of LB occurrences between Hard Science and Soft Science, as shown in Table 4. This finding implies that Hard Science and Soft Science differ

noticeably, at least when it pertains to the quantity of LB occurrences. In addition, Table 5 and 6 present an overview of LBs in each corpus by listing the three most common of three- and four-word bundles in this study.

Table 5. Most frequent three-word LBs in Hard Science and Soft Science (top 3)

Rank	Hard Science		Soft Science	
	LBs	Freq.	LBs	Freq.
1	patients with COVID-19	24	as well as	15
2	coronavirus disease (covid)	22	individual differences in	9
3	of patients with	16	based on the	7

**Table 6**. Most frequent four-word LBs in Hard Science and Soft Science (top 3)

Rank	Hard Science		Soft Science	
	LBs	Freq.	LBs	Freq.
1	severe acute respiratory syndrome	17	as well as the	5
2	acute respiratory distress syndrome	8	on the basis of	4
3	in patients with covid	7	are discussed in terms	3

Table 5 and 6 show that Hard Science lexical bundles are primarily composed of specific medical terms, including "severe acute respiratory syndrome" and "patients with COVID," which represent the discipline's exacting and technical nature. On the other hand, Soft Science lexical bundles, including "as well as" and "on the basis of," are typically less specialized and more expansive, which fits with the discipline's more interpretive and generic approach. This result

indicates the differing linguistic and rhetorical demands of Hard Science and Soft Science.

LBs Structures in Hard Science and Soft Science

This subsection examines the structural variation of lexical bundles in Hard Science and Soft Science using the framework by Biber (2004). This framework divides LBs into three categories: verb phrases, noun phrases, and prepositional phrases. Additionally, this analysis incorporated the other categories presented by Nasrabady et al. (2020), which include adjectival phrases, adverbials, be+adjective/adverb+to structures, and bundles beginning with conjunctions, to gain deeper understanding into LBs in both corpora.
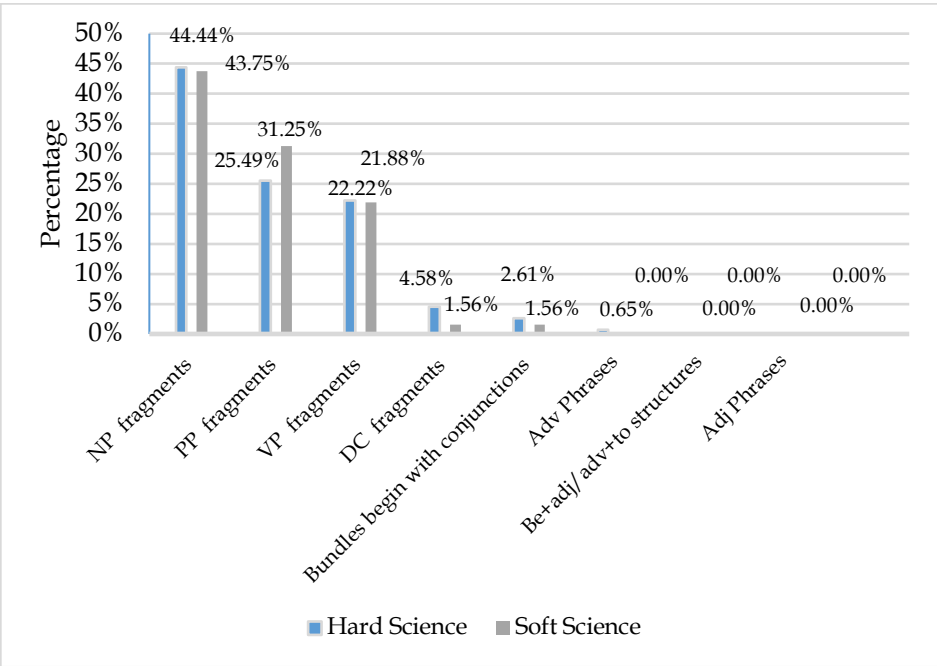
Figure 1. Distribution of LBs structures in Hard Science and Soft Science

The results of this study are aligned with a number of earlier studies that highlight the prevalence of phrasal lexical bundles (LBs) in academic abstracts, include those by Abdollahpour and Gholami

(2018), Niu (2014), and Varghaei and Khodadadi (2022). Figure 1 revealed that noun phrase (NP) fragments are used most frequently in both disciplines (44.44% in Hard Science and 43.75% in Soft Science), followed by prepositional phrase (PP) fragments (31.25% in Soft Science and 25.49% in Hard Science) and verb phrase (VP) fragments (22.22% in Hard Science and 21.88% in Soft Science). These results are in line with those of Abdollahpour and Gholami (2018) and Varghaei and Khodadadi (2022), which emphasize the predominance of NP and PP fragments in medical abstract. However, this analysis also shows a new finding that in Soft Science, the frequency of PP is higher than in Hard Science.

In contrast, other categories like dependent clause (DC) fragments, adjectival phrases, and bundles that start with conjunctions, are either barely represented or not at all in both corpora. DC fragments (e.g., to describe the) served as a modifier. (1)

1) We aimed **to describe the** CT findings across different time points throughout the disease course. (HS)

Those DC-based bundles with to-clause fragments used to explain an action (Kurniawan & Haerunisa, 2023).

Table 7. Z-test calculation of LBs main structures

Structures	Hard Science	Soft Science	z score	*p* value
VP fragments	22.22%	21.88%	0.056	0.955
DC fragments	4.58%	1.56%	1.074	0.283
NP fragments	44.44%	43.75%	0.094	0.925
PP fragments	25.49%	31.25%	-0.869	0.3845
Adjectival phrases	-	-	-	-
Adverbials	0.65%	-	0.65	0.516
Be+adj/adv+to structures	-	-	-	-
Bundles begin with conjunctions	2.61%	1.56%	0.468	0.639

Moreover, only Hard Science uses adverb phrases (e.g., 'more likely to'), adding details about degree or likelihood. Even no adjectival phrases and bundles with be+adj/adv+structures were found in either corpus. Lastly, bundles beginning with conjunction also appeared in low percentage in both corpora (e.g., '*and discriminant validity*' and '*and ground glass*'), introducing additional or complementary information. As seen in Table 7, the most common lexical bundle structures were noun phrase (NP) fragments with comparable percentages in Hard Science (44.44%) and Soft Science (43.75%). However, with an indicator p-value > 0.05, the data showed a consistent usage of structural categories in lexical bundles across Hard and Soft Science with no structure showing a significant proportion difference.

Lexical bundle (LB) substructures were further examined to allow for a more thorough comparison, as seen in Table 8. To improve structure and clarity, the classification underwent a number of adjustments. Based on research by Berūkštienė (2018) and Pearson (2021), the first modification was changing the subcategory "connector + 3rd person pronoun + VP fragments" to "pronoun/noun + VP fragments." Another modification added adverb and adjective clauses as new substructures for bundles with DC fragments, which was initiated by Nasrabady et al. (2020). Bundles that had adverbs, adjectival phrases, be+adjective/adverb+to structures, and those that began with conjunctions, however, were not included in the following analysis because they had no substructures.

Table 8. LBs substructures distribution.

Structures	Substructures	Hard Science	Soft Science	Z score	*p* value
VP fragments	Pronoun/noun + VP fragments	50.00%	40.00%	0.646	0.516
	VP with non-passive verb	8.82%	-	1.1874	0.2351
	VP with passive verb	41.18%	60.00%	-1.216	0.2248
	WH-question	-	-	-	-
DC fragments	To-clause	71.43%	100.00%	-0.617	0.537
	That-clause	14.29%	-	0.404	0.686
	Adverbial clause	14.29%	-	0.404	0.686

	Adjective clause	-	-	-	-
NP fragments	NP + of-phrase fragments	36.76%	50.00%	-1.2	0.23
	NP + other post modifier fragments	22.06%	25.00%	-0.312	0.755
	Other NP expressions	41.18%	25.00%	1.497	0.134
PP fragments	PP expressions	92.31%	90.00%	0.301	0.763
	Comparative	7.69%	10.00%	-0.301	0.763

Table 8 shows that there is no statistically significant difference between the two groups under analysis in any of the lexical bundle substructures, as all p-values exceed the threshold of 0.05. In other words, no single substructure exhibits a clear pattern or distribution, it appears that the groups' use of substructures is generally consistent. Thus, H0 was failed to be rejected for all the structures.

NP-based Bundles
According to NP-based bundle analysis, fascinatingly, Hard Science and Soft Science have different dominant NP substructures. The most common NP expressions in Hard Science are 'other NP expressions' (41.18%), with bundles like "science and technology", "an intensive care unit" and "the median age," (2) helping to succinctly indicate specific topics (Abdollahpour & Gholami, 2018). This is not in line with Kim and Lee (2021) and Niu (2014) which focused on medical abstract. Their study showed that the most frequent bundle is NP with of-phrase fragments.

2) **The median age** of deceased patients (68 years) was significantly older than recovered patients (51 years). (HS)

In comparison, NP + of-phrase fragments account for 50.00% of the total in Soft Science. Often, these bundles outline relationships and areas of concentration. (3)

3) Discussion focuses on **the role of** behavioral regulation in early academic achievement and preparedness for kindergarten. (SS)

The last point is that NP with post-modifier fragments, like "ways in which," contribute 22.06% in Hard Science and 25.00% in Soft Science, indicating their function in giving the central noun more specificity. This is consistent with Niu (2014) findings that these bundles provide academic abstracts more context and precision.

PP-based Bundles

The analysis of prepositional phrase (PP)-based bundles showed that 'Prepositional phrase (PP) expressions' are the most common substructure of bundles based on PPs, which are found to be highly prevalent in both Hard Science and Soft Science. PP expressions make up 92.31% of PP-based bundles in Hard Science, while they make up 90.00% in Soft Science. Bundles like "in patients with" (4) which underline the links between variables or components in research, can offer crucial locative, descriptive or comparative information. This predominance is consistent with research by Kim and Lee (2021) and Niu (2014), who highlighted the importance of prepositional words in abstracts, especially in interdisciplinary and medical research where clear links between components are crucial.

4) Both helper T (Th) cells and suppressor T cells **in patients with** COVID-19 were below normal levels, with lower levels of Th cells in the severe group. (HS)

Moreover, Comparative phrases, the second substructure, make up 7.69% of Hard Science and 10.00% of Soft Science. These bundles, such as "as well as," (5) highlight comparisons or contrasts. This is in line with the findings of Abdollahpour and Gholami (2018) and Varghaei and Khodadadi (2022), who emphasized the value of comparison constructions in medical and scientific presentations when presenting findings and talking about consequences. Also, the bundle "as well as" was used to designate two components as being equally significant (Kurniawan & Haerunisa, 2023).

5) An individual's self-efficacy and outcome expectations were found to be positively influenced by the encouragement of others in their work group, **as well as** others' use of computers. (SS)

VP-based Bundles

Table 8 analyses are consistent with a finding from a previous study, Niu (2014) which concentrates on LBs in English abstracts of Chinese and International Journals. The table reveals an intriguing dominance in the VP fragment substructures, where VP with pronoun/noun + VP fragments appeared as the most frequent substructure in Hard Science (50.00%). This substructure is frequently employed to explain goals, conclusions, or circumstances. (6)

6) **Results suggest that** a phonological deficit can appear in the absence of any other sensory or motor disorder, and is sufficient to cause a literacy impairment, as demonstrated by five of the dyslexics. (HS)

These results show a difference from previous research in which passive VP was the most frequently occurring in medical RAs (Kim & Lee, 2021; Varghaei & Khodadadi, 2022). In this study, VP with passive verbs appeared as the second most frequent occurring in Hard Science (7) while in Soft Science, VP with a passive verb was the most frequently occurring structure (60%), followed by pronoun/noun + VP fragment (40%).

7) Older age **was associated with** greater risk of development of ARDS and death likely owing to less rigorous immune response. (HS)

In addition, two other substructures of VP fragments namely VP with non-passive verbs and VP with WH-questions appeared less frequently in both corpora. Even no VP with WH-questions occurred in both corpora.

DC-based Bundles

Analyzing bundles based on dependent clauses (DC) shows intriguing trends in how they are used in both soft and hard sciences. With 71.43% in Hard Science and 100% in Soft Science, 'to-clauses' are the most prevalent substructure among DC fragments, dominating this category. These bundles express purposes or objectives, which reflects their function in outlining the goals and intentions of study. (8)

8) Based on the mode analysis, an approximation method is developed **to estimate the** peak gain point, which is useful in LLC design.

Only 14.25% of academic abstracts in the Hard Sciences and none at all in the Soft Sciences contain adverbial phrases (such as "who were admitted to") or that-clauses (such as "those of the"). Adverbial clauses helped develop connections between concepts for the creation of cohesive writings (Fang, 2006), whereas that-clauses offer information by referencing or comparing features. Their restricted and specialized importance in these sectors is reflected in their low usage.

LBs Function is Hard Science and Soft Science
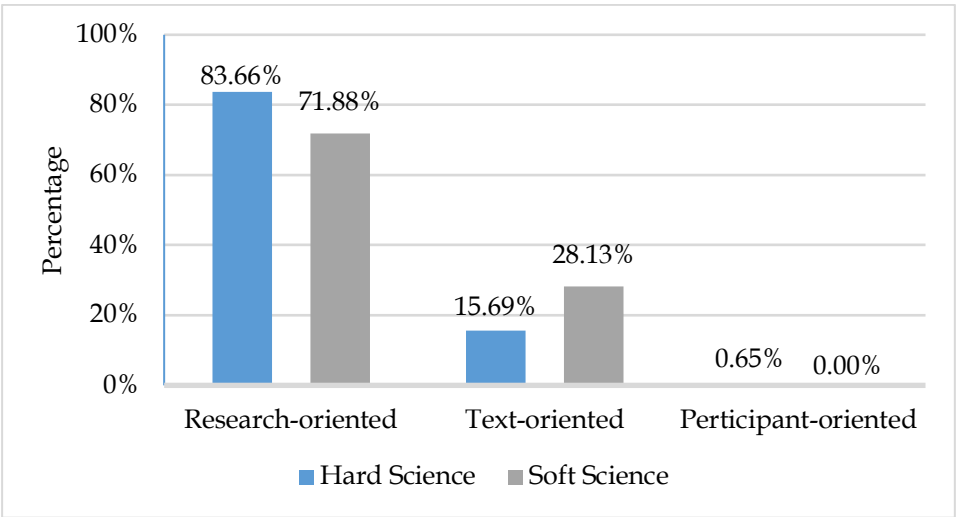
Figure 2. Distribution of LBs functions in Hard Science and Soft Science

Initially, the final lists of lexical bundles (LBs) were functionally classified using Hyland's (2008b) framework. The chart illustrates that research-oriented bundles predominate in both corpora, making up 83.66% in Hard Science and 71.88% in Soft Science. This is consistent with earlier research, including Kim and Lee (2021), Varghaei and Khodadadi (2022), Niu (2014) and Abdollahpour and Gholami (2018), which also found that research-oriented bundles are commonly used in academic writing to describe procedures, present results, or contextualize research within a particular framework.

On the other hand, as a result of the greater interpretive and explanatory demands of Soft Science writing, text-oriented bundles, which aid in discourse organization and reader guidance are more common in Soft Science (28.13%) compared to Hard Science (15.69%). This finding is consistent with observations made by Niu (2020) and Hyland (2008b), who pointed out that text-oriented bundles are frequently used in Soft Science disciplines to build relationships and make arguments more understandable for readers. Moreover, participant-oriented bundles, which indicate the author's position or interaction with the audience, are hardly present—they make up 0.65% of Hard Science abstracts and 0% of Soft Science abstracts.

The previous section, Table 8 presents the sub-functions of LBs that were examined to have a better understanding of their presence in the two corpora. In order to examine bundles that did not suit Hyland's (2008b) framework, other frameworks were also used. The other sub-functions were grouping, citation, generalization, and objective from Salazar (2014), as well as doubling, exemplifier, and questioning from Nasrabady et al. (2020). Additionally, the categorization of Hyland's (2008b) transition and resultative signal sub-functions into additive, comparative, inferential, and causal categories by Salazar (2014) was also used.

Table 9. LBs sub-functions distribution.

Functions	Sub-functions	Hard Science	Soft Science	Z score	*p* value
Research-oriented	Location	0.78%	10.87%	-3.216	0.001
Text-oriented	Procedure	12.50%	15.22%	-0.467	0.641
	Quantification	5.47%	2.17%	0.915	0.36
	Description	8.59%	47.83%	-5.822	0
	Topic	66.41%	19.57%	5.467	0
	Grouping	1.56%	2.17%	-0.273	0.785
	Doubling	4.69%	2.17%	0.744	0.457
	Additive	4.17%	5.56%	-0.209	0.834
	Comparative	20.83%	11.11%	0.837	0.403
	Inferential	8.33%	27.78%	-1.673	0.094
	Causative	8.33%	16.67%	-0.825	0.409
	Structuring signals	25.00%	11.11%	1.134	0.257
	Framing signals	-	11.11%	-1.673	0.094
	Citation	4.17%	-	0.877	0.381
	Generalization	-	5.56%	-1.169	0.243
	Objectives	29.17%	11.11%	1.411	0.158
	Exemplifier	-	-	-	-
	Questioning	-	-	-	-
Participant-oriented	Stance features	100.00%	-	-	-
	Engagement features	-	-	-	-

Table 9 shows the apparent statistical differences in the distribution of LB functions and sub-functions between research articles in Hard Science and Soft Science. Each sub-function appeared differently in Hard Science and Soft Science. The results are discussed in greater detail in the ensuing subsections.

Research-oriented Bundles

As the figure shows, the examination of research-oriented bundles identifies significant distinctions between Hard Science and Soft Science in their sub-functions. In Hard Science, topic bundles account for 66.41% of research-oriented bundles, while in Soft Science, they only make up 19.57%. Bundles like "patients with Covid-19" (9) or

"respiratory distress syndrome" are commonly used in Hard Science abstracts to precisely identify and highlight the primary research topic. This is consistent with findings by Niu (2020), who highlighted the use of topic-oriented bundles in Hard Science to convey research focal areas.

9) The lungs from **patients with Covid-19** also showed distinctive vascular features, consisting of severe endothelial injury associated with the presence of intracellular virus and disrupted cell membranes.

Unlike hard science abstracts, which tend to emphasize procedural or result-oriented bundles, soft science abstracts make extensive use of descriptive bundles, such as "the development of", (10) to offer richer contextual and interpretive information (Varghaei & Khodadadi, 2022).

10) This article presents **the development of** a brief, self-report measure of female sexual function. Initial face validity testing of questionnaire items, identified by an expert panel, was followed by a study aimed at further refining the questionnaire.

The Z-test study shows that Hard Science and Soft Science differ significantly in certain sub-functions of research-oriented bundles: Topic, description and location sub-functions (data marked in bold). The prevalence of topic bundles is much higher in Hard Science (66.41%) compared to Soft Science (19.57%), with a p-value of 0.000 and a Z score of 5.467, highlighting Hard Science's emphasis on precisely specifying study topics and areas. (9) The p-value of 0.000 and the Z score of -5.822 show that, on the other hand, description bundles are substantially more common in Soft Science (47.83%) than in Hard Science (8.59%). This is consistent with Soft Science's interpretative focus on detailed and contextual descriptions. (10)

Additionally, with a Z score of -3.216 and a p-value of 0.001, location bundles (11) also exhibit a significant difference, being more

prevalent in Soft Science (10.87%) compared to Hard Science (0.78%), highlighting the contextual framing frequently observed in Soft Science abstracts. These notable distinctions show how the two fields employ different rhetorical methods, with Soft Science emphasizing in-depth explanations and contextualization and Hard Science concentrating on succinct, topic-driven discourse.

11) Yet Anthony also used his discursive inquiry to "trouble the water" in his classroom and **in the study** group workshops.

Text-oriented Bundles

Different distributions between sub-functions in Hard Science and Soft Science are revealed by the examination of text-oriented bundles. More **objective bundles**, like "to evaluate the," (12) are found in Hard Science (29.17%) than Soft Science (11.11%), indicating that Hard Science places a greater focus on clearly articulating study objectives. Comparably, **structuring bundles**, like "in this paper," (13) make up 25.00% of Hard Science and 11.11% of Soft Science, indicating their function in directing readers and structuring discourse (Hyland, 2008b). Moreover, **comparative bundles**, such as "was the most," (14) are more common in Hard Science than Soft Science, highlighting Hard Science's emphasis on accurate comparisons to properly convey study findings.

12) This study aimed **to evaluate the** clinical characteristics of COVID-19 in pregnancy and the intrauterine vertical transmission potential of COVID-19 infection.
13) **In this paper,** we introduce a framework for VHR scene understanding.
14) On admission, ground-glass opacity **was the most** common radiologic finding on chest computed tomography (CT) (56.4%).

In contrast, the other sub-functions of text-oriented bundles reveal notable variations between Hard Science and Soft Science. The fact that inferential bundles, such "findings indicate that," (15) are far more

common in Soft Science (27.78%) than in Hard Science (8.33%) highlights the interpretive character of Soft Science and its emphasis on making connections and deductions from data. Moreover, Soft Science (16.67%) places a greater emphasis on causative bundles such as "as a result of," (16) than in Hard Science (8.33%) which indicate cause-and-effect relationships (Salazar, 2014).

15) Furthermore, the **findings indicate that** a classification system that is based on the simple view has advantages over standard systems that focus only on word recognition and/or reading comprehension. (SS)

16) **As a result,** much research is still based on the old Kučera and Francis frequency norms. (SS)

Furthermore, framing bundles like "in terms of" are commonly used in the Soft Sciences to contextualize arguments, and additive bundles (e.g., 'and the results'), which are less prevalent overall, are used slightly more in the Soft Sciences than in the Hard Sciences to help with idea connections. Certain text-oriented bundles, such exemplifier and generalization bundles, are absent from both fields, suggesting their limited use in academic abstracts.

Lastly, the Z-test results suggest that none of the text-oriented sub-functions exhibit statistically significant differences between Hard Science and Soft Science, with all p-values over 0.05, despite these apparent distributional discrepancies. This lack of statistical significance implies that both fields use text-oriented bundles in a similar way overall, even though there are differences in the presence or absence of certain sub-functions.

Participant-oriented Bundles

Participant-oriented bundles show a significant lack of variance between Hard Science and Soft Science, with engagement features completely missing from both corpora and stance features making up 100% of Hard Science participant-oriented bundles. Stance bundles,

like "more likely to," (17) highlight the author's role in delivering the research by expressing the authors' views, level of confidence, or assessment of their findings. This supports the findings of Niu (2014), who observed that participant-oriented bundles—in particular, posture features—are more common in fields that place a high value on findings that are straightforward and easy to understand.

17) The 2019-nCoV infection was of clustering onset, is **more likely to** affect older males with comorbidities, and can result in severe and even fatal respiratory diseases such as acute respiratory distress syndrome.

The lack of engagement elements, which usually entail communication with the audience (such as instructions or inquiries), highlighted that academic abstracts typically maintain an impersonal tone, minimizing direct interaction with readers.

CONCLUSION

This study explored the use of lexical bundles in abstracts of high-impact research articles across disciplines, especially comparing Hard Science and Soft Science. The analysis revealed information about the frequency, structural patterns, and functional distribution of lexical bundles by utilizing Biber's (2004) structural taxonomy and Hyland's (2008b) functional classification, along with other categories to support the investigation. The results demonstrated distinct disciplinary similarities and variations in both structural and functional aspects. The Hard Science and Soft Science were dominated structurally by bundles based on noun phrases, followed by PP-based bundles and VP-based bundles. This finding indicates that both corpora utilize more phrasal bundles than clausal ones. However, the frequency of PP-based bundles is higher in Soft Science. This higher use of PP-based bundles in hard science abstracts is indicative of the field's demand for exact locational, temporal, and relational markers in order to effectively communicate technical and factual information. Functionally, both fields had the highest concentration of research-

oriented bundles, especially in topic bundles in the Hard Sciences and description bundles in the Soft Sciences. The higher use of topic bundles indicates a greater emphasis on defining the topic or focus of the investigation, while the higher use of description in Soft Science indicates a focus on elaborating attributes. This study emphasizes the variety of ways lexical bundles (LBs) are employed in academic writing, pointing out that their functions and forms are influenced by discourse communities and might not be appropriate in all contexts. Despite the limited size of its corpus, the study identifies recurrent patterns that serve as the basis for recommendations on how abstracts can align with disciplinary standards. For more accurate and thorough insights of LB usage, future studies should employ sophisticated analytical tools and expand the corpus size.

ACKNOWLEDGMENTS

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES

While preparing this paper, the authors utilized Grammarly to reduce writing errors and ChatGPT to help with data calculation. The writers have used these tools to thoroughly evaluate and revise the content as needed, and they are solely responsible for the final version of this publication.

REFERENCES

Abdollahpour, Z., & Gholami, J. (2018). Building blocks of medical abstracts: Frequency, functions and structures of lexical bundles. *The Asian ESP Journal*, *14*(1), 83-111.

Alamri, B. (2020). A comparative study of Saudi and international journals of applied linguistics: The move–bundle connection approach. *Journal of Language and Education*, *6*(2), 9-30. https://doi.org/10.17323/jle.2020.10531

Anthony, L. (2024). *AntConc (4.3.1)* [Computer Software]. Tokyo, Japan: Waseda University https://www.laurenceanthony.net/software/antconc/

Baker, P. (2006). *Using Corpora in Discourse Analysis*. Bloomsbury Academic.

Becher, T., & Trowler, P. (2001). Academic tribes and territories. McGraw-Hill Education (UK).

Berūkštienė, D. (2018). A corpus-driven analysis of structural types of lexical bundles in court judgments in English and their translation into Lithuanian. *Kalbotyra*, *70*(70), 7-31. https://doi.org/10.15388/Klbt.2017.11181

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), 375-311. 10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *Lexical bundles in university spoken and written registers*, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

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

Biber, D., Johannsson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman Grammar of Spoken and Written English*. Longman. (Biglan, 1973)

Biglan, A. (1973a). The Characteristics of Subject Matter in Different Academic Areas. *Journal of applied psychology*, 57(3), 195-203. https://psycnet.apa.org/doi/10.1037/h0034701

Chen, Y.-H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, *37*(6), 849–880. https://doi.org/10.1093/applin/amu065

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, *27*(1), 1-18. 10.1016/s0346-251x(98)00046-3

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

Deng, L., & Liu, J. (2023). Move–bundle connection in conclusion sections of research articles across disciplines. *Applied Linguistics*, *44*(3), 527-554. https://doi.org/10.1093/applin/amac040

Fang, Z. (2006). The language demands of science reading in middle school. International Journal of Science Education, 28(5), 491–520. https://doi.org/10.1080/09500690500339092

Farhang-Ju, M., Jalilifar, A., & Keshavarrz, M. H. (2024). Specificity and generality of lexical bundles in the rhetorical moves of applied linguistics research article introductions. *Journal of English for Academic Purposes*, *69*, 1-12. https://doi.org/10.1016/j.jeap.2024.101387

Fauzan, U., Lubis, A. H., & Kurniawan, E. (2020). Rhetorical moves and linguistic complexity of research article abstracts in international applied linguistics journals. *The Asian E.S.P. Journal*, *16*(5.2), 219-247.

Fryer, D. L. (2012). Analysis of the generic discourse features of the English-language medical research article: A systemic-functional approach. *Functions of Language*, *19*(1), 5-37. 10.1075/fol.19.1.01fry

Ghasempour, B., & Farnia, M. (2017). Contrastive move analysis: Persian and English research articles abstracts in law. *The Journal of Teaching English for Specific and Academic Purposes*, *5*, 739-753. 10.22190/JTESAP1704739G

Goodarzi, R., Ghonami, J., & Abdollahpour, Z. (2024). Lexical bundles in the discussion sections of medical sciences articles: frequencies, syntactic structures, and discourse functions.

Language Teaching Research Quarterly, *40*, 1-28. https://doi.org/10.32038/ltrq.2024.40.01

Hirsch, J. E. (2005). An index to quantify an individual's scientificresearch output. *The Proceedings of the National Academy of Sciences (PNAS)*, *102*(46), 16569. 10.1073pnas.0507655102

Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins Publishing Company.

Hyland, K. (2000). *Disciplinary discourses: social interactions in academic writing*. Longman.

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*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Jasim, Y. B. (2023). Restating research findings in research articles discussion section: A corpus analysis of linguistic cues and lexical bundles. *GEMA Online Journal of Language Studies*, *23*(1), 225-243. http://doi.org/10.17576/gema-2023-2301-12

Kafes, H. (2012). . Cultural traces on the rhetorical organization of research article abstracts. *International Journal on New Trends in Education and Their Implications*, *3*(3), 207-220. http://www.ijonte.org/FileUpload/ks63207/File/20.kafes.pdf

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, *24*(3), 269-292. 10.1016/j.esp.2004.08.003

Kashiha, H. (2023). Beyond words in evaluation: Formulaic language in critical reviews of research articles across disciplines. *Russian Journal of Linguistics*, *27*(2), 251-275. https://doi.org/10.22363/2687-0088-34320

Kim, E.-S., & Lee, E.-J. (Spring 2021). A corpus-based analysis of lexical bundles between English L1 and L2 writers in medical journal abstracts: Focusing on structures and functions. *The Journal of*

Asia TEFL, *18*(1), 142-160. http://dx.doi.org/10.18823/asiatefl.2021.18.1.9.142

Kurniawan, E., Dallyono, R., & Cahyowati, A. (2019). Exploring logical connectors in journals with different indexing levels: A comparison between international and national indexed journals. *Indonesian Journal of Applied Linguistics*, *9*(1), 77. 10.17509/ijal.v9i1.16088

Kurniawan, E., & Haerunisa, Z. F. (2023). A comparative study of lexical bundles in accepted and rejected applied linguistic research article introductions. *Studies in English Language and Education*, *10*(2), 628-648. https://doi.org/10.24815/siele.v10i2.28119

Lorés, R. (2004). On RA abstracts: from rhetorical structure to thematic organisation. *English for Specific Purposes*, *23*(3), 280-302. https://doi.org/10.1016/j.esp.2003.06.001

Maswana, S., Kanamaru, ,. T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, *2*, 1-11. 10.1016/j.amper.2014.12.002

Nasrabady, P., Shirvan, M. E., & Golparvar, S. E. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages*, *6*(3), 175-197. https://doi.org/10.1080/21698252.2020.1797992

Niu, G. (2014). Structurally and functionally comparative analysis of lexical bundles in the English abstracts of Chinese and international journals. In X. Su & T. He (Eds.), *Chinese lexical semantics* (pp. 349-357). New York: Springer International Publishing. https://doi.org/10.1007/978-3-319-14331-6_35

Qi, H., & Pan, F. (2020). Lexical bundle variation across moves in abstracts of medical research articles. *Southern African Linguistics and Applied Language Studies*, *38*, 109-128. https://doi.org/10.2989/16073614.2020.1763814

Richter, K. G., Gaskaree, B. L., & Mirza, M. (2022). A functional analysis of lexical bundles in the discussion sections of applied linguistics research articles: A cross-paradigm study. *Russian*

Journal of Linguistics, *26*(3), 625-644. https://doi.org/10.22363/2687-0088-27752

Rochma, A. F., Triastuti, A., & Ashadi. (2020). Rhetorical styles of Introduction in English language teaching (ELT) research articles. *Indonesian Journal of Applied Linguistics*, *10*(2), 304-314. https://doi.org/10.17509/ijal.v10i2.28593

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

Samraj, B. (2024). Disciplinary differences in lexical bundles use: A cautionary tale from methodological variations. *Journal of English for Academic Purposes*, *70*, 1-12. https://doi.org/10.1016/j.jeap.2024.101399

Shahmoradi, N., Jalali, H., & Ghadiri, M. (2021). Lexical bundles in the abstract and conclusion sections: The case of applied linguistics and information technology. *Applied Research on English Language*, *10*(3), 47-76. 10.22108/are.2021.128024.1703

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method and results sections of the research article. *Corpora*, *12*(1), 1-22. 10.3366/cor.2017.0107

Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambrigde University Press.

Tocalo, A. W. I. (2021). Move structures and their rhetorical verbs of research article abstracts across Englishes. *Indonesian Journal of Applied Linguistics*, *11*(1), 1-10. https://doi.org/10.17509/ijal.v11i1.34593

Varghaei, E., & Khodadadi, G. (2022). Comparing lexical bundles in medical research article abstracts of Iranian and foreign journals. *GEMA Online Journal of Language Studies*, *22*(3), 86-103. http://doi.org/10.17576/gema-2022-2203-05

Verma, M. K., & Yuvaraj, M. (2022). Measuring the online attention to business research papers: An altmetric study of selected journals with high impact factor. *Journal of Business & Finance Librarianship*, *27*(4), 279. https://doi.org/10.1080/08963568.2022.2100089

Wannaruk, A., & Amnuai, W. (2016). A comparison of rhetorical move structure of applied linguistics research articles published in international and national Thai journals. RELC Journal, 47(2), 193-211. 10.1177/0033688215609230

Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecol*, *26*, 1345-1349. https://doi.org/10.1007/s10980-011-9674-3

Yang, M. (2022). Connecting the functions of lexical bundles and moves in published research articles: the case of developmental and educational psychology. Nordic Journal of English Studies, 21(1), 141-189. 10.35360/njes.743