



P-ISSN 2355-2794
E-ISSN 2461-0275

Discourse Functions of Lexical Bundles in Indonesian EFL Learners' Argumentative Essays: A Corpus Study

Ikmi Nur Oktavianti*¹
Icuk Prayogi²

¹Department of English Education, Faculty of Teacher Training and Education,
Universitas Ahmad Dahlan, Yogyakarta 55161, INDONESIA

²Department of Indonesian Education, Faculty of Language and Arts Education,
Universitas Persatuan Guru Republik Indonesia (PGRI) Semarang, Semarang 50166,
INDONESIA

Abstract

Lexical bundles are functional units that are essential to building texts. However, lexical bundles vary due to some aspects, e.g., nativity, professionalism, text genre, etc. This study explored functional categories of lexical bundles in EFL written production, i.e., argumentative essays, focusing on 3-, 4-, and 5-word sequences employed by English major students in Indonesia. The data were taken from a learner corpus comprising 169 argumentative essays with 87.939 tokens. The lexical bundles were identified by using computerized and manual procedures. The identified bundles were then classified into functional categories and subcategories by Hyland. The results show that all functional categories were identified in the learner corpus proving the importance of those functions in students' essays. Regarding the distribution, research-oriented bundles are the most frequent bundles in the corpus, while text-oriented bundles are the least frequent. Although all functional categories were found, the structuring signals (a subcategory of text-oriented bundles) were absent in the corpus. Moreover, this study found the gap between tokens and types of each bundle function, indicating the restricted variants of bundles used by the learners. Considering the low frequency of text-oriented bundles, the absence of structuring bundles, and restricted bundles, thus the exposure of more lexical bundles serving multiple

* Corresponding author, email: ikmi.oktavianti@pbi.uad.ac.id

Citation in APA style: Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education*, 9(2), 761-783.

Received December 23, 2021; Revised March 15, 2022; Accepted April 23, 2022; Published Online May 23, 2022

<https://doi.org/10.24815/siele.v9i2.23995>

functions in writing materials is necessary, especially bundles used for organizing and structuring texts.

Keywords: Functional category, learner corpus, lexical bundles, students' essays.

1. INTRODUCTION

Lexical bundles are the combinations of three, four, or more words that are repeatedly used in discourse based on the genre or register (Cortes, 2004; Hyland, 2012). Biber et al. (1999) stated that lexical bundles are unrelated to idiomaticity or structural status, which means that lexical bundles are not meaning units; instead, they are functional units to build up and characterize a discourse (Wood, 2015). As proven by Biber et al. (1999), 30% of the words in conversations and 21% of words in academic writing are in the form of lexical bundles. In addition, Erman and Warren (2000) showed that 58.6% of spoken English discourse and 52.3% of written English discourse consist of lexical combinations. van Lancker-Sidtis and Rallon (2004) also investigated language use in the screenplay of a classic American film, 'Some Like it Hot', and figured out that 25% of the dialogues were formulaic. These findings emphasize that language is inevitably formulaic.

Lexical bundles, compared to other formulaic language concepts, are the most widely studied (Cortes, 2004) because these word sequences are frequent and measurable (Biber et al., 2004; Lee, 2020), and their occurrence is pervasive in discourse due to their vital role (Wood, 2015). Besides, the use of lexical bundles in the EFL context is crucial because they belong to the element of nativelike linguistic knowledge (Meunier, 2012). Lexical sequences, including lexical bundles, are necessary for language fluency since using correct word combinations is part of good speaking and writing (Salazar, 2011). This is because there is a tendency to use typical combinations despite the infinite potential of linguistic constructions in using language (Wood, 2015). In other words, the ability to use lexical bundles is a salient aspect to be proficient in the target language (Lee & Kim, 2017). Thus, considering the importance of lexical bundles, they should be introduced rigorously to the students, especially those in the ESL and EFL context.

Two aspects that should be concerned related to lexical bundles are the structure and the functional categories. Many studies have focused on the structure of the lexical bundles (Oktavianti & Sarage, 2021), which show that particular bundle structures might be used more often in a specific register or text genre. However, because lexical bundles are functional units that serve as the building blocks of the discourse, it is necessary to explore the functional aspects of lexical bundles. Previous studies have demonstrated that the functions of lexical bundles vary across different registers, i.e., spoken and written registers (Biber et al., 2004), text genres (Yang, 2017), and disciplines (Kwary et al., 2017; Ren, 2021). In addition, lexical bundles might have different usage tendencies related to the author's professionalism and nativity (Chen & Baker, 2010; Fajri et al., 2020). Therefore, it is important to explore the functional categories of lexical bundles comprehensively.

In the academic context, lexical bundles are pervasive because they are used at high frequency (Hyland, 2008, 2012; Hyland & Jiang, 2018). Furthermore, Gray

(2015) includes lexical bundles as the prominent features of the academic genre, which can vary across different fields. Some factors, such as professional level and nativity, hold vital roles related to the use of lexical bundles. Several studies have claimed the discrepancies in lexical bundle usage between L1 and L2 writers (Pan et al., 2016; Salazar, 2011, 2014; Yakut et al., 2021) and between professional and novice writers (Cortes, 2004). Considering that nativity and professionalism affect the usage of lexical bundles, English L2 and FL writers need to understand how lexical bundles are used in their writing to show their writing proficiency.

Previous studies on lexical bundles in the EFL academic context have mainly focused on theses or dissertations (Fitrianasari et al., 2018; Yakut et al., 2021), published articles (Pan et al., 2016; Salazar, 2014) or books (Alquraishi, 2014; Hussain et al., 2021). These studies are important to show the academic language mastery of the learners or to identify particular bundles in specific fields. However, it is also important to study the initial stage of undergraduate academic writing, e.g., students' argumentative essays.

An argumentative essay is a kind of academic prose commonly written by undergraduate students (Shin, 2018; Wingate, 2012). This writing belongs to the academic genre because writing an argumentative essay needs logical and critical thinking and the capability to connect arguments and evidence coherently (Parkinson & Musgrave, 2014). This genre was chosen because lexical bundles are more frequently used in argumentative texts than the other genre, e.g., narrative texts (Yang, 2017). In addition, some research has demonstrated that academic writing, including argumentative text, is packed with multi-word combinations (Gray, 2015; Hyland & Jiang, 2018). Regarding the aforementioned rationale, the investigation of lexical bundles in argumentative essays should be done more intensively. Common writing at the undergraduate level is an essential text type that is important for novice writers (Wingate, 2012). The investigation results can contribute to understanding the learners' writing proficiency and identify the learners' formulaic sequence knowledge, which is prominent for language mastery. However, only a few studies have been about lexical bundles in early undergraduate writing (Ädel & Erman, 2012; Bychkovska & Lee, 2017; Yang, 2017). These studies emphasized the importance of investigating lexical bundles in a more initial stage of academic writing to understand students' writing (Bychkovska & Lee, 2017) and increase language fluency and accuracy (Yang & Fang, 2021).

In the Indonesian context, lexical bundles have not been widely discussed. Most research on lexical bundles has been conducted to analyze the structure (Oktavianti & Sarage, 2021) and to examine lexical bundles theses/dissertations (Fitrianasari et al., 2018; Wachidah et al., 2020) and research articles (Budiwiyanto & Suhardijanto, 2020; Kwary et al., 2017). Only a few studies explored the functional categories of lexical bundles (Budiwiyanto & Suhardijanto, 2020; Wachidah et al., 2020), and these studies did not analyze students' essays; little is known about the discourse functions in the early academic writing stage which is crucial to find out the learners' development and to identify what should be improved or emphasized in the teaching materials. Therefore, this study aims to contribute to a better understanding of lexical bundles and their functional categories as well as provide educators and Indonesian EFL students with an insight into the effective use of lexical bundles in argumentative writing. Furthermore, this study is a corpus study that provides larger data and a more

accurate investigation of lexical bundles since lexical bundles are computer-generated (Salazar, 2014). Particularly, this research attempts to answer these questions:

- What are the functional categories of lexical bundles identified in the students' essays?
- How are the functional categories distributed in the students' essays?

2. LITERATURE REVIEW

2.1 Lexical Bundles

The lexical bundle was first proposed by Biber et al. (1999) when describing American and British English grammar using the corpus approach in their book. They successfully identified word sequences frequently occurring in natural language as the building blocks of discourse. Lexical bundles refer to the word sequences (three or more sequences) that co-occur syntactically and usually characterize particular types of discourse (Biber et al., 1999; Cortes, 2004). The main procedure for studying lexical bundles is the identification process. Lexical bundles are not predefined linguistic units having multi-word combinations. Instead, they have an empirical basis by relying on the frequency criteria (Salazar, 2014; Wood, 2015). Word sequences can be recognized as lexical bundles with a particular cutoff frequency. However, frequency is not adequate and should be followed up by looking at the distribution of the bundles across texts to avoid individual idiosyncrasies (Lee, 2020; Salazar, 2014).

Another prominent characteristic of lexical bundles is their fixed structure (Cortes, 2004; Salazar, 2014). This is due to the characteristic of lexical bundles as computer-generated bundles, so they are strongly related as structure, not as a semantic unit. The structure of lexical bundles consists of sequences with embedded fragments, and most of them are not structural units. Therefore, most lexical bundles are not idiomatic since the meanings are fully retrievable from the individual words (Cortes, 2004; Wood, 2015). Although lexical bundles are associated with recurrent expressions in natural language, there is the assumption that they will be acquired easily. However, Wood (2015) underlined that acquiring lexical bundles does not happen naturally.

2.2 Functions of Lexical Bundles

Previous studies have explored lexical bundles' functional categories in different registers (spoken vs. written) and enormous genres (academic vs. narrative). Biber et al. (2004) proposed a functional taxonomy to investigate bundles' functions in spoken and written academic registers by classifying bundles into four primary functions, namely stance expressions (i.e., to express the attitude of the writer or certainty toward a proposition), discourse organizers (i.e., to organize discourse), referential bundles (i.e., to refer to any entities or experiences), and special conversation functions (e.g., politeness). For more general use, Biber et al.'s (2004) taxonomy cover lexical bundles in spoken academic register and textbooks with some classifications that are particularly assigned for spoken language, namely, special conversational functions (e.g., politeness and simple inquiry). In addition, the discourse organizers focus on the topic introduction and topic elaboration, which are more relevant in spoken language.

Later, this classification was developed by [Simpson-Vlach and Ellis \(2010\)](#) by adding some new categories and subcategories, e.g., meta-discourse, textual reference, cause and effect, and discourse markers that belong to the discourse organizer category. However, Simpson-Vlach and Ellis's taxonomy is also more suitable for spoken academic registers ([Liu & Chen, 2020](#)).

Another perspective of lexical bundle functions was proposed by [Hyland \(2008\)](#). [Hyland \(2008\)](#) introduced an alternative taxonomy that focuses on the characteristics of written language. In this taxonomy, Hyland classifies the bundles into research-oriented bundles, text-oriented bundles, and participant-oriented bundles. The present study employs the functional taxonomy by [Hyland \(2008\)](#) because it is more suitable for written registers, especially for academic prose ([Liu & Chen, 2020](#)). For example, some previous studies employed Hyland's functional category, e.g., [Fajri et al. \(2020\)](#) and [Yakut et al. \(2021\)](#). [Yang and Fang \(2021\)](#) also used this framework to analyze students' essays in their studies. Table 1 presents the details of Hyland's taxonomy, along with some examples of the lexical bundles.

Table 1. Functional categories and subcategories.

Functional category	Explanation	Subcategory	Example
Research-oriented	To refer to entity, activity, and experience	Location	'in the present study', 'at the beginning'
		Procedure	'the use of', 'the role of', 'the purpose of'
		Quantification	'one of the', 'the magnitude of'
		Description	'the structure of', 'the surface of'
		Topic	'the currency board system', 'in the Hongkong'
Text-oriented	To organize text and its meanings	Transition signals (e.g., addition, contrast)	'in contrast to', 'in addition to', 'on the other hand'
		Resultative signals (i.e., inferential or causative relations)	'these results suggest', 'the results show', 'as a result'
		Structuring signals (organize parts of discourse and direct readers to certain parts of the texts)	'as shown in', 'in the next section'
		Framing signals (state argument, limited condition, etc.)	'with respect to the', 'in regard to', 'with the exception of'
Participant-oriented	To exhibit writers' positions and address readers	Stance features	'may be due', 'it is possible'
		Engagement features	'as can be seen', 'it should be'

Because some research used [Biber et al.'s \(2004\)](#) functional taxonomy and other studies used [Hyland's \(2008\)](#) functional taxonomy, this research frequently utilized

the terms from both taxonomies. Regarding the nature, Biber et al.'s (2004) category and Hyland's category share similar concepts related to the functions of bundles in discourse. In the first category, research-oriented and referential bundles share a similar aim, i.e., to attribute entities in the real world. Meanwhile, participant-oriented and stance bundles are related to speakers' attitudes and the reader's engagement. Lastly, text-oriented and discourse organizer bundles help organize the texts. Thus, although this study employs Hyland's (2008) functional category, the functional category terms proposed by Biber et al. (2004) are not replaced and are perceived as relatively and simply equal to Hyland's category.

2.3 Previous Studies

There have been some studies about functional categories of lexical bundles in academic prose (e.g., Pan et al., 2016; Yakut et al., 2021). Some concerns have been addressed regarding this subject, such as the nativity (native vs. non-native writers), professionalism (professional vs. novice writers), and academic prose genre (e.g., research articles, theses/dissertations, essays). Regarding the nativity (L1 vs. L2, FL), Salazar (2014) examined lexical bundles in the biomedical field written by native vs. non-native English writers. Salazar's (2014) study showed that non-native writers overuse particular bundles, resulting in repetitiveness and lack of variation. Regarding lexical bundles' functions, the L2 writers use participant-oriented bundles restrictedly compared to the other functional categories, which indicates a lack of awareness of this bundle function and lack of exposure to the bundles. In line with Salazar, Pan et al. (2016) also identified the discrepancies in lexical bundle use between L1 and L2 English writers and found that the proportions of functional distribution are relatively similar. However, L2 professional writers utilize fewer research-oriented bundles and more stance-oriented bundles, while L1 professionals use more research-oriented ones. Moreover, some discrepancies in lexical bundle use were found between L1 and L2 English professional writers.

Different from Salazar (2014) and Pan et al. (2016), Bychkovska and Lee (2017) have claimed the different lexical bundle distribution between L1 and L2 English learners. The study revealed that L2 students more frequently use stance bundles in their argumentative essays, and there were some misuses of bundles by the L2 students, which led to some pedagogical suggestions. Similar to Bychkovska and Lee (2017), Shin (2018) proved the same proportion of lexical bundles between L1 and L2 students' argumentative essays. The study showed the domination of stance expressions as many as 47.9% in the native corpus and 45.5% in the non-native corpus, followed by referential expressions (41.1% in the native corpus, and 38.5% in the non-native corpus), and discourse organizers (10.9% in native corpus, and 15.4% in non-native corpus). Yakut et al. (2021) also demonstrated that both L1 and L2 English writers tend to use the same bundle. Both groups used more text-oriented bundles than research-oriented bundles in dissertation writing and also used fewer participant-oriented bundles. In other words, they used more stance bundles than engagement bundles.

Although the comparison of bundles in L1 and L2 writers are intriguing, some studies also solely focus on foreign language (FL) learners (e.g., Fitrianasari et al., 2018; Yang & Fang, 2021). Fitrianasari et al. (2018) investigated lexical bundles in EFL students' theses and found that undergraduate students more frequently use

research-oriented bundles and text-oriented bundles are more common in graduate students' theses. Later, [Yang and Fang, \(2021\)](#) also analyzed EFL students' essays in China and showed that research-oriented bundles are the most frequently used bundles concerning the type and frequency, followed by participant-oriented and text-oriented bundles. Both [Fitrianasari et al. \(2018\)](#) and [Yang and Fang \(2021\)](#) employed Hyland's functional categories to classify lexical bundles' functions. [Fajri et al. \(2020\)](#) conducted a more recent study contrasting lexical bundles produced by L1 English professional writers vs. L2 English professional writers. The study revealed the different functional categories found in both groups since L2 writers employ fewer quantification bundles than L1 writers. Both groups also utilize the text-oriented bundles more significantly (more than 50%). A different framework was employed by [Sadeghi \(2015\)](#). [Sadeghi \(2015\)](#) used the functional categories by [Biber et al. \(2004\)](#) to examine the EFL students' theses. According to the findings, the most frequent function is the referential expression (50.8%), organizing bundles (41.5%), and stance bundles (7.6%). Referential bundles in the students' theses are very frequent and significant. Regarding the proficiency of foreign language learners, [Chen and Baker \(2014\)](#) demonstrated similar distribution and patterns of bundles used across learners' proficiency levels.

Lexical bundle studies have also focused on the professionalism aspect. For example, [Chen and Baker \(2010\)](#) contrast L1 English professional writers with L1 English students. They examined lexical bundles in two corpora: The British Academic Written English (BAWE) corpus and the FLOB corpus. The study showed some differences regarding lexical bundles in L1 English professional vs. novice writers corpora emphasizing the structures and functions. For the structure, VP-based bundles outnumber the other structural categories in the native students, while NP-based bundles are the most commonly used by L1 professionals. As for the functions, native students use more discourse organizer bundles, while native professionals utilize referential bundles more recurrently. It is noticeable that advanced learners are plausible to achieve native-like writing competence.

Lexical bundle research has also been done by considering the disciplines (e.g., [Pan et al., 2016](#); [Ren, 2021](#)). [Ren \(2021\)](#) explored lexical bundles in research papers published in two fields: applied linguistics and pharmaceutical science. The study demonstrated the difference in lexical bundle functions in both disciplines because referential expressions are the most prevalent in applied linguistics articles, while stance expressions are the most dominant in pharmaceutical science articles. We can notice the different functional categories used in both fields, which exhibit the prominence of disciplines in lexical bundle distribution. Concerning a particular field, [Nasrabad et al. \(2020\)](#) discovered some new functional categories of lexical bundles used in applied linguistics published papers that were not classified in the existing functional taxonomies. Based on the findings of those studies, variations of lexical bundles are not only subject to different fields, but they might also occur within a discipline marking the linguistic characteristics of the discipline.

Based on those previous studies, it is then evident that some factors affect the use of lexical bundles in academic writing, such as nativity (L1, L2, FL learners), professionalism (professional vs. novice writers), and disciplines. Thus, researching lexical bundles is of salience and interest, especially in those areas less covered. Those studies primarily focused on published articles or theses or dissertations, university students' essays as the more initial stage of academic writing mastery are somewhat

overlooked. This study then focuses on EFL novice writers' state, i.e., university students' argumentative essays.

3. METHODS

This study employed a corpus approach since it compiled a learner corpus of students' essays, used a corpus tool, and conducted corpus analyses. Below are the detailed descriptions of the methods.

3.1 Learner Corpus Design

The corpus was compiled from the students' argumentative essays assigned in the Writing in Academic Context course for fourth-semester students of an English education department at one of the universities in Indonesia. The participants of this research were required to have taken several writing courses in the department, namely Paragraph Writing, Essay Writing, and Writing in Professional Context. The essays were collected from the participants by conducting some writing tasks related to argumentative essays. The design of the writing task addressed several issues on the design of learner corpus as proposed by [Granger \(1998, 2008\)](#). There are learner variables and task variables, as in Table 2, which were necessary for selecting the participants and designing the writing task.

Table 2. Learner corpus variables.

Variables of learners		Variables of task	
General variables	Specific variables	General variables	Specific variables
Age: 18-20-year-old	Learning context: English as a foreign language	Written text	Type of text: argumentative essay
Sex: Female and male students	Proficiency level: intermediate	Topic: education, technology	Setting: timed
Country: Indonesia	FL: English	Genre: academic	Number of words: 500-1000 words
L1: local language(s) or Indonesian L2: Indonesian			

In addition to the variables in Table 2, some features related to the learners should also be concerned for constructing the corpus. Table 2 shows the shared features of the learners (i.e., the participants of this study), such as age, learning context, region, medium, the genre of the text, and task setting. Besides, there are the variable features of the learners, e.g., sex and L1 backgrounds.

3.2 Corpus Tool

A corpus tool is necessary for corpus-driven and corpus-based research since it enables computerized procedures to collect and analyze data. The corpus tool selected for this study is LanCSBox ([Brezina et al., 2020](#)) because it can read numerous file formats and has the needed features to identify bundles and conduct analysis (*n*-gram,

frequency, and dispersion measurement). In addition, it works well with the English language data, so the learner corpus of this study is in an unannotated format.

3.3 Data Collection

3.3.1 *Collecting the essays*

The data collection procedure involved distributing the writing task to the students. They were assigned to write argumentative essays covering one of the following topics: (1) digital minimalism helps students stay focused, (2) women should not focus on higher education, (3) online learning is more effective than offline learning, and (4) becoming viral is an important goal for millennials. The online writing task was timed to be completed within three days ranging from 500 to 1000 words. The submission was done on Google Drive for a more accessible procedure for the researchers and the participants. After the submission, as many as 169 essays were compiled into a learner corpus with 87.939 tokens.

3.3.2 *Identifying lexical bundles*

The identification stage followed the corpus building as part of the data collection. Since lexical bundles are identified empirically (Cortes, 2004), there should be clear criteria for the identification. There are two identification steps in the present study, i.e., the computerized procedure and the manual procedure.

The computerized procedure was done with the assistance of a corpus tool, LanksBox. This present study explored the lexical bundles containing 3- to 5-word sequences due to several rationales: the size range is the most researched length, and it is a manageable size for both computerized and manual procedures (Lee, 2020). Many studies have focused on 3-word bundles because they display a broad range of productive expressions; thus, they are included in this study. As for the 4-word bundles, they are still manageable and more commonly produced than 5-word bundles. Besides, 4-word bundles have more apparent structures (3-word bundles are usually embedded in them) and have more evident functions (Hyland, 2008). This study also included 5-word bundles to yield a larger size of lexical bundles for the analysis to complete the investigation, as has been done by Gil and Caro (2019).

Since lexical bundles are corpus-driven procedures, some criteria are applied for the identification stage. The requirements include frequency cutoff and dispersion threshold, both of which were processed using LanksBox. The main criterion for lexical bundle identification is frequency. The commonly used cutoff frequency is ten times per million words, but the frequency can be adjusted from two to ten times for a smaller corpus (Hyland, 2012; Lee, 2020). The present study employed ten occurrences to identify the 3- and 4-word sequences. However, the cutoff frequency was changed to five hits in the corpus because 5-word bundles contain more word combinations, decreasing the frequency of use (Cortes, 2013).

The second criterion is the dispersion threshold, which refers to the distribution of bundles in multiple texts in the corpus. Many earlier studies set the occurrence of the bundles in five texts in the corpus as the dispersion threshold, but the present study utilizes a statistical measurement to calculate the dispersion. Several statistical measurements for dispersions of lexical bundles are ‘standard deviation’, ‘variation

coefficient', 'Juilland et al.'s, and some others (Gries, 2008). This study employed Gries' DP for the dispersion threshold since it is suitable for a corpus comprising many texts having different sizes (Biber et al., 2016). The dispersion values are unitless, ranging from 0 (minimum value) to 1 (maximum value). Zero value shows that the lexical bundles are only used in one corpus text, and the maximum values indicate that the lexical bundles are used in multiple corpus texts (Burch et al., 2017).

As for the following procedure, there is manual identification. The manual identification of lexical bundles was done by confirming the bundles with the following criteria: (1) the fulfillment of certain discourse functions, (2) the compositional meanings of the bundles (meaningless and functionless bundles were excluded), (3) the exclusion of proper nouns, and (94) the exclusion of free combinations (Biber et al., 2004; Hyland, 2012; Lee, 2020). Utilizing computerized and manual procedures, the identified bundles are accurate and valid for the analysis. Each author worked independently to identify the bundles manually, and the inter-rater reliability was 95%. The discrepancies were then discussed to reach 100% agreement based on their contexts.

3.4 Data Analysis

The data that have been collected were classified based on the functional categories and subcategories adopted from Hyland (2008). This study employed Hyland's functional taxonomy by classifying lexical bundles into research-oriented, text-oriented, and participant-oriented categories. The bundles were classified based on Hyland's functional categories independently by each author and discussed to meet the agreement. However, each bundle might occupy more than one function in various contexts, e.g., 'at the end of' can serve as a time or place reference (Biber et al., 2004). An attempt to overcome this issue is by analyzing the concordances of potentially multifunction bundles and categorizing them based on their typical use (Salazar, 2014). After each bundle has been assigned to its functional category, this study investigated the frequency of category and subcategory. The findings of types and frequency of functional categories were then interpreted and connected to the previous studies to discuss and elaborate on the pedagogical implications.

4. RESULTS

4.1 The Results of Lexical Bundles Identification

This section presents the results of lexical bundles identification. This study has identified 3-, 4-, and 5-word bundles, as displayed in Table 3. Since frequency and dispersion threshold are the criteria for inclusion, the information is also included in Table 3.

Table 3. List of 3-word bundles.

3-word bundles	Frequency	Dispersion	3-word bundles	Frequency	Dispersion
'the quality of'	43	0.78	'the use of'	105	0.62
'the spread of'	34	0.84	'a lot of'	96	0.55
'the development of'	33	0.79	'we have to'	21	0.89

Table 3 continued...

'the importance of'	19	0.91	'and many more'	20	0.90
'the existence of'	14	0.91	'as well as'	20	0.90
'one of them'	14	0.90	'I agree with'	18	0.88
'the number of'	14	0.93	'as long as'	17	0.89
'the most important'	10	0.93	'as a result'	14	0.91
'be able to'	61	0.68	'not only that'	10	0.93
'take advantage of'	18	0.88	'it will be'	25	0.85
'according to the'	14	0.90	'there are also'	23	0.88
'take care of'	12	0.93	'there is no'	21	0.91
'stay focused on'	18	0.91	'there are some'	18	0.90
'to sum up'	10	0.93	'and so on'	18	0.90
'in the future'	65	0.73	'as we know'	15	0.92
'in my opinion'	39	0.80	'various kinds of'	16	0.92
'in addition to'	28	0.84	'around the world'	10	0.94
'in order to'	20	0.88	'due to the'	10	0.92
'in the world'	24	0.89			

As many as 37 bundles were identified in 3-word combination, ranging from the most frequently used 'the use of' (with 105 occurrences) into the least frequent bundles 'to sum up', 'not only that', 'due to the', 'the most important', and 'around the world' (with 10 occurrences). This study also identified 4- and 5-word bundles, presented in Table 4.

Table 4. List of 4- and 5-word bundles.

4-word bundles	Frequency	Dispersion	5-word bundles	Frequency	Dispersion
'one of the important'	30	0.85	'most value to your life'	8	0.94
'must be able to'	16	0.90	'will not be able to'	5	0.95
'that can be done'	16	0.93	'be known by many people'	6	0.96
'in the form of'	15	0.91	'to be seen and recognized'	5	0.97
'at the same time'	10	0.93	'we must be able to'	6	0.95
'as efficiently as possible'	12	0.93	'there are still many people'	5	0.96
'a lot of people'	11	0.95			
'if you want to'	10	0.94			

Table 4 shows eight bundles belonging to the 4-word combinations and six bundles belonging to the 5-word category. These numbers are significantly smaller than the number of 3-word bundles because they have longer word sequences which restrict the number of occurrences. Overall, 51 bundles (51 types) were identified in the learner corpus with 1122 tokens of bundles.

4.2 Functions of Bundles

Those identified bundles were categorized based on their functional categories and subcategories. In general, all functional categories and subcategories can be found in the learner corpus. Table 5 presents the complete classification of the bundles.

Table 5. Functional category.

Functional category	Subcategory	Lexical bundles
Research-oriented	Location	'in the future', 'at the same time', 'in the world', 'around the world'
	Procedure	'the development of', 'take advantage of', 'the use of', 'as efficiently as possible', 'that can be done'
	Quantification	'one of them', 'the number of', 'a lot of', 'and many more', 'various kinds of', 'a lot of people', 'there are still many people', 'one of the important'
	Description	'the quality of', 'the spread of', 'the importance of', 'the existence of', 'the most important', 'in the form of', 'most value to your life', 'to be seen and recognized', 'be known by many people'
	Topic	'take care of', 'stay focused on'
Text-oriented	Transition signals	'to sum up', 'in addition to', 'in order to', 'not only that'
	Resultative signals	'as a result'
	Structuring signals	-
	Framing signals	'according to the'
Participant-oriented	Stance features	'be able to', 'in my opinion', 'I agree with', 'will not be able to', 'must be able to'
	Engagement features	'we have to', 'as we know', 'if you want to', 'we must be able to'

For a more rigorous discussion, each subfunction of research-oriented, text-oriented, and participant-oriented is presented with examples from the learner corpus in the next sub-sections.

4.2.1 Research-oriented

Several research-oriented bundles were identified in the learner corpus. They refer to entities, activities, and experiences in the real world to be described in texts. All subfunctions of research-oriented bundles are presented below.

a) Location

Referential bundles indicate time/place in the writing, such as 'in the future' and 'at the same time' that were found in the corpus.

- (1) ...so, I hope that in the future this kind of learning can continue...
But at the same time, this era also has some negative sides...

b) Procedure

Procedure bundles mark the procedure, step, or implementation of something. There were 'the use of' and 'the development of' as the examples of procedure bundles.

- (2) The use of gadgets is very necessary from the smallest thing to the urgent needs such as work.
The development of technology in the digital direction is currently growing rapidly.

c) *Quantification*

These bundles are to quantify something being described. There were some variants of quantification of bundles identified in the corpus, e.g., ‘the number of’ and ‘various kinds of’.

- (3) The photo and video sharing pictures on Instagram make users try to display photos and videos as good as possible so that people are interested, and it can increase the number of followers. Various kinds of efforts have been made by millennial circles just to go viral.

d) *Description*

Description bundles are to describe or give attributes to something. Some variants found in the corpus were ‘the quality of’ and ‘the existence’.

- (4) ...it is clear that higher education is very important for women because it will greatly affect the quality of their lives in the future.
But the existence of the Covid-19 pandemic has transformed the education system into an online system.

e) *Topic*

Some topic bundles were found in the corpus, such as ‘take care of’ and ‘stay focused on’. Topic bundles are associated with the essay topic or related to the field of study being written.

- (5) People are asked to remain calm and always take care of their health to fight this virus and to maintain their body’s immune health.
...even though students are forced to study after playing on smartphones with their tired eyes, they will sleepy and feel hard to stay focused on studying.

4.2.2 *Text-oriented*

These bundles are utilized to organize text and its meaning to deliver the content of the text. Transition bundles, resultative bundles, and framing bundles were found in the corpus, but surprisingly, no structuring bundle was identified. Each subfunction is presented below.

a) *Transition signals*

The bundles aim to show addition or contrast between elements, e.g., ‘in addition to’ that was found in the learner corpus.

- (6) In addition to signal constraints, online learning causes a lack of understanding of the material presented.
...everyone must adapt and be patient in order to overcome this crisis together.

b) *Resultative signals*

These bundles mark inferential or causal relationships, as in ‘as a result’ that is found in the learner corpus.

- (7) As a result, our learning activities become more efficient.

c) *Framing signals*

Framing bundles attempt to frame ideas/opinions based on something/someone, and situate arguments, e.g., 'according to the' as found in the learner corpus.

- (8) According to the data in the last five years, the use of Information and Communication Technology (ICT) in Indonesia has shown rapid development.

4.2.3 Participant-oriented

The other bundle functional category is the participant-oriented bundle that focuses on the texts' writers and readers. Both stance and engagement bundles were found in the learner corpus and are exemplified below.

a) *Stance features*

These bundles are to express the writer's attitudes and evaluations. One of the examples in the learner corpus was 'in my opinion'.

- (9) In my opinion, online learning has some advantages and disadvantages.

b) *Engagement features*

These bundles address readers directly. Some examples found in the learner corpus was 'if you want to'.

- (9) If you want to be a content creator then you should be a content creator who can be a good role model for people.

4.3 Distribution of the Functional Categories

Lexical bundles were also examined based on the frequency of use in the learner corpus. Table 6 presents the bundle usage proportion.

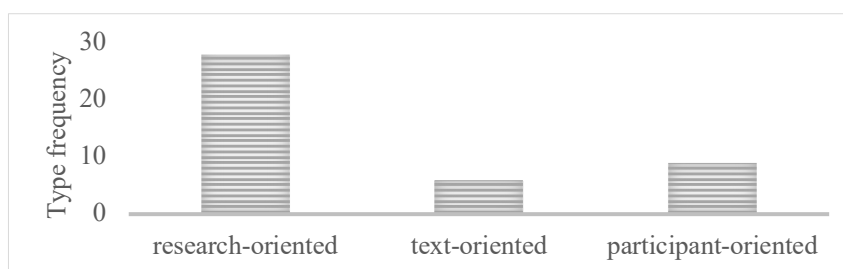
Table 6. Frequency of lexical bundle usage.

Functional category	Type		Token	
	Raw frequency	%	Raw frequency	%
Research-oriented				
Location	4	9.30	109	12.39
Procedure	5	11.63	184	20.91
Quantification	8	18.60	206	23.41
Description	9	20.93	154	17.50
Topic	2	4.65	30	3.41
	28	65.11	683	77.62
Text-oriented				
Transition signals	4	9.30	30	3.41
Resultative signals	1	2.33	14	1.59
Structuring signals	0	0.00	0	0
Framing signals	1	2.33	14	1.59
	6	13.96	58	6.59

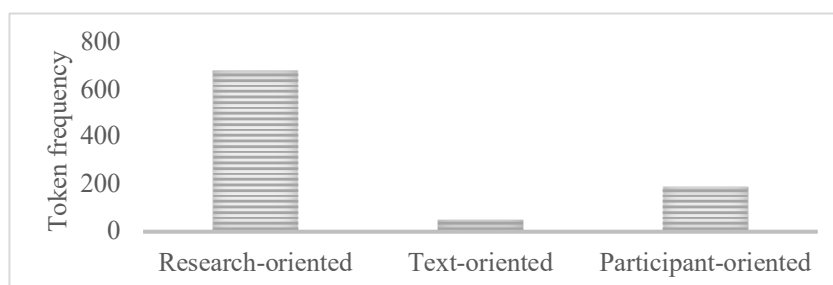
Table 6 continued...

Participant-oriented				
Stance features	5	11.63	139	15.80
Engagement	4	9.30	52	12.39
	9	20.93	191	28.19

Table 6 shows the type and token frequency of the functional categories and subcategories, except for the structuring signal that is not found in the corpus. The type frequency showed that research-oriented was the most frequent type which is plausible because this category has more subcategories than the others. The second most frequent type was participant-oriented, while the least frequent type was text-oriented, as displayed in Figure 1.

**Figure 1.** Type distribution of functional categories.

As for the token frequency, there are linear results with the type frequency, showing that research-oriented bundles were the most frequently used, followed by participant-oriented bundles and text-oriented bundles. The frequency comparison is shown in Figure 2.

**Figure 2.** Token distribution of functional categories.

Based on the type and token distribution of functional categories, a similar state between the variants and the general use was noticed. Research-oriented bundles had the most variants of bundles, and they were also the most frequently used. Meanwhile, the participant-oriented and text-oriented bundles were the second most and the least number of bundle variants and the second most and the least frequently used. As for the subfunctions, description bundles (research-oriented) had the most variants compared to other subfunctions in the same functional category and across functional categories, followed by quantification bundles and procedure bundles.

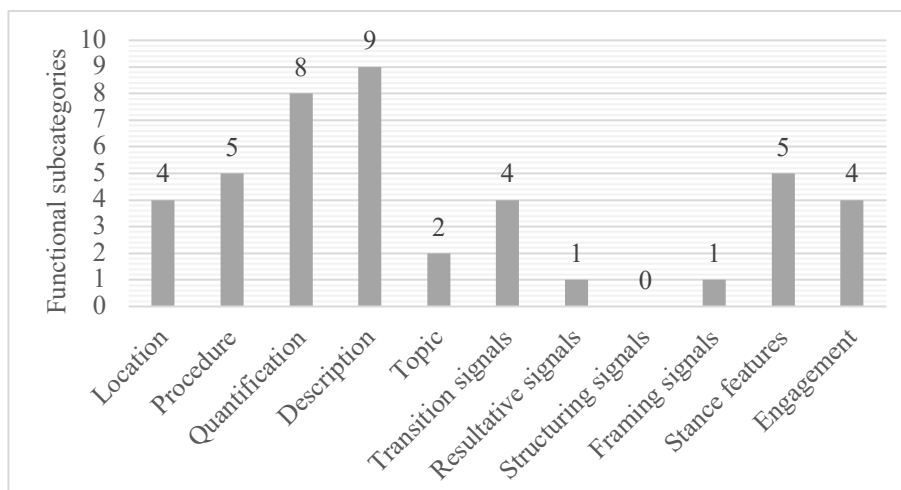


Figure 3. Type distribution of functional subcategories.

Regarding the token distribution, there were some differences in the functional categories. The most frequently used subfunction was quantification (23.41%), followed by procedure (20.91%) and description (17.50). Although there were more description bundles, they were not used more frequently than quantification bundles and procedure bundles. However, the three of them still belong to the same functional category, i.e., research-oriented bundles.

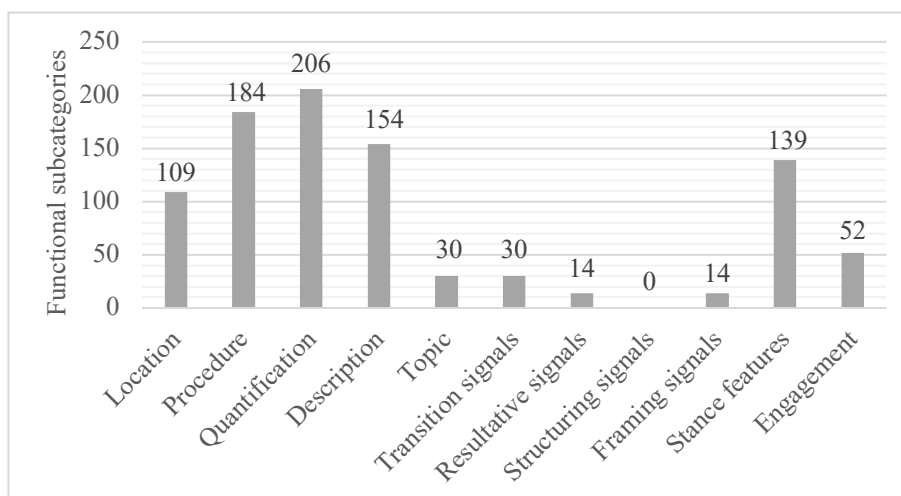


Figure 4. Token distribution of functional subcategories.

5. DISCUSSION

This study revealed that, despite all functional categories being identified in the corpus, the use of research-oriented bundles outnumbers the other functional categories and text-oriented bundles were underused. The findings of all function categories are relevant to many other studies with relatively similar results (Fajri et al., 2020; Salazar, 2014). It indicates that these three functions are fundamental in constructing a discourse, especially in academic prose. However, what should be

emphasized is the distribution of the bundles in enormous texts depending on nativity, professionalism, and proficiency.

In terms of frequency, this study demonstrated the most frequent bundles are research-oriented bundles related to the substantial aspects of the discourse (e.g., time, location, description). This result corresponds to several previous studies (Biber et al., 2004; Sadeghi, 2015; Yang & Fang, 2021), emphasizing the use of research-oriented bundles (or referential bundles) in academic contexts. The results confirmed that these bundle types are prominent and salient in academic prose. However, Pan et al. (2016) stated that L1 professionals utilize more research-oriented bundles than L2 professionals showing the discrepancies that might occur due to the nativity of the writers. Interestingly, a different result was demonstrated by Salazar (2014), claiming that text-oriented bundles are the most frequently used, in contrast to the present study's finding. Like Salazar (2014), Wachidah et al. (2020) also demonstrated the high frequency of text-oriented bundles used in the results and discussion sections. It might happen due to the learners' proficiency levels, the disciplines, and the different needs of specific academic paper sections (i.e., findings and discussion). Another contrasting result was reported by Bychkovska and Lee (2017), with stance bundles as the most commonly used bundles and referential bundles as the least frequently used ones.

As for the second most frequent function, there were participant-oriented bundles. This result is in accordance with Liu and Chen (2020) and Yang and Fang (2021), who also identified participant-oriented bundles as the second most frequent, followed by text-oriented bundles as the least frequent bundles. Liu and Chen (2020) proved that referential bundles and stance bundles are prominent in academic lectures. This finding emphasized the salience of engagement with the readers and the writer's stance. As for the least frequent bundles, this study showed that the learners rarely used text-oriented bundles. There was some restricted use of text-oriented bundles, either the types or the tokens, resulting in the minimum frequency of use.

In contrast, Salazar (2014) claimed that participant-oriented bundles are the least used by non-native writers. The different result is plausible due to varying levels of learners' proficiency, different learning quality, and various needs for the academic discipline. Similarly, Wright (2019) who examined lexical bundles in the literature review section, also shared a contrasting result, showing that the second most frequent bundles are discourse organizers or text-oriented bundles, and participant-oriented bundles are the least frequent. Yakut et al. (2021) emphasized the least important bundles in the study are the participant-oriented bundles. The contrasting findings can also be found in the studies of textbooks reported by Lee (2020), showing that discourse-organizing bundles are more dominant than stance expression bundles in linguistics textbooks. Another study was conducted by Hussain et al. (2021) which demonstrated the different results arguing that discourse organizing bundles (text-oriented bundles) are frequent. However, considering the distinct characters of textbooks and scientific papers, the discrepancies might be unavoidable.

More specifically, this study showed the absence of structuring signals as the subcategory of text-oriented bundles. Structuring signals are discourse organizers or text-oriented bundles to help connect each part of the writing into a more structured whole. In the text-oriented category, there are relatively smaller members compared to research-oriented and participant-oriented bundles. For instance, transition signals only have four types, the resultative signal only has one type, and the framing signal

has one type, leaving the structuring signal with no subcategory member. The minimal use of text-oriented bundles and the absence of structuring signal bundles indicated that the learners were not fully mastered or knowledgeable in organizing their writing. Although earlier studies proved the usage deficiency of text-oriented bundles, this present study specifically addressed structuring signals as the subcategory.

Concerning the type/token ratio, the results of this study demonstrated a small ratio since the variants were limited or restricted. The number of occurrences was pretty high, but the number of variants was limited, which can be seen from, for example, research-oriented bundles. It proves that the learners already had the repertoire of lexical bundles, but it did not vary to some extent. There were finite variants of the bundles, and there was the domination of research-oriented bundles in writing. [Chen and Baker \(2010\)](#) found out the use of specific connector bundles in students writing, which shows the repetitiveness. Similarly, [Salazar \(2014\)](#) argued that non-native English writers overuse specific bundles yielding to a lack of variation.

On the contrary, [Shin \(2018\)](#) and [Yakut et al. \(2021\)](#) proved that L1 and L2 writers use the same proportion of lexical bundles. In [Shin's \(2018\)](#) study, the corpus was collected from first-year university students during the initial higher education stage. Meanwhile, [Yakut et al.'s \(2021\)](#) study focused on the doctoral dissertation as the highest level of education. These studies demonstrate the similarity of lexical bundle use between L1 and L2 English writers showing the plausible linear development of lexical bundle mastery for native and non-native writers, either in the initial or the final stage. However, it is premature to state the previous statement certainly, and thus this result invites further studies on L1 vs. L2 English written production.

The study's findings not only informed us and helped us understand the features of academic writing, but they could also inform pedagogy ([Hyland 2012](#); [Meunier, 2012](#)). Regarding the frequency of use, research-oriented bundles being used the most frequently across disciplines either by L1, L2, or FL learners indicate that these bundles are crucial in discourse construction. Regardless of the nativity and professionalism, research-oriented bundles outnumber other word sequences in the texts, so research-oriented bundles should be perceived as the basic bundles of most discourses. Based on the frequency findings, the topic subcategory had the lowest token and type frequency compared to all subcategories in the research-oriented type. Topic bundles are related to the field of research ([Hyland, 2008](#)) or associated with the topic of the academic essays. It might be due to the lack of mastery of the topic being written or the learners' ignorance. Although further studies need to be done regarding this issue, it is relevant to state that learners need to master the topic well before writing.

As for the other two bundle functions (i.e., participant-oriented and text-oriented bundles), they were used more restrictedly, as seen by the lower type and token frequencies, compared to the research-oriented bundles. Based on the result, text-oriented bundles were negligible in the study, which is quite surprising because these bundles were important to keep the writing flow. Therefore, the minimum use of text-oriented bundles should be considered seriously by writing instructors. They should introduce this bundle function and its concrete use to help learners organize their writing.

Interestingly, the absence of structuring bundles (as part of text-oriented bundles) might indicate the students' low writing proficiency. The absence of certain

bundles reveals the lack of novice writers' fluency (Hyland, 2012). Writing an essay is about combining sentences into paragraphs and combining paragraphs as a whole part. Thus, the ability to structure the essays or texts is crucial, and it can be done by using several markers that might be in the form of sequences. The absence of structuring bundles emphasized the students' inability to organize their writing. It is then crucial to expose learners to structuring bundles and train them to develop and structure their essays.

Another point to revisit is the variants of lexical bundles. Given the type and token frequency results, there was an evident gap between variants and the frequency of use. As discussed, the corpus's TTR ratio of lexical bundles was small. The variants were minimal, but the use was high, showing the learners' lack of variants in word combinations. Although lack of variation commonly happens in the L2 and EFL contexts, this should be solved by providing adequate language input related to discourse chunks or word combinations to the learners. Learners must be familiar with various bundles to enrich their lexical profile and writing quality.

Based on the study's results, it is necessary to facilitate learners with sufficient writing materials. Realizing that word bundles construct texts, students must be familiar with them. To begin with, teachers can use the Academic Formulas List I (Simpson-Vlach & Ellis, 2010) to introduce lexical bundles to the learners. By doing so, they are exposed to lexical bundles more intensively. Teachers can also use a corpus, such as the BAWE corpus, or academic sub-corpus in the COCA, to facilitate authentic and massive examples of bundles used naturally in academic prose. Some lexical bundles-related materials include the introduction to various bundles serving various functions. The emphasis should be placed on text-oriented bundles, which are absent while important in texts.

All in all, consulting corpus in designing writing materials should be taken into account. Teachers or instructors can design teaching materials comprehensively by checking the language used in an appropriate corpus (in this context, written academic corpus). Some corpora are highly recommended, such as the general reference corpus as the general information of lexical bundles usage. There are COCA and BNC2014 that are quite updated and massive so teachers (and learners) can better understand using lexical bundles. Specifically, BAWE can be recommended as an academic written corpus that comprises students' writing so teachers can explore what lexical bundles are prominent among native writers. Teachers can then focus on what should be emphasized and repeated or omitted in their teaching materials.

Besides referring to native speaker corpora, it is essential to consult a learner corpus, such as ICLE and MICUSP. Teachers might have some insights into the learners' language development by using a learner corpus, especially on lexical bundles. By doing so, teachers might find the shared principles of lexical bundle use across learners with various L1 backgrounds. In addition, learners also need to be more familiar with academic terms, which means teachers should refer to the Academic Word List and Academic Formulas List in designing and delivering writing materials. Regarding the learner corpus utilized in the present study, it can be a guide to consult teaching materials for the writing instructors or lecturers in the department. The learner corpus can be developed further into a bigger one for wider purposes.

6. CONCLUSION

The study's findings show that all three major functions were identified in the learner corpus. They are identified in the corpus along with the subfunctions of each function category. However, the functional categories' distribution varied as the use of certain bundles was lower than the others, with research-oriented bundles outnumbering all bundles. In contrast, text-oriented bundles were underused by the learners. This bundle distribution should be considered because in writing argumentative texts, it is also essential to use text-oriented bundles to organize the arguments. More specifically, structuring signals (the subcategory of text-oriented bundles) were absent in the learner corpus, indicating the learners' lack of discourse markers knowledge, which should be highlighted for improvement. Despite the functions, the variants of the bundles were also used limitedly compared to the bundle tokens. The low frequency and the absence of particular bundles might signal the need to introduce learners to those bundles more intensively in the writing courses.

This study involved the fourth-semester students who had passed several writing courses in the previous semesters. However, since they were not final-year students, they have not fully learned and mastered academic writing. The findings of the present study can be used to map the learning progress, considering that the fourth semester is the medium phase of the whole study period in a university. To map the learning achievement, it is expected for future researchers to conduct the study in the final year of the study to see the complete mastery of the learners. Concerning the corpus, it is also necessary to collect exemplary sources of data to get a clearer picture of the learners' language development. Furthermore, this research does not compare the learner corpus with any other control corpus; the results are solely based on the investigation of the learner corpus. Thus, it is also intriguing to have a control corpus to compare the findings from a learner corpus or a general reference corpus.

REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Alquraishi, M. A. (2014). *Lexical bundles in an advanced INTOCSU writing class and engineering texts: A functional analysis* [Master's thesis, Colorado State University]. CSU Theses and Dissertation. <http://hdl.handle.net/10217/83876>
- 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*. Longman.
- Biber, D., Reppen, R., Schnur, E., & Ghanem, R. (2016). On the (non)utility of Juilland's *D* to measure lexical dispersion in large corpora. *International Journal of Corpus Linguistics*, 21(4), 439-464. <https://doi.org/10.1075/ijcl.21.4.01bib>
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *LancsBox* (v.5.x) [Computer software]. Lancaster University. <http://corpora.lancs.ac.uk/lancsbox>

- Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, 10(2), 292-303. <https://doi.org/10.17509/ijal.v10i2.28592>
- Burch, B., Egbert, J., & Biber, D. (2017). Measuring and interpreting lexical dispersion in corpus linguistics. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 3(2), 189-216. <https://doi.org/10.1558/jrds.33066>
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 20, 38-52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Chen, Y.-H., & Baker, P. (2014). 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>
- 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>
- 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>
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk: An Interdisciplinary Journal of Language, Discourse, & Communication Studies*, 20(1), 29-62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Fajri, M. S. A., Kirana, A. W., & Putri, C. I. K. (2020). Lexical bundles of L1 and L2 English professional scholars: A contrastive corpus-driven study on applied linguistics research articles. *Journal of Language and Education*, 6(4), 76-89. <https://doi.org/10.17323/jle.2020.10719>
- Fitrianasari, N. I., Sulistyorini, T., & Sugiarti, T. R. (2018). Lexical bundles in academic writing by undergraduate and graduate students of English Language Education Program. *Loquen: English Studies Journal*, 11(02), 1. <https://doi.org/10.32678/loquen.v11i02.1102>
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA. In S. Granger (Ed.), *Learner English on computer*. Longman.
- Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook volume I* (pp. 259-274). Walter de Gruyter.
- Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story* (Vol. 71). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.71>
- Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437. <https://doi.org/10.1075/ijcl.13.4.02gri>
- Hyland, K. (2008). 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>
- 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. (Kevin). (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>
- Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse functions of lexical bundles in Pakistani chemistry and physics textbooks. *GEMA Online® Journal of Language Studies*, 21(1), 221-238. <https://doi.org/10.17576/gema-2021-2101-13>
- Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, 7(1), 131. <https://doi.org/10.17509/ijal.v7i1.6866>
- Lee, H.-K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research*, 37(1), 121-145. <https://doi.org/10.17250/KHISLI.37.1.202003.005>
- Lee, H.-K., & Kim, H.-O. (2017). Features of lexical bundles in academic writing by non-native vis-à-vis native speakers of English. *English Language and Linguistics*, 23(2), 67-88. <https://doi.org/10.17960/ELL.2017.23.2.003>
- Liu, C.-Y., & Chen, H.-J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, 58, 122-137. <https://doi.org/10.1016/j.esp.2019.12.003>
- Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32, 111-129. <https://doi.org/10.1017/S0267190512000128>
- 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>
- Gil, N. N., & Caro, E. M. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language & Literature*, 12(1), 65. <https://doi.org/10.5565/rev/jtl3.794>
- Oktavianti, I. N., & Sarage, J. (2021). Lexical Bundles in Students' Argumentative Essays: A Study of Learner Corpus. *Indonesian Journal of EFL and Linguistics*, 6(2), 510-534. <http://dx.doi.org/10.21462/ijefl.v6i2.421>
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48-59. <https://doi.org/10.1016/j.jeap.2013.12.001>
- Ren, J. (2021). Variability and functions of lexical bundles in research articles of applied linguistics and pharmaceutical sciences. *Journal of English for Academic Purposes*, 50, 100968. <https://doi.org/10.1016/j.jeap.2021.100968>
- Sadeghi, M. (2015). Investigation of EFL learners discourse functions of lexical bundles: A case of written register. *International Journal of English and Literature*, 6(1), 9-15. <https://doi.org/10.5897/IJEL2014.0689>
- Salazar, D. (2011). *Lexical bundles in scientific English* [Unpublished doctoral dissertation]. Universitat de Barcelona.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.65>

- Shin, Y. K. (2018). *Lexical bundles in argumentative essays by native and nonnative English-speaking novice academic writers* [Doctoral dissertation Georgia State University]. ScholarWorks. <https://doi.org/10.57709/11994284>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>
- van Lancker-Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language & Communication*, 24(3), 207-240. <https://doi.org/10.1016/j.langcom.2004.02.003>
- Wachidah, W. D. N. A., Fitriati, S. W., & Widhiyanto, W. (2020). Structures and functions of lexical bundles in findings and discussion sections of graduate students' thesis. *English Education Journal*, 10(2), 131-142.
- Wingate, U. (2012). 'Argument!' Helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11(2), 145-154. <https://doi.org/10.1016/j.jeap.2011.11.001>
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1-14. <https://doi.org/10.1016/j.esp.2018.09.001>
- Yakut, I., Yuvayapan, F., & Bada, E. (2021). Lexical bundles in L1 and L2 English doctoral dissertations. *Journal of Teaching English for Specific and Academic Purposes*, 9(3), 475-493. <https://doi.org/10.22190/JTESAP2103475Y>
- Yang, W., & Fang, Q. (2021). The structures and functions of lexical bundles in argumentative essays by Chinese EFL students at the tertiary level. *International Journal of TESOL Studies*, 3(3), 41-60. <https://doi.org/10.46451/ijts.2021.09.03>
- Yang, Y. (2017). Lexical bundles in argumentative and narrative writings by Chinese EFL learners. *International Journal of English Linguistics*, 7(3), 58-69. <https://doi.org/10.5539/ijel.v7n3p58>