

Structures and Functions of Lexical Bundles in the Discussion Section of Research Articles: An Interdisciplinary Study

¹Leila Shoja*

²Logman Hoveizavi

Research Paper

IJEAP- 2411-2102 DOR: [20.1001.1.24763187.2024.13.4.4.5](https://doi.org/20.1001.1.24763187.2024.13.4.4.5)

Received: 2024-11-01

Accepted: 2024-12-27

Published: 2024-12-30

Abstract: Many studies have focused on lexical bundles, examining the generic profile of research article (RA), the most important scholastic genre for knowledge production and dissemination. However, lexical bundles have not been thoroughly investigated in different academic disciplines comparatively while disciplinary features do impact the genre in a variety of ways. This study, therefore, aimed to examine structures and functions of lexical bundles in the discussion of RAs, one of the most significant academic part-genres. To this end, 50 English RAs of applied linguistics and 50 RAs of geology published from 2017 to 2022 were selected from several peer-reviewed journals of the fields as the corpus. The RAs, representing soft and hard disciplines, were analyzed based on Biber et al. (1999, 2004)'s models of lexical bundles using AntConc software. The results revealed that the most frequent lexical bundles in both disciplines were phrasal (passive and prepositional phrases) and of referential function. It was also found that there are marked similarities between the two subcorpora regarding the structural and functional categories of lexical bundles. The findings promise implications for discourse analysis and the field of English for Academic Purposes in relation to syllabus design and materials development for academic reading and writing.

Keywords: Applied Linguistics, Disciplines, Discussion, Geology, Lexical Bundles, Research Article

Introduction

In order to integrate into their respective disciplinary communities and achieve acceptance, academic writers must familiarize themselves with the defining characteristics of these communities (Flowerdew, 2000). Writing practices constitute a significant aspect of these characteristics and conventions; therefore, researchers must attain academic literacy to effectively communicate with peers both locally and globally in order to be acknowledged as members of their communities (Belcher, 2007).

Academic literacy encompasses an understanding of various elements of academic discourse, including linguistic, textual, social, and cultural dimensions, as well as familiarity with the English language utilized across different academic disciplines. This proficiency is crucial for the creation of suitable academic texts (Ferenz, 2005). After engaging in "knowledge-telling" tasks and producing corresponding texts during their postgraduate studies, students are typically expected to undertake more intricate "knowledge-transforming" tasks, which provide them with the opportunity to generate new knowledge (Tardy, 2005, p. 325). This level of literacy allows students to take an active role in their academic communities, facilitating their involvement in writing, presenting, and disseminating their research findings (Tardy, 2005). Thus, beyond possessing school-oriented expertise—namely, content knowledge in a specific field—students must cultivate academic expertise that includes an understanding of the rhetorical practices pertinent to their disciplines. This latter form of expertise is

¹ Assistant Professor of TEFL (Corresponding Author), l.shoja@ilam.ac.ir; Department of English Language and Literature, Ilam University, Ilam, Iran.

² MA Student of TEFL, loqmanhoveizavi@gmail.com; Department of English Language and Literature, Ilam University, Ilam, Iran.

fundamental to their integration into disciplinary communities and enhances the relevance of their knowledge production and assertions (Geisler, 1994, as cited in Peters, 2011).

The role of English as the dominant language of science and the primary language for global scholarly communication is crucial in this context as multilingual scholars increasingly face pressure to publish their academic work in English (Belcher, 2007; Curry & Lillis, 2004; Flowerdew, 2000). Publications in English are highly regarded by institutions and organizations, influencing researchers' careers through opportunities for rewards and promotions (Lillis & Curry, 2006; Polo & Varela, 2009). Furthermore, these publications enable multilingual scholars to share their research with a broader audience and contribute diverse perspectives to the global research community (Curry & Lillis, 2010). Consequently, a considerable amount of research has been done on academic literacy to identify both discourse and non-discourse resources crucial for successful publication in English, thereby informing the teaching circles and academic communities. Some investigations have concentrated on linguistic and rhetorical resources while others have examined practices and experiences.

Linguistic and textual characteristics are fundamental components of academic literacy, being examined through various analytical frameworks, including discourse analysis. To address challenges such as enhancing access to academic communities for writers and mitigating obstacles researchers face in publishing their work, discourse analysts have investigated macro and micro structures. These include different genres and part genres along with the linguistic and rhetorical features inherent to the genres. Among the latter, lexical bundles (LBs) have found their place for more than half a century (Simpson-Vlach & Ellis, 2010). Being studied under different names such as 'formulaic sequences' (Wray, 2002), 'formulaic expressions' (Simpson, 2004), 'fixed expressions' (Moon, 1998), 'lexical phrases' (Nattinger & Decarrico, 1992), or 'multiword lexical units' (Cowie, 1992), LBs characterized as a group of words used together more frequently than chance, defining the meanings and coherence of a text (Hyland, 2008), are regarded as the distinguishable linguistic feature of academic writing. Many scholars believe that competency and incompetence in academic writing highly rely on exploiting LBs since the frequent presence of these bundles in writing denotes proficiency in language use and their absence signifies a lack of experience (Chen & Baker, 2010; Cortes, 2004).

Literature Review

Researchers and scholars have examined various characteristics of LBs. They have identified several key features of these bundles, including their frequency-based nature, continuity, grammatical incompleteness, functional completeness, and transparency of meaning (Biber & Barbieri, 2007). The primary defining attribute of LBs is their frequency of occurrence; a bundle must appear more than 20 times per million words to be classified as such, although many bundles exceed 100 occurrences per million words in certain registers. In contrast to other word combinations, such as pure idioms, which appear only 0.5 times per million words, this frequency is notably high (Cortes, 2004).

Furthermore, LBs can be characterized by their idiomaticity and fixedness, which are essential properties in the analysis of word combinations (Cortes, 2002). Of course, a significant number of them lack idiomaticity; instead, their meanings are clear (Cortes, 2004). Durrant (2015) regards LBs as crucial features of discourse, highlighting three significant properties: they can be automatically identified, they serve functional roles, and they exhibit high sensitivity to variations among text types. Additionally, for multiword expressions to qualify as LBs, they must meet specific frequency and dispersion criteria, occurring at least 20-40 times per million words and across diverse texts (Biber et al., 2004).

To better understand LBs, several researchers have examined their structural and functional characteristics to propose categories and classifications. About their structures, a distinction is made between phrasal and clausal bundles by Biber et al. (1999), with the former including noun phrases as well as prepositional phrases and the latter involving simple verb phrases or main clauses. Cortes (2013) divides lexical bundles into four major groups: (1) lexical bundles with a noun prepositional phrase; (2) lexical bundles with a verb phrase; (3) lexical bundles with a dependent clause; and (4) lexical bundles with noun and verb phrases. From a functional perspective, Biber et al. (2004) divide LBs into three

categories of referential bundles, stance expressions, and discourse organizers while they specify different subcategories for each. Special conversational functions are also a part of the model for spoken genres and contexts. Similarly, Hyland (2008) classifies LBs as research-oriented, text-oriented, and participant-oriented types. Research-oriented bundles are used for expressing activities and experiences, text-oriented lexical bundles for organizing texts, and participant-oriented lexical bundles for presenting perspectives and ideas about claims.

LBs are useful in academic writing since they are, in part, characterized by their formal structural requirements, which vary based on academic disciplines (Wood, 2015). Variations observed in the practices within discourse communities “influence both the preferred modes of communication in different disciplines and the rhetorical characteristics of genres that students are expected to manage in becoming competent members of the discourse community” (Bhatia, 2004, p. 36). Thus, intra- and inter-disciplinary studies on LBs in different genres can be helpful with understanding the profile of both different disciplines and various genres. In this regard, the current study was an attempt to investigate LBs in the discussion section of applied linguistics and geology research articles (RAs). The discussion section was the focus of this study as its importance has been underscored by various scholars (Dujcik, 2013; Moyetta, 2016). Among the different sections of a research article, the discussion section poses significant challenges for both novice and seasoned writers, as noted by Amnuai (2017). In this part of the writing, it is essential for authors to possess strong persuasive writing abilities in order to effectively convince their audience that their claims are significant (Pojanapunya & Todd, 2011). To do this successfully, writers should have a good knowledge of the formulaic language so that they can develop and organize their writing; possessing a solid grasp of LBs could aid inexperienced writers in enriching their mental lexicon, and formulating their ideas in the discussion section to achieve its communicative functions. Thus, the study tried to answer the following questions:

Research Question One: What are the most frequent LBs in the discussion section of applied linguistics RAs?

Research Question Two: What are the most frequent LBs in the discussion section of geology RAs?

Research Question Three: Is there any significant difference between LBs usage in the discussion section of applied linguistics and geology RAs?

The last research question is addressed through a null hypothesis, that is, there is no significant difference between LBs usage in the discussion section of applied linguistics and geology RAs.

Methodology

Design of the Study

Using a discourse analysis approach, this corpus-based study involved first identifying structures and functions of four-word LBs in the discussion section of RAs, and then using both descriptive and inferential statistics to examine their frequencies and the potentially significant variable of discipline in relation to them. Employing two analytical frameworks by Biber et al. (1999, 2004), this descriptive-analytic study did not involve any manipulation of the variables.

Corpus

The corpus of the present study consisted of the discussion sections of 100 English RAs of applied linguistics and geology published in several peer-reviewed high-quality journals. The sample of articles in applied linguistics was randomly selected from Applied Linguistics, TESOL Quarterly, English for Specific Purposes, Journal of Second Language Writing, Journal of Teaching Language Skills, Iranian Journal of Applied Linguistics, Iranian Journal of Applied Language Studies, and Research in Foreign Languages. Tunneling and Underground Space Technology, Bulletin of Engineering Geology and the

Environment, Bulletin of Engineering Geology and the Environment, Earthquake Engineering and Structural Dynamics, and Engineering Geology were navigated to choose RAs of geology.

The reason behind choosing the two disciplines was a lack of sufficient studies comparing RAs of soft and hard disciplines in terms of LBs. Applied linguistics was selected as a soft discipline and geology as a hard one. We also considered the previous studies on lexical bundles in research articles to choose the fields. Thus, considering the fact that geology is not investigated in lexical bundles studies and our intention to include applied linguistics in the research, these two were chosen as the focus. 50 articles for each field were selected because, as noted by Biber (2006), it is essential for a corpus to be sufficiently extensive to accurately reflect the prevalence of the features under investigation.

The articles were limited to 2017-2022 to not only learn about the most recent patterns of LBs use but control the variable of time which is important in diachronic studies. The corpus was selected through a simple random sampling procedure from among the journals. Having selected the articles, the researchers checked to see if they contained discussion sections or not. If not, they were removed from the analysis.

Procedure

After collecting all the electronic copies of the articles, all sections were removed and just the discussion sections remained for later analysis. As AntConc (Anthony, 2007) requires plain texts, discussion sections of all the articles were saved as plain texts before being uploaded to the program. Then, to extract LBs from those files, frequency counts of 4-grams using the N-grams command in AntConc (Anthony, 2007) were run by specifying 'n' to perform a full retrieval of any n-grams from the corpus. N-grams are "contiguous sequence [s] of 3 or 4 words identified purely through automatic means using a frequency-driven approach" (Flowerdew, 2015, p.105).

The criterion adopted for the definition of LBs was developed by Biber et al. (1999) who characterized these sequences as the most frequently occurring multi-word combinations within a specific register. The research concentrated on units comprising four-word sequences as Hyland (2008) noted that four-word bundles are more common than five-word bundles and typically exhibit clearer structures and functions compared to three-word bundles. Setting a frequency of five instances in five discussion sections of the corpus was done in AntConc. A cut-off frequency of 20 in one million words was also associated as a criterion. The criterion used to determine the cut-off points is in line with Cortes (2008)'s statement that a four-word combination should be seen twenty times in one million words and five or more texts to qualify as a lexical bundle. Equipped with these features, AntConc presented clusters of words linked to a search term, ordering them either by their frequency or in alphabetical order.

In the following stage, a thorough manual examination of each expression in the list was necessary to ascertain whether it was present in more than five texts from the corpus. Any expressions identified in fewer than five texts were not classified as LBs and were, therefore, eliminated due to their low frequency. AntConc provided a frequency list required for the qualitative analysis of the results: the description of structural and functional categories of LBs extracted from the corpus according to Biber et al. (1999, 2004)'s models (Tables 1 and 2). Based on the analyses, similar grammatical structures and functions among bundles led to their grouping. It is essential to emphasize that a second rater contributed to the identification and classification process, thereby enhancing the reliability of the conclusions drawn.

Table 1*Structural Taxonomy of LBs (Biber et al., 1999, pp.1015-1024)*

Categories	Examples
Noun phrase with of-phrase fragment	the beginning of the
Noun phrase with other post-modifier fragments	the way in which
Prepositional phrase with embedded of-phrase fragment	at the end of
Other prepositional phrase fragments	as in the case
Anticipatory it + verb phrase/adjective phrase	it is possible to
Passive verb + prepositional phrase fragment	is based on the
Copula be + noun phrase/ adjective phrase	is one of the
(verb phrase +) that-clause fragment	has been shown that
(verb/adjective +) to-clause fragment	are likely to be
Adverbial clause fragment	as soon as you
Pronoun/noun phrase + be (+...)	there was no significant
Other expressions	may or may not

Table 2*Functions of LBs (Biber et al., 2004, pp. 384-388)*

Categories	Subcategories	Examples
Stance expressions	Epistemic stance Attitudinal/modality	are more likely to, it is clear that
Discourse organizers	Topic introduction/focus Topic elaboration/clarification	in the first place, first of all the on the other hand
Referential expressions	Identification/focus Imprecision Specification of attributes Time/place/text reference	one of the most, is one of the and things like that in the form of, at the same time

Results

The findings of the study are presented according to the research questions.

LBs in the Discussion Section of Applied Linguistics RAs

The first research question concerns the most frequent LBs in the discussion section of applied linguistics RAs based on Biber et al. (1999)'s model. The analysis revealed 40869 types and 43058 tokens of them. It should be pointed that not all LBs found in the corpus were investigated. As the criterion for investigation was a minimum occurrence of 5 instances, those LBs following this criterion, that is, 722 bundles were included in the study, and the rest were excluded from the analysis. The most three common LBs found were **of the present study**, **the findings of the**, and **the results of the**, with the frequency of 23, 23, and 23 tokens respectively. The results of structural and functional analysis are reported in Tables 3 and 4.

Table 3

Structural Categories of LBs in the Discussion of Applied Linguistics RAs

Structures	Examples	Frequency
Passive+ prepositional phrase fragment	was explored by its, was found from the	136
Other prepositional phrases	of the present study, in the present study, in line with the	112
Noun phrase+ of	the findings of the	101
Verb phrase+ that- clause fragment	should be mentioned that, should be noted that	90
Noun phrase with other post-modifier fragments	the difference in the, the relationship between interlocutors	81
Prepositional phrase+ of	in the case of	69
Anticipatory it+ noun/adjectival phrase/ verb phrase	it is worth mentioning that, it is unlikely that, it may be argued	53
Copula be + noun phrase/ adjective phrase	is due to the	42
Adverbial clause fragment	as viewed in the, as we were familiar	38

For further analysis and comparison, the results are grouped into phrasal and clausal categories. Out of 722 LBs, 25.2% are noun-based phrasal LBs, 25.06 preposition-based phrasal LBs, and 31.99 verb-based phrasal bundles. Thus, the majority of them are phrasal. The remaining, 17.72%, are clausal.

The most frequent type of VP-based phrasal bundle is **Passive+ prepositional phrase fragment**. Two examples from the corpus are as follows:

- "According to Li (2014), this list could be called a "list of pragmatic routines" since **it was explored by** its pragmatic functions in semi- structured role-play tasks".
- "**It was found that** there was a general tendency for conventionally indirect head acts in all situations by all the participants".

The second frequent type was **other prepositional phrases**, which shows the good knowledge of the writers of prepositions and how to use them. Two examples from the corpus are:

- "This finding, **in line with the** previous literature, illustrates that explicit information about the target structure plays little role in Processing instruction".
- "The findings **of the present study** indicate that Iranian EFL learners' attitudes toward learning English play a very significant role in shaping their motivational preferences".

The third frequent type was **Noun phrase+ of** used to indicate the existence or presence of relationships.

- "Based on **the findings of the** present study, a new model of L2WTC from an ecological perspective emerged. The model is shown in Figure 3, which is a graphical **summary of the results** of this study".

The fourth frequent category was **Verb phrase+ that- clause fragment** to express an opinion.

- "**It should be mentioned that** classroom environment also influenced L2WTC indirectly through communication confidence, motivation, and attitudes".
- "However, **it must be noted that** the two aforementioned studies were conducted at the university level while our study focused on high school students".

The fifth frequent category was **Noun phrase with other post-modifier fragments** used to describe how a process occurs or to show relations among things.

- "The results of the two phases did contradict about **the difference between the** effect of background knowledge activation and related vocabulary instruction on L2 listening performance".

The sixth frequent category was **Prepositional phrase+ of** used for topic introduction, topic clarification and elaboration as well as referential expressions which are used to make direct reference to physical or abstract entities.

- "The students mostly acknowledged and confirmed their partners' proposals **in the case of** non-CCEs".
- "Regarding the efficacy **of the presence of** additional language in promoting pragmatic competence, the analysis did not delineate better performance of bilingual EFL learners".

Anticipatory it+ noun/adjectival phrase/ verb phrase was the seventh frequent category reporting possibility/likelihood, importance, and necessity.

- "**It is noteworthy that** when investigating the superior performance of the intermediate bilinguals compared to the other two groups, large effect sizes were found".
- "**It could be argued** that this finding is echoing the suggestions made by some of the researchers in the field".

Copula be + noun phrase/ adjective phrase was the eighth type which was frequent in the corpus used to identify causative relations or comparative relations.

- "This **might be due to** the differences existing between L1 and L2. It means that L2 reading is cognitively more demanding than L1 reading".

The last frequent type was **Adverbial clause fragment** used for making deictic reference to other discourse elements.

- "**As viewed in the** analyzed CCEs, the composition encouraged the students to pay more attention to various relationships between form and meaning".

Having presented the structural categories of the LBs in the applied linguistics discussions, now we present the findings regarding their functional categorization. LBs are assigned to one of the three major categories of stance expressions, discourse organizers, and referential expressions based on their functions (Table 4).

Table 4

Functions of LBs in the Discussion of Applied Linguistics RAs

Category	Subcategory	Frequency
Stance expressions	Epistemic stance	62
	Attitudinal/modality stance	98
Discourse organizers	Topic introduction/focus	89
	Topic elaboration/clarification	103
Referential expressions	Identification/focus	39
	Imprecision	28
	Specification of attributes	96
	Time/place/text reference	207

The referential expressions that most frequently occurred were those making direct references to either physical or abstract entities, along with the textual context. 370 instances of LBs were specified for this specific function.

- "**In the case of** Iranian language teachers, vicarious experiences are precious potential for self-efficacy beliefs".

Most LBs of this category play the function of text deixis.

- "As shown in the moderator analysis section, the effect size of the ISA is far larger for advanced learners than for beginners".

The role of multifunctional reference was also common. An example is as follows:

- "This section discusses the findings in three subsections according to the main research questions posed **at the beginning of** the study which concern the pragmalinguistic and sociopragmatic differences among learners with different proficiency levels".

Following referential markers, discourse organizers reflecting the relationships between parts of discourse were ranked second with 192 instances of LBs. Most of them were employed primarily for elaboration and clarification, as demonstrated in the following example.

- "**On the other hand**, a striking finding of the current study, which is at odds with conventional wisdom, is the lack of a significant difference between B.A. and M.A. teachers with regard to their overall mean scores".

However, stance expressions stating feelings, ideas, and perspectives were the least frequent group in the study with 160 LBs. Epistemic stance and attitudinal/modality stance were used in the applied linguistics discussions, with the latter being more common. The following examples illustrate their functions:

- "Iranian teachers in particular sound more emotional, and hence **are more likely to be** affected by positive and negative emotions".
- "In our study, the findings may be **related to the fact** that Baluch bilinguals as opposed to Persian monolinguals in Iran do not receive formal education and literacy in their first language".

Among the sub-categories of functions, time/place/text reference was ranked first in terms of frequency with 207 instances. An example is as follows:

- "**As reported in Table 3**, results of t- test revealed that there was a significant difference in the mean item difficulty estimates of MC and CR items ($t(51.54) = -4.399, p = .000$)".

LBs in the Discussion Section of Geology RAs

The data analysis of the geology corpus revealed 39353 types and 42100 tokens of LBs. The three most frequent LBs found were **in the study area**, **the chemical composition of**, and **according to the results**, with the frequency of 15, 13, and 8 tokens respectively accounting for 0.035%, 0.030%, and 0.019% of the total number of LBs found. 560 bundles meeting the criterion of five instances were classified according to their structural and functional features (Tables 5 and 6).

Table 5

Structural Categories of LBs in the Discussion of Geology RAs

Structures	Examples	Frequency
Other prepositional phrases	in the study area, in the magma chamber, on the other hand	106
Passive+ prepositional phrase fragment	are shown in figure, are related to areas	96
Anticipatory it+ noun/adjectival phrase/ verb phrase	it should be noted, it should be mentioned	79
Prepositional phrase+ of	as a result of, at the base of	66
Copula be + noun phrase/ adjective phrase	are due to the, is due to practice	64
Noun phrase+ of	the results of each, the findings of descriptive	48

Verb phrase+ that- clause fragment	can be comprehended that, can be proposed that	41
Noun phrase with other post-modifier fragments	the difference among them, the relationship between scientific	32
Adverbial clause fragment	as shown in figure, as shown by the	28

For further analysis, the results are grouped into phrasal and clausal categories. Out of 560 LBs, 14.28% are noun-based, 30.71 preposition-based, and 42.67 verb-based phrasal bundles. Thus, the majority of them are phrasal. The remaining 10.71 percent are clausal.

The most frequent type of VP-based category was **other prepositional phrases**. Some of the bundles of this type are used to determine a certain place or time.

- "Stratigraphical, structural and petrofabric studies **in the study area** show that three metamorphic events and four deformation phases have affected the rock units".
- "Also, the Synneusis texture is observed in the rim of the crystals, which is the result of the dynamic motion of the crystal during its growth **in the magma chamber**".

The second frequent type was **Passive+ prepositional phrase fragment**. Two examples are as follows:

- "The thickness of the lithosphere or low velocity layer in each region **is related to areas** in which areas are piercing points".
- "The relationships between the highest CBR values and soil-wastes stabilized mixtures **are shown in Figure 8**".

The third frequent type was **Anticipatory it+ noun/adjectival phrase/ verb phrase** reporting possibility/likelihood, importance, and necessity.

- "**It should be noted** that the high sulfate ion in the eastern parts is due to the presence of gypsum soils and gypsum veins in the existing fractures".

Prepositional phrase+ of, used for topic introduction and topic clarification and elaboration to make direct reference to physical or abstract units was the fourth frequent category.

- "Presence of various percentages of anorthite in microlite can be due to the recharge of basic magma **at the base of** the magma chamber on the magmatic system".
- "Secondary epic minerals mostly formed **as result of hornblende** alteration and they are seen in the margin and center of minerals".

Following prepositional phrase + of, **Copula be + noun phrase/ adjective phrase** was the next category.

- "The lower pH values of the soils in the tea-cultivated area **are due to the** intensive application of phosphorus fertilizers".

Noun phrase+ of was the sixth frequent type of LBs used to show relationships.

- "**The findings of descriptive** statistical and geo-statistical analysis of soil particle sizes distribution (sand, silt and clay) are presented in Tables 1 and 2".
- "Prosser highlighted **the importance of involvement** of local communities and stakeholders to the geomorphosite management".

The following frequent type was **Verb phrase+ that- clause fragment** used to talk about one's opinion.

- "**It can be comprehended that** after developing the CS morphology, micro-textures like FS, FOZ, GLO, and SY are developed from a complex and repeated magma process".

- "It is proposed that these dykes may have originated in a back-arc basin tectonic setting".

Noun phrase with other post-modifier fragments was the next frequent category used to explain the progression of a process or highlight the associations between distinct entities.

- "... but **the difference among them** was not significant.
- "It is a similar situation with **the relationships between scientific** and conservation values and scientific and educational values".

Finally, the least frequent type was **Adverbial clause fragment** used for deictic reference to other discourse segments.

- "the deformation patterns cannot be satisfactorily reproduced with the inverse rupture, **as expected in this area**".

Now, the results for the functional categorization of LBs in the geology RAs are presented.

Table 6

Functions of LBs in the Discussion of Geology RAs

Category	Subcategory	Frequency
Stance expressions	Epistemic stance	61
	Attitudinal/modality stance	69
Discourse organizers	Topic introduction/ focus	78
	Topic elaboration/ clarification	80
Referential expressions	Identification/focus	58
	Imprecision	44
	Specification of attributes	66
	Time/place/text reference	104

The predominant referential expressions were those that explicitly referred to physical or abstract entities, in addition to the textual context. 272 instances of LBs were specified for this specific function.

- "The reason for these changes can be related to **the size of the** magmatic chamber".

It should be pointed out that the most frequent instances played the function of time/place reference.

- "**In this framework here**, our 149.6 ± 0.5 Ma Ar-Ar dating of biotite appears somewhat younger".

Discourse organizers reflecting the connections between previous and forthcoming discourse were ranked second with 158 instances, and most instances belonged to the function of topic elaboration/clarification.

- "CAS **on the other hand** increased in its average maximum dry density after 0.5% content in the soft soil and remains almost constant at 1.5% content in the soil".

However, stance bundles which convey individual emotions, viewpoints, beliefs, confidence, and doubt. were the least frequent group in the study with 130 instances.

- "It **can be stated that** IOT displays good characteristic behavior in the soil because of its larger grain sizes in the soil sample".
- "The initial reduction in angle of internal friction for SLG stabilized soil could be attributed **to the fact that** the quantity of SLG was small and could not disperse well in the soil".

Comparison Between the Two Subcorpora

The third research is about possible significant differences in the use of LBs in the discussion section of applied linguistics and geology RAs. To answer this question, two chi-square tests were run to examine both structural and functional categories. Based on their results, the null hypothesis is accepted. About the former, the results presented in Table 7 (i.e., Sig. 0.998 > ρ . 0.05) reveal that there are no significant differences between applied linguistics and geology regarding the structures of LBs in the discussion of RA.

Table 7

Chi-square Test on Differences in Structures of LBs between Geology and Applied Linguistics RAs

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1.037	8	.998
Likelihood Ratio	1.037	8	.998
Linear-by-Linear Association	.280	1	.597
N of Valid Cases	1282		

The results of structural analysis for the subcorpora show that in both, the phrasal bundles greatly outnumber clausal ones. Moreover, among the phrasal bundles, the verb-based phrasal ones are most frequent. More specifically, in both fields, passive + prepositional phrase fragments and other prepositional phrases are the first two most frequent categories.

Similarly, the results in Table 8 show that the RAs of the two fields are not significantly different with regard to the functions of LBs in the discussion section as the significance value (0.635) is larger than ρ . 0.05.

Table 8

Chi-square Test on Differences in Functions of LBs between Geology and Applied Linguistics RAs

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.909	2	.635
Likelihood Ratio	.909	2	.635
Linear-by-Linear Association	.672	1	.412
N of Valid Cases	1282		

In both corpora, referential bundles with subcategories such as time/place reference and identification are the most frequently used bundles, followed by discourse organizers and then stance expressions.

Discussion

Many corpus studies (e.g., Biber et al., 1999; Hyland, 2008, 2012) have demonstrated that the LBs approach is highly effective in illustrating the distinctive features of various disciplines within academic writing. In fact, several interdisciplinary studies have found significant differences between soft and hard disciplines with regard to LBs use (Damchevska, 2019; Durrant, 2017; Hyland & Jiang, 2018, to name a few) and the results of this study are in line with those and add to the evidence regarding the importance of LBs for disciplinary research. LBs are recognized by language users and serve established pragmatic or discursal purposes. Consequently, the disciplinary connections suggested by LBs may reveal the deviations from conventional discourse structures (Hyland, 2012).

The findings revealed that in both subcorpora, phrasal LBs heavily outnumber clausal ones. This is in keeping with the results of many previous studies (e.g., Zare & Naseri, 2019; Zare &

Valipouri, 2021) and confirms Biber (2009)'s statement that in academic writing, phrasal bundles, instead of clausal ones, are more frequently used. Moreover, both subcorpora used VP-based phrasal bundles more frequently than the other two categories. This is partly due to the high number of passive structures in both subcorpora, which aligns with the nature of scientific writing. Yet, this is in contrast to the results of a study by Ritcher, Lotfi, and Mirzai (2022) who found noun-based and preposition-based LBs to be more frequent in the discussion section of applied linguistics articles. This contrast may be explained with reference to the criteria used in that study for the extraction of LBs; the mentioned researchers considered three occurrences as the cut-off for extraction while in this study, five occurrences was the rule. This underscores Samraj (2024)'s point argument about the importance of changes in criteria for corpus composition and extraction in research on LBs, especially in disciplinary studies.

Despite the high frequency of verb-based LBs in both subcorpora, a subtle difference was detected between the subcorpora with regard to the structural categories; in the applied linguistics discussions, the numbers of noun-based and preposition-based LBs were very close (25.2% and 25.06%, respectively), but in the geology articles, the preposition-based ones vastly outnumbered the noun-based LBs (30.71% and 14.28%, respectively). This inclination to utilize prepositional language features may be attributed to their primary function of discussing and presenting factual information, as well as explaining ideas and arguments, which is in keeping with the goal of the discussion section. Noun phrases are used to mention various elements of the research procedure, indicating either extents or qualities (Shirazizadeh & Amirfazlian, 2021). Consequently, it seems that in the discussion of the applied linguistics RAs, almost equal spaces were dedicated to referring to various sides of the research and explaining arguments while in the geology corpus, the latter was prioritized.

Considering the functional categories of LBs, the results revealed the similarity of subcorpora as in both, referential bundles were used most frequently (51.24% in the applied linguistics corpus and 48.57% in the geology corpus). This aligns with the findings of many previous studies (e.g., Adel & Erman, 2012; Appel, 2022; Biber et al., 2004; Chen & Baker, 2010; Liu & Pan, 2023) which showed most of the four-word LBs in academic writing are referential bundles because communicating facts is the priority in this kind of discourse (Goodarzi, Gholami, & Abdollahpour, 2024). Interestingly, among the referential bundles' subcategories, in both subcorpora, time/place/text references were the majority. The reasons behind the greater frequency of deixis can be related to the fact that, according to Cairns (1991), deixis is very important since it creates a certain point in space and time for participants so they do not lose the track of content; some readers face challenges with understanding the intention of the writer, and deixis will make it simpler for the writers to communicate their messages and for the readers to understand them. In the realm of hard sciences, referential bundles predominantly emphasize the physical environment, specific locations, and measurement. Conversely, in the domain of soft sciences, the focus shifts towards abstract concepts and their contextual placement within historical frameworks or processes (Durrant, 2017).

Conclusion

This study was a corpus-based analysis of the occurrence and roles of LBs within the discussion section of research articles in the fields of applied linguistics and geology. The findings showed strong disciplinary similarities with regard to both structural and functional categories. Few minor differences were detected. The dependence of both fields on the referential LBs indicates that the discussion sections of RAs within both aim to achieve a similar primary function. This function, according to the findings of this study, is writers' organizing their experiences and determining their points of view (Cortes, 2013; Shin, 2019) since "discussion sections mainly focus on arguing and drawing conclusions" (Ritcher et al., 2022, p.640). The dominance of phrasal LBs and then verb-based ones in both corpora also reveals the frequency of passive structures in scientific writing as well as the nature of this kind of discourse.

The findings of this study may promise implications for both LBs research and language teaching. Regarding the former, the findings can highlight the significance of choosing and setting

criteria for research on LBs. Concerning the latter, the study can increase our knowledge of their use, both structural and functional, in the discussion section of research articles of different disciplines. This knowledge can, in turn, be utilized in designing courses and developing materials for EAP students since the significance of LBs in academic writing is already established and differences between native speakers and non-native speakers of English in using LBs are also detected (Vaziri et al., 2023). This can be done by focusing on certain part-genres like the discussion section of RAs whose significance (Dujsik, 2013; Moyetta, 2016) and challenges have been known (Amnuai, 2017). In this regard, previous research has shown the effectiveness of the explicit instruction of LBs in academic contexts (Bagherkazemi & Rabi, 2024).

Academic writing is essential in disciplinary communities. Thus, the members must master the key academic genres. This entails having a good command of different aspects of genres including their lexico-grammatical features like LBs which foster coherence in a written work and reflect the level of proficiency a writer possesses in their discipline. Besides enhancing the quality of second language writing, formulaic sequences offer writers a tool for being communicative (Schmitt & Carter, 2004). Research on LBs and feeding the findings to language teaching, therefore, are significant for academic writing.

This study can contribute to our knowledge of academic writing and student needs. As the number of L2 writers continues to increase, becoming acquainted with the research article genre has emerged as a critical objective for students across various academic disciplines. Understanding the lexical elements pertinent to a specific genre and their application can facilitate fluency within that genre. Given that lexical bundles are specific to genres, they serve as markers of success within discourse communities. Therefore, it is essential for learners to focus on these characteristics to effectively generate a cohesive discourse.

We examined the discussion section of RAs in two disciplines. Future studies may investigate lexical bundles in under-researched genres and part-genres in other disciplines and their subdisciplines so that we would have a more complete picture of their disciplinary usage. Besides, focusing on various writer-related factors like experience and first language, which did not receive consideration in this study, can be insightful.

Acknowledgement

The authors would like to thank Professor Laurence Anthony for providing researchers worldwide with free access to his corpus analysis software.

Declaration of Conflicting Interests

There are no conflicting interests to report.

Funding Details

This study did not receive any funding.

References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and nonnative speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Anthony, L. (2007). *AntConc* (Version 3.2.3) [Computer software]. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software/AntConc>
- Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, 110, 102899. <https://doi.org/10.1016/j.system.2022.102899>

- Bagherkazemi, M., & Rabi, A. (2024). Effect of Explicit vs. Meaning-Focused Instruction on Implicit and Explicit Knowledge of Lexical Bundles: Focus on Research Articles' Discussions in Applied Linguistics. *Iranian Journal of English for Academic Purposes*, 13(3), 18-34. https://journalscmu.sinaweb.net/article_211627.html
- Bal-Gezegin, B. (2019). Lexical bundles in published research articles: A corpus-based study. *Journal of Language and Linguistic Studies*, 15(2), 520-534. <https://doi.org/10.17263/jlls.586188>
- Belcher, D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1), 1-22. <https://doi.org/10.1016/j.jslw.2006.12.001>
- Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum. <https://doi.org/10.5040/9781474212038>
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin. <https://doi.org/10.1075/scl.23>
- Biber, D. (2009). A corpus-drive approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3), 275-311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education. <https://doi.org/10.1075/z.232>
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49. <https://eric.ed.gov/?id=EJ895972>
- Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131-145). Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.9>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3, 43-57. <https://doi.org/10.3366/E17495032080000863>
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33-43. <https://doi.org/10.1016/j.jeap.2012.11.002>
- Cowie, A. (1992). Multiword lexical units and communicative language teaching. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 1-12). Macmillan. <https://doi.org/10.1007/978-1-349-12396-4>
- Curry, M. J., & Lillis, Th. (2004). Multilingual Scholars and the Imperative to Publish in English: Negotiating Interests, Demands, and Rewards. *TESOL QUARTERLY*, 38(4), 663-688. <https://doi.org/10.2307/3588284>
- Curry, M. J., & Lillis, Th. (2010). Academic research networks: Accessing resources for English medium publishing. *English for Specific Purposes*, 29, 281-295. <https://doi.org/10.1016/j.esp.2010.06.002>
- Damchevska, V. (2019). Structure of lexical bundles in economics research articles. *Journal of Teaching English for Specific and Academic Purposes*, 7(2), 225-235. <https://doi.org/10.22190/JTESAP1902225D>

- Dujsik, D. (2013). A Genre Analysis of Research Article Discussions in Applied Linguistics. *Language Research*, 49(2), 453-477. <https://hdl.handle.net/10371/86507>
- 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>
- Ferenz, O. (2005). EFL writers' social networks: Impact on advanced academic literacy development. *Journal of English for Academic Purposes*, 4, 339-351. <https://doi.org/10.1016/j.jeap.2005.07.002>
- Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL Quarterly*, 34, 127-150. <https://doi.org/10.2307/3588099>
- Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 48(1), 99-116. <https://doi.org/10.1017/S0261444813000037>
- Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing and knowing in academic philosophy*. Hillsdale, NJ: Erlbaum. <https://doi.org/10.4324/9780203812174>
- Goodarzi, R., Gholami, 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://eric.ed.gov/?id=EJ1425221>
- Hyland, K. (2008). As Can be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, 27, 1-21. <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. <https://doi.org/10.1075/ijcl.17080.hyl>
- Lillis, Th., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication*, 23(1), 3-35. <https://doi.org/10.1177/0741088305283754>
- Liu, C., & Pan, F. (2023). Connecting lexical bundles and moves in medical research articles' Methods section. *Southern African Linguistics and Applied Language Studies*, 42(1), 111-127. <https://doi.org/10.2989/16073614.2023.2226171>
- Moon, R. E. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon. <https://doi.org/10.1093/oso/9780198236146.001.0001>
- Moyetta, D. (2016). The discussion section of English and Spanish research articles in psychology: A contrastive study. *ESP Today*, 4(1), 87-106. https://esptodayjournal.org/pdf/current_issue/3.6.2016/DANIELAMOYETTA-full%20text.pdf?form=MG0AV3
- Nattinger, J. R., & Decarrico, J. S. (1992). *Lexical phrases and language teaching*. OUP. https://books.google.ca/books/about/Lexical_Phrases_and_Language_Teaching.html?id=VeB_luuoZ1wMC&redir_esc=y
- Peters, S. (2011). Asserting or deflecting expertise? Exploring the rhetorical practices of master's theses in the philosophy of education. *English for Specific Purposes*, 30, 176-185. <https://doi.org/10.1016/j.esp.2011.02.005>
- Pojanapunya, P., & Todd, R. W. (2011). Relevance of findings in results to discussion sections in applied linguistics research. In *Proceedings of the International Conference Doing Research in Applied Linguistics* (pp. 21-22). Thammasat University.

- Polo, F. J., & Varela, M. C. (2009). English for research purposes at the University of Santiago de Compostela: A survey. *Journal of English for Academic Purposes*, 8, 152-164. <https://doi.org/10.1016/j.jeap.2009.05.003>
- Richter, K. G., Lotfi Gaskaree, B., & Mirzai, 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://journals.rudn.ru/linguistics/article/view/32088>
- Safarzade, M. M., Monfared, A., & Sarfeju, M. (2013). Native and Non-native Use of Lexical Bundles in Discussion Section of Political Science Articles. *Iranian Journal of Applied Language Studies*, 5(2), 137-166. <https://doi.org/10.22111/IJALS.2015.1881>
- Samraj, (B. (2024). Disciplinary differences in lexical bundles use: A cautionary tale from methodological variations. *Journal of English for Academic Purposes*, 70, 101399. <https://doi.org/10.1016/j.jeap.2024.101399>
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.9>
- Shin, Y. K. (2019). Do native writers always have a head start over non-native writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, 40, 1-14. <https://doi.org/10.1016/j.jeap.2019.04.004>
- Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, 49, 100946. <https://doi.org/10.1016/j.jeap.2020.100946>
- Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In A. T. Upton & U. Conner (Eds.), *Discourse in the Professions. Perspectives from Corpus Linguistics* (pp. 37-64). Amsterdam and Philadelphia: John Benjamins. <https://doi.org/10.1075/sc1.16>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list (AFL). *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>
- Tardy, Ch. M. (2005). It's like a story: Rhetorical knowledge development in advanced academic literacy. *Journal of English for Academic Purposes*, 4, 325-338. <https://doi.org/10.1016/j.jeap.2005.07.005>
- Vaziri, A., Barjesteh, H., & Nasrollahi Mouziraji, A. (2023). Formulaic Sequences in Learners' Spoken English: A Comparative Corpus-based Study between Native and Non-Native Speakers of English. *Iranian Journal of English for Academic Purposes*, 12(3), 56-72. https://journalscmu.sinaweb.net/article_182279.html
- Wood, D. (2015). *Fundamentals of Formulaic Language*. London: Bloomsbury Academic. <https://doi.org/10.5040/9781474218771>
- Wray, A. (2002). *Formulaic language and the lexicon*. CUP. <https://doi.org/10.1017/cbo9780511519772>
- Zare, J., & Naseri, Z. S. (2020). Lexical bundles in English review articles. *Iranian Journal of English for Academic Purposes*, 9(1), 41-56. 20.1001.1.24763187.2020.9.1.4.1
- Zare, J., & Valipouri, L. (2021). Lexical bundles in research articles in chemistry: a structural analysis. *Iranian Journal of English for Academic Purposes*, 10(2), 90-105. 20.1001.1.24763187.2021.10.2.7.3