Indonesian Journal of EFL and Linguistics
Vol. 6 No. 2, 2021
eISSN: 2503-4197, pISSN: 2527-5070
www. indonesian-efl-journal.org
doi: http://dx.doi.org/10.21462/ijefl.v6i2.421

Lexical Bundles in Students' Argumentative Essays: A Study of Learner Corpus

Ikmi Nur Oktavianti
Universitas Ahmad Dahlan
ikmi.oktavianti@pbi.uad.ac.id

Japen Sarage
Universitas Ahmad Dahlan
japen.sarage@pbi.uad.ac.id

Abstract:
This study investigates the structures of 3- to 5-word bundles used by EFL learners at the university level in writing argumentative essays. This qualitative corpus-based study focuses on answering the structural category and frequency of lexical bundles in students' essays. The data of this study are sentences containing lexical bundles in the students' essays, and the data were collected from students' essays compiled as a learner corpus. The lexical bundles were identified from the corpus with the assistance of a corpus tool, LancsBox, using the n-gram feature. This study used ten occurrences as the cutoff frequency and Gries' DP as the dispersion threshold to identify the lexical bundles. The bundles were then classified into structural taxonomies, and the frequency of use of the lexical bundles was also investigated to complete the analysis. Academic Formulas List comprising bundles commonly used in the academic context was used to validate the bundles. The results show enormous structures of NP-based, VP-based, PP-based, and others identified in the learner corpus with NP-based bundles as the most frequent bundles and 'the use of' as the most frequent individual bundle. However, the variants of the bundles in the learner corpus are still dominated by fixed frames. In addition, apart from the shared bundles between the corpus and Academic Formulas List, there are some discrepancies related to the registers. It indicates that writing courses and writing

materials should provide learners with more variants of lexical bundles and the appropriate context of use.

***Keywords**: argumentative essay, corpus linguistics, learner corpus, lexical bundles*

1. INTRODUCTION

In using language, language users combine words to construct phrases or clauses. Chomsky (1965, 2015) states that humans have inherent capacities to generate and understand sentences they have never heard before. Language users are creative, and our cognitive system has the ability for novelty. However, language users also utilize a prefabricated multi-word combination of the existing words (Erman & Warren, 2000; Meunier, 2012). The prefabricated multi-word combination means that the sequence is stored or retrieved mentally as a single word (Wood, 2015). The latter is plausible since the human cognitive system has to store and process information, including language, effectively and efficiently by using prefabricated multi-word combinations (Hye-Kyung Lee, 2020).

The term multi-word combination is also known as *formulaic language,* indicating that language is patterned. The cognitive system stores and retrieves information in association; thus, lexicon storage and retrieval are done associatively (Szudarski, 2017). In reality, it is easier and more efficient for language users to retrieve the lexicons in association. Szudarski (2017) states that we are psychologically primed to use vocabulary in specific collocational patterns, meaning language users construct language in formulaic sequences.

Wood (2015) describes several categories of formulaic languages, such as phrasal verbs, collocations, idioms, lexical phrases, lexical bundles, and others. Nevertheless, this research focuses on lexical bundles, defined as the category of formulaic language characterized by the combination of three or more words identified in a corpus using a corpus analysis software program (Biber et al., 1999; Wood, 2015). Byrd & Coxhead (2010) add the definition of lexical bundles with the aspect of frequency, indicating that lexical bundles are about the length of the combination and the frequency of use or frequency of occurrence in a particular discourse. By concerning frequency, it is relevant to say that lexical bundles can distinguish certain text types. Lexical bundles are frequently used in academic writing such as journal articles, and particular kinds of bundles are characteristic of specific disciplines or discipline-bound (Cortes et al., 2002). Biber (2006) discovered that academic disciplines use lexical bundles differently, with social and natural sciences using them more frequently than humanities.

This research focuses on lexical bundles based on several rationales. First, lexical bundles are measurable (with the typical combination of 3- to 4-word strings) and frequent (Biber et al., 2004; Hye-Kyung Lee, 2020; Wood, 2015). The lexical bundles' length and the possibility of observing the frequency are the fundamental reasons for selecting lexical bundles, among other concepts in formulaic language. Thus, lexical bundles are observable and can be analyzed more comprehensively. Second, lexical bundles have discourse functions that are important to build up the discourse. Formulaic language, including lexical bundles, indicates the learners' native-like fluency. In other words, the use of formulaic language is part of native speakers' competence (Meunier, 2012; Wray & Perkins, 2000). Wood (2015) points out that formulaic language is likely the key to second language fluency and native-like selection: the tendency to use typical ways of expressing things, despite the supposed infinite potential of language. Thus, formulaic language is an essential element of proficient language use (Crossley et al., 2014; Kremmel et al., 2015).

Research shows that formulaic language, including lexical bundles, is not used appropriately by foreign language learners or non-native speakers (Granger, 2018; Meunier, 2012; Paquot & Granger, 2012). There are some differences between non-native (L2) and native (L1) or expert production on the formulaic sequence. Moreover, some research has investigated the difficulties experienced by L2 writers in using lexical bundles (Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008). In language learning, it is also found that L2 learners' have some difficulties learning lexical bundles. The results of several studies show that there are some difficulties and discrepancies in using lexical bundles among L2 or FL learners (Ädel & Erman, 2012; Chen & Baker, 2010; Shin, 2018). Chen & Baker (2010), who compared L1 and L2 of students' academic writing, found that L2 learners used fewer lexical bundles in their writing and had particularly limited ability to use bundles for certain discourse functions such as hedging. On the other hand, many studies show that FL learners rely more on restricted lexical bundles than native writers (Ädel & Erman, 2012). Unlike collocations, the use of lexical bundles decreases as proficiency or the time spent in an English-speaking country increases (Groom, 2009).

Given the results of the previous studies, it is a pivotal move to introduce or strengthen the materials of lexical bundles for EFL learners. Formulaic sequences, including lexical bundles, are prevalent in academic discourse and offer a necessary means of differentiating disciplinary practices (Biber et al., 2004; Durrant, 2017; Hyland, 2008). Moreover, in terms of the academic context, lexical bundles are frequently used in published academic writing, such as journal articles. Particular types of bundles are characteristic of specific disciplines (Cortes et al., 2002). Since lexical bundles are ubiquitous in academic writing, they can show the students how to be fluent and meet academic reader expectations (Wright, 2019). Thus, the teaching of lexical bundles should have been taken into account in the ELT practice. The teaching materials of writing courses should cover adequate explanations on

lexical bundles. However, it is also essential to map the students' proficiency in using the formulaic sequences, especially lexical bundles.

Many studies on lexical bundles have been focusing on the use of lexical bundles in academic prose, such as journal articles (Fitrianasari et al., 2018; Nasrabady et al., 2020; Pan et al., 2016), theses/dissertations (Fitrianasari et al., 2018; Wachidah et al., 2020) and book (Hye-Kyung Lee, 2020). In the academic context, the studies of lexical bundles have also been conducted based on proficiency level by comparing the use of lexical bundles between novice writers and professional writers (Fajri et al., 2020). Furthermore, the use of lexical bundles in the academic genre has been investigated in specific parts of academic essays, such as the introduction section (Jalali & Moini, 2014). Some previous studies also examined the use of lexical bundles in various academic disciplines (Durrant, 2017; Kwary et al., 2017), proving that lexical bundles across fields are not homogeneous. Several studies also compare lexical bundles of native writers to non-native writers (Chen & Baker, 2010; Navarro Gil & Martínez Caro, 2019; Pan et al., 2016).

In the Indonesian context, there are some relevant studies of lexical bundles analyzing students' writing, predominantly analyzing theses (Fitrianasari et al., 2018; Samodra & Pratiwi, 2018; Wachidah et al., 2020) and journal articles (Fajri et al., 2020; Kwary et al., 2017). However, little is known about the lexical bundles in students' essays (non-theses) in Indonesia, especially at the university level. This study aims at identifying lexical bundle use, which can contribute to mapping university students' fluency in using lexical bundles as a critical aspect of written academic English. The research questions are as follows: (1) what are the structural taxonomies of the lexical bundles in students' argumentative essays? (2) what are the most frequent lexical bundles? and (3) what are the pedagogical implications of the study's findings?

2. LITERATURE REVIEW

2.1 Lexical Bundles

Lexical bundles can be defined as a sequence of words or other elements stored and retrieved as a whole from memory at the time of use (Wray, 2002). Lexical bundles are characterized by their distinctive feature, i.e., they do not mean per se, but they are units of function serving as the characteristics of a specific text genre (Wood, 2015). Lexical bundles comprise three or more word sequences identified in a corpus of natural language using the assistance of a corpus tool. Another characteristic of lexical bundles is the occurrence in various text types and disciplines in academic contexts (Wood, 2015). Biber & Conrad (1999) argue that the word sequences are persistent, and they might be assumed as simple expressions and can be acquired easily.

Lexical bundles are intensively studied in register-specific phraseology research (Szudarski, 2017). The ease primarily causes this since lexical bundles can be

identified using corpus tool assistance. Gray & Biber (2015) point out that the characteristic of corpus study of phraseological patterns (including lexical bundles) is inductive processing. In inductive processing, the corpus is analyzed inductively with frequency as the main criterion. Certain units that cannot meet the basic requirement (frequency) will be removed from the identification. It shows that the computer identifies the noteworthy word sequences (Gray & Biber, 2015; Szudarski, 2017).

Regarding structural taxonomy, Biber et al. (1999) have presented 12 structural classifications of lexical bundles, which becomes the fundamental basis of further development of the structural category. They are: (1) noun phrase with of-phrase fragment, (2) noun phrase with other post-modifier fragment, (3) prepositional phrase with embedded of-phrase fragment, (4) other prepositional phrase fragment, (5) anticipatory *it* + verb phrase/adjective phrase, (6)passive verb + prepositional phrase fragment, (7) copula *be* + noun phrase/adjective phrase, (8) (verb phrase +) *that*-clause fragment, (9) (verb/adjective +) *to*-clause fragment, (10) adverbial clause fragment, (11) pronoun/noun phrase + *be*, and (12) other expressions. However, Chen & Baker (2010) modify and categorize them into bigger taxonomies, namely NP-based, PP-based, and VP-based bundles. A further development, Navarro Gil & Martínez Caro (2019) and Hye-Kyung Lee (2020) classify lexical bundles into NP-based, PP-based, VP-based, and *Others* (such as adverbial fragments). For rigorous analysis, this study employed various classifications for the need of research by modifying the categories proposed by Biber et al. (1999), Chen & Baker (2010), Hye-Kyung Lee (2020), Navarro Gil & Martínez Caro (2019).

2.2 Argumentative Essays

Argumentative essays can be categorized as expository texts since they present, explain, analyze, and interpret the facts, and they argue for a particular perspective (Shin, 2018). Writing an argumentative essay requires critical and logical thinking and incorporating sources coherently (Parkinson & Musgrave, 2014). Regarding text genre, argumentative essay is the most common genre in undergraduate students writing (Mei, 2006; Wingate, 2012), especially in the humanities, arts, and social sciences (Hewings, 2010; Shin, 2018).

Studies have confirmed that each text genre has its linguistic features (Biber & Conrad, 2009; Biber & Egbert, 2018). For instance, previous studies have shown the characteristics of verb choice are specific to types of texts, such as in news texts and editorial newspaper articles (Oktavianti & Adnan, 2020; Oktavianti & Ardianti, 2019). Academic essays also have distinctive features compared to other text types. In formulaic language, lexical bundles are ubiquitous in academic texts, showing that these word sequences are necessary (Kwary et al., 2017; Navarro Gil & Martínez Caro, 2019; Wright, 2019). Therefore, it is intriguing to investigate the use of lexical bundles among EFL learners at the university level.

2.3 Learner Corpus

Learner corpus is a specialized corpus comprising language data from learners of a second or foreign language (Flowerdew, 2012, 2014; Granger, 2008). It has all the characteristics of a general corpus, and the only difference is the data (Granger, 2008). The criteria of a learner corpus are that the language should be neither the first language of the learners or an institutionalized additional language in the country they live in (Granger, 2008). Historically, the first learner corpus was the *International Corpus of Learner English* (ICLE) compiled in a project initiated by Sylviane Granger at the University of Louvain, Belgium, in 1990. These projects are the counterpart of the large corpus project ICE (*International Corpus of English*), having the data from English language learners (Flowerdew, 2012). ICLE contains the sub-corpus of academic argumentative essays written in the enormous L1 background of the advanced learners, namely French, German, Polish, Greek, etc. In a different continent, a learner corpus compiling academic essays of English learners in Hongkong was built by John Milton at Hongkong University of Science and Technology (HKUST), and a learner corpus of English learners in Japan was established by Yukio Tono consisting of academic essays written by students of junior high, high school, and university level (Flowerdew, 2012).

The aim of building a learner corpus is to facilitate research in second language acquisition. Granger (1998) states that second language research is to uncover the principles that govern the process of learning a second or foreign language. In other words, learner corpus can assist the mapping of the learner language development in learning a second or foreign language (Flowerdew, 2012; Granger, 2008). Research in second or foreign language learning should comprehensively use learner corpus to understand the learners' language development. Thus, considering its crucial function, the corpus's design and size should be concerned with building a learner corpus. Sinclair (1991) points out that the quality of the corpus determines the results of corpus analysis. There should be a clear design criterion for learner language corpus by considering the L1 background, the level of the learners, the sex, the region, and the task setting. In regard to the size, Flowerdew (2014) argues that it is not as important as the design. Many learner corpora have large-scale data comprising of million words. However, when the qualitative methods are selected, smaller learner corpora consisting of l50.000-150.000 words can be used to examine a specific linguistic unit or feature in a particular genre.

2.4 Previous Studies

In the recent development, the studies of lexical bundles have been focused on the analysis toward various registers (spoken and written) and text genres which primarily focus on academic prose. Research on lexical bundles has been done in terms of the register, i.e., spoken or written register, and those research prove that

the use of lexical bundles is distinct in spoken and written context (Biber et al., 2004; Biber & Barbieri, 2007). Biber et al. (2004) prove that lexical bundles are unique linguistic constructs since they occur differently in different texts, e.g., lexical bundles used in classroom teaching (spoken) serve more various functions than those used in academic prose (written). Similarly, Biber & Barbieri (2007) point out the distinctive use of lexical bundles in different university registers, i.e., core instructional and course syllabi. Their study shows that lexical bundles are most commonly used in course syllabi (written) than in core instruction (spoken). As for the classroom context, Csomay (2013) identifies that the use of lexical bundles is slightly different in different classroom teaching sessions (e.g., opening phase, instructional phase).

Despite the apparent differences of lexical bundles in spoken and written texts, according to (Biber & Barbieri, 2007), the genre is more important in distributing lexical bundles (Durrant, 2017; Hyland, 2012). As they are strongly related to the academic context, lexical bundles have been investigated frequently in the academic genre (Durrant, 2017; Fitrianasari et al., 2018; Samodra & Pratiwi, 2018; among others). To this extent, the studies have been focused on several aspects, including the use of lexical bundles among students as novice writers (Durrant, 2017; Fitrianasari et al., 2018; Samodra & Pratiwi, 2018), the use of lexical bundles between novice writers and professional writers (Fajri et al., 2020), the use of lexical bundles between native writers and non-native writers (Navarro Gil & Martínez Caro, 2019), the use of lexical bundles across academic disciplines (Durrant, 2017; Kwary et al., 2017), and the use of lexical bundles in academic texts and narrative texts (Yang, 2017).

Several studies investigated lexical bundles among novice writers (Durrant, 2017; Fitrianasari et al., 2018; Samodra & Pratiwi, 2018; Ulfa & Muthalib, 2020; Wachidah et al., 2020). Those studies show that students use lexical bundles to various extents and are distinct from the use of lexical bundles of professional writers. Durrant (2017) shows that students from multiple disciplines use lexical bundles differently. Meanwhile, in the level of education, Fitrianasari et al. (2018) argue that graduate students use more varied lexical bundles than undergraduate students along with distinct functions of lexical bundles. In terms of the patterns, Ulfa & Muthalib (2020) examine the n-gram of lexical bundles in students' essays and determine that 3-word combinations were more frequent than 4-word bundles. As for the variety, Wachidah et al. (2020) claim that students use all forms of lexical bundles serving all three functions. Although there are some differences across disciplines and levels of education, linguistic background slightly influences the use of lexical bundles in academic texts. Samodra & Pratiwi (2018), who compared English and Indonesian abstracts, identify no difference in lexical bundles type in English and Indonesian abstracts.

Concerning academic writing, the expertise (i.e., professional writing) and nativity play a vital role in the use of lexical bundles, as observed in the studies done by Navarro Gil & Martínez Caro (2019), Chen & Baker (2010), Pan et al. (2016), and Fajri et al. (2020), among others. Chen & Baker (2010) find out that L2 student writers use more limited lexical bundles than professional L1 writers. However, Navarro Gil & Martínez Caro (2019) argue that there is no significant difference between novice versus professional writers using lexical bundles. They examine lexical bundles in dissertation and published research articles and point out that the use of the bundles is relatively similar. The main difference lies in the academic disciplines, which means each academic discipline has its preferable bundles (regarding the structure and the function).

Regarding nativity, lexical bundles in academic contexts vary. Several studies prove that there are some discrepancies in the use of lexical bundles between native writers and non-native writers (Ädel & Erman, 2012; Chen & Baker, 2010; Fajri et al., 2020; Hye-Kyung Lee & Hyeon-Okh Kim, 2017; Lou, 2012; Pan et al., 2016; Shin, 2018). Fajri et al. (2020) claim that L2 writers use more lexical bundles than L1 writers, except that the lexical bundles are underused (compared to the lexical bundles in L1 English writing). Similarly, Pan et al. (2016) state that L2 professional writers use lexical bundles differently from L1 professional writers. Shin (2018) argues that there are differences in lexical bundles among native and non-native students. More specifically, the results of Ädel & Erman (2012)'s study shows that native writers have a more significant number of types of lexical bundles. These findings are similar to the tradition in SLA research. Nevertheless, Lou (2012), who examined the lexical bundles in L2 students' writing (L1 Chinese) and native English students' writing, points out that Chinese students share several types of bundles with native English speakers. This study proves that there is something in common in terms of lexical bundles in the academic context.

In academic texts, lexical bundles are used distinctively regarding several aspects, such as nativity, level of education, expertise, and academic discipline. As Navarro Gil & Martínez Caro (2019) mentioned, the main difference of lexical bundle use is in the varied fields. Several studies (Hyland, 2012; Hyland & Jiang, 2018; Kwary et al., 2017) claim that some discrepancies exist in using lexical bundles across academic disciplines. Kwary et al. (2017) investigate the lexical bundles among three disciplines, i.e., health science, physical science, social science. This study shows that physical science and social science shared lexical bundles the most. Surprisingly, no lexical bundles were shared between health science and physical science and neither between health science and social science. Those results strongly prove that each academic discipline might use bundles distinctively. In a specific field, such as linguistics, Hye-Kyung Lee (2020) and Nasrabady et al. (2020) have conducted research to produce subject-related lexical bundles. These studies successfully demonstrate that linguistics has some bundles specific to the discipline.

Lexical bundles are also interesting to investigate in another text genre, e.g., narrative texts. Following the statement from Biber & Barbieri (2007) and Hyland (2012) in which genre is the vital role for lexical bundle selection or distribution, Yang (2017) mentions that students used more 4-word bundles in argumentative essays than those in narrative essays. However, no significant structural difference was found in the 4-word lexical bundles between the two text types. Yang (2017) also argues that students utilized stance bundles more frequently than the other functional categories of bundles in their argumentative writings. Still, they used referential expressions more in the narrative writing. It shows that bundles are used differently in another genre, such as in narrative texts. In the context of literary works, Kaur et al. (2019) argue that lexical bundles are used in children's fiction. The study results show that prepositional and verb phrases dominate lexical bundles in the selected children's fiction. As for the function, referential lexical bundles are the most frequent, followed by some other functions, such as action-related expressions and stance. The results indicate the presence and the use of lexical bundles in another text genre.

Concerning the pedagogical context, lexical bundles have been studied related to learners' competence. Staples et al. (2013) show that lower-level learners used lexical bundles more frequently. The study also proves the similar use of stance and discourse organizing bundles across learners' proficiency levels. In contrast to Staples et al. (2013), Chen & Baker (2014) find a contradictory finding. They prove that learner writing at lower levels tends to share lexical bundles of conversation and the lexical bundles used by proficient writing are closer to that of academic writing.

As for the Indonesian context, the study of lexical bundles is not new but not widely recognized. Most of the studies focused on lexical bundles production among students (learners of English), such as in essays, theses, or dissertations (Fitrianasari et al., 2018; Samodra & Pratiwi, 2018; Ulfa & Muthalib, 2020). Some other studies focus on using lexical bundles in academic and professional contexts, e.g., journal articles (Fajri et al., 2020; Kwary et al., 2017). Some of them have been conducted using corpus-based analysis (Fajri et al., 2020; Kwary et al., 2017; Ulfa & Muthalib, 2020).

3. RESEARCH METHODOLOGY

To comprehensively answer the research questions, this study employed a corpus-based analysis. Therefore, the method section consists of corpus construction and the corpus tool selection alongside data collection, lexical bundle identification, and data analysis.

3.1 Corpus Construction

The corpus collected in this research is a learner corpus comprising students' argumentative essays. The participants of this study were all fourth-semester students (169 students) of an English Education Department enrolled in *Writing in*

Academic Context course. The participants of this study fulfilled the following criteria: (1) fourth-semester students enrolled in the *Writing in Academic Context* course, (2) having passed/taken the previous writing courses in semesters 1—3, (3) having joined the *Writing in Academic Context* course for at least seven meetings.

For the compilation of the student essays, there was an instrument in the form of a writing task. The instrument's design followed the criteria of the learner corpus *International Corpus of Learner English* (Granger, 1998). As for the variables, Granger (2008) proposes two types of variables, namely *learner variables* and *task variables* (table 1) which have been included in the design criteria of the learner corpus (table 1 and table 2). Table 1 shows the variables involved, namely learner and task variables, which should be concerned in designing the learner corpus.

Table 1: Variables in learner corpus (Granger, 2008)

Learner variables		Task variables	
General	Specific	General	Specific
Age: 18-20-year-old	Learning:	Medium: written	Task type: argumentative essay
Gender: F, M	Proficiency: intermediate	Field: technology, education	Conditions: timed
Region: Indonesia	L2: English	Genre: academic	
Mother tongue: local language(s) or Indonesian	Other Foreign Languages		

In terms of the participants (i.e., the learners), two significant points should be considered: the learners' shared features and the variable features. Below is table 2, presenting those features to design the learner corpus (Granger, 1998, 2008).

Table 2: Design criteria of the learner corpus

Shared features	Variables features
Age: 18-20	Sex: male, female
Learning context: EFL	Mother tongue: local language(s) in Indonesia or Indonesian language
Level: intermediate (fourth-semester students of an English Education Department)	Region: Indonesia
Medium: written	
Genre: argumentative	
Task setting: non-exam, allocated time, more than 500 words, less than 1000 words, with the topics of online learning, being viral, digital minimalism, women, and higher education	

3.2 Corpus Tool

Since the n-gram analysis of corpus software assists the categorization of lexical bundles, it is essential to describe the corpus tool used in this study. This study used *LancsBox* (Brezina et al., 2018) due to several rationales, including the flexibility to deal with many file formats (e.g., pdf, doc, txt), the ability to work well with the English language, and the availability of the necessary analytical tools (e.g., KWIC, concordance, n-gram, frequency, dispersion).

3.3 Data Collection

In collecting the data, the writing task was distributed to the participants with one of the following topics: (a) *online learning is more effective than offline learning*, (b) *being viral is one of the important goals*, (c) *women should not focus on higher education, or* (d) *digital minimalism helps students stay focused.* The writing task was non-exam type and was required to be completed in 3 days. The number of words should be more than 500 words and less than 1000 words. The students then submitted the essays on *Google Drive* for a more accessible compilation. There are 169 essays from 169 participants with 87.939 tokens. Due to the limited time and technical barriers, this corpus is not annotated (raw corpus). This circumstance, however, does not affect the analysis since the corpus tool used in this study, *LancsBox* is designed to work well with the English language.

3.4 Lexical Bundles Identification

After the corpus was fully compiled, the following procedure is the identification of the lexical bundles. The procedure to identify the LBs was done with the assistance of *LancsBox*, using the n-gram feature. This research includes 3- to 5-word combinations since they are the most researched length and within manageable size for manual categorization and concordance checks (Hye-Kyung Lee, 2020). The 3-word bundles are included in this research because many trials have shown that these bundles display a wide range of productive expressions (Hye-Kyung Lee, 2020). Meanwhile, 4-word bundles were included because these bundles are more common than 5-word bundles and have more evident structures and functions than 3-word bundles (Hyland, 2008). However, to provide a broader range of lexical bundles for the analysis, this study includes 5-word bundles as done by Navarro Gil & Martínez Caro (2019).

The criteria applied to the identification are frequency cutoff and dispersion, which were calculated by using the corpus tool. Biber (2006) points out that the critical measure of a lexical bundle is frequency. Word sequences have to recur ten times per million words and appear in more than five texts to be categorized as lexical bundles. For a small corpus, a raw cutoff frequency ranging from two to ten times is adopted (Hye-Kung Lee, 2020). Following Hye-Kung Lee (2020), this study employed ten occurrences for the identification of lexical bundles. However, for the 5-word bundles, the cutoff frequency was adjusted to five occurrences in the corpus.

Different frequency cutoff points are that the frequency of lexical bundles decreases as they contain more words (Cortes, 2013).

Meanwhile, the dispersion thresholds also differ from research to research. This study employs Gries' *DP* (Gries, 2008) because it is reliable for a corpus that consists of many parts (more than 100 parts) (Biber et al., 2016), and it can handle different sized corpus parts (Gries, 2008). Since this corpus comprises 169 essays (or 169 parts), Gries' DP is more suitable for this study. The dispersion values are unitless, with a minimum value of zero and a maximum value of one. The minimum value of zero indicates the lexical bundle falls in only one part of the corpus, and the maximum value marks the occurrence of the lexical bundle across all parts of the corpus (Burch et al., 2017).

After conducting the automated procedure, the next step was done manually by checking each lexical bundle according to several identification criteria (Biber et al., 2004; Fajri et al., 2020; Hye-Kyung Lee, 2020). First, the bundles should have particular discourse functions. Word combinations that do not have meaning and discourse function were excluded. Second, word sequences containing proper nouns were excluded. Third, free combinations in which the meaning is compositional or transparent enough were excluded. Lastly, topic-related bundles were excluded from the list (e.g., *online learning system, teachers, and students*).

3.5 Data Analysis

There are several steps of analysis in this study. After all the lexical bundles were identified and sorted, they were classified based on the structural types and patterns (Biber et al., 1999; Chen & Baker, 2010; Hye-Kyung Lee, 2020; Navarro Gil & Martínez Caro, 2019). The structural taxonomies were then described to fully understand the phenomena of lexical bundle use among FL learners. This study conducted a quantitative analysis by calculating the frequency and the percentage of the lexical bundles in the corpus. After having the frequency and percentage, it is plausible to describe the learners' most frequent lexical bundles and structures. In addition, the lexical bundles under investigation were also verified. This study utilized the *Academic Formulas List* (AFL), the list of the most functional word combinations for teaching English for academic purposes (EAP) designed by Simpson-Vlach & Ellis (2010). This list is the most common formulaic sequence, consisting of 3- to 5-word combinations used in English Academic discourse and can be seen as the equivalent list to the *Academic Word List* (Simpson-Vlach & Ellis, 2010). By interpreting the results of analysis and confirming the AFL, this research's pedagogical implication can be explained comprehensively and provide more accurate suggestions for teaching lexical bundles.

4. FINDINGS

After identifying the bundles and excluding the word combinations that do not meet the criteria, there are 1124 tokens of lexical bundles (see appendix). Those bundles are classified based on the structure, and the use of the bundles is also investigated.

4.1 Structural Taxonomy

Regarding bundle identification, this study figures out several bundles under the criteria of 3-word, 4-word, and 5-word sequences. Those bundles are then classified into their structural classifications: NP-based, VP-based, PP-based, and others (Hye-Kyung Lee, 2020; Navarro Gil & Martínez Caro, 2019). Table 3 shows that, for 3-word bundles, the variants for NP-based and VP-based bundles are equally the same, followed by *other* and PP-based bundles.

Table 3: Structural taxonomy of 3-word bundles

NP-based	VP-based	PP-based	Others
the spread of	be able to	in the future	as we know
the development of	take advantage of	in my opinion	we have to
the importance of	according to the	in addition to	and many more
the existence of	take care of	in order to	as well as
one of them	stay focused on	in the world	I agree with
the number of	to sum up	around the world	as long as
the most important	it will be		as a result
various kinds of	there are also		not only that due to the
the use of	there is no		and so on
a lot of	there are some		

As for the 3-word bundles, there are NP-based (e.g., *the number of, the spread of*), VP-based (e.g., *take advantage of, stay focused on*), PP-based (e.g., *in addition to, in my opinion*), and other bundles (e.g., *as well as, as long as*). Examples of the 3-word bundles in sentences are below.

(1) Digital minimalism can manage *the use of* technology effectively.
(2) Therefore, it is important to limit *the use of* social media in daily life.
(3) *To sum up*, there are three big negative effects if one of your important goals is going viral.
(4) *In my opinion*, if online learning will continue to be done in the future, it will get better.
(5) *In my opinion*, there are several factors why online learning system is very ineffective
(6) I think nothing is wrong *as long as* they do not harm others

As for the 4-word bundles, there are also NP-based, VP-based, PP-based, along with *others*.

Table 4: Structural taxonomy of 4-word bundles

NP-based	VP-based	PP-based	Others
one of the important	must be able to	in the form of	as efficiently as possible
a lot of people	that can be done	at the same time	if you want to

In the 4-word bundles, there are all types of structural taxonomy, including NP-based (*one of the important, a lot of people*), VP-based (*must be able, that can be done*), PP-based (in the form, at the same time), and others (*as efficiently as possible, if you want to*). Given the data for 4-word bundles and compared to the 3-word bundles, the difference is noticeable in the variants. There are more bundles in the 3-gram category than in the 4-gram category, proving that the longer the bundles, the lesser the number. Some examples of the 4-word bundles are below.

(7) If we are famous, *a lot of people* will see our story or post on social media
(8) Lately, we have seen *a lot of people*, especially millennials, who are interested in being viral.
(9) This being a woman, we *must be able to* manage both career and education
(10) We *must be able to* address this wisely
(11) doing all of this *at the same time* can be stressful for parent.
(12) digital minimalism is a lifestyle to use technology only according to our needs and *as efficiently as possible.*

As the word-combination gets bigger, interestingly, the type of structural classification is getting more limited. As shown in table 5, the use of the bundles is limited, and the number is negligible.

Table 5: Structural taxonomy of 5-word bundles

NP-based	VP-based	PP-based	Others
most value to your life	will not be able to	-	we must be able to
	be known by many people	-	there are still many people
	to be seen and recognized	-	

Table 5 shows the NP-based bundles, the VP-based bundles, and others, but there is no PP-based bundle. In other words, the 5-word PP-based bundle is absent from the corpus. Given the data, it is noticeable that the number of the 5-word bundles is lesser than the other n-gram bundles under study. Below are two examples of 5-word bundles in sentences.

(13) they *will not be able to* focus on learning activities

(14) *there are still many people* who underestimate the danger of covid-19

In terms of the detailed patterns, table 6 shows the subcategory of the structural classification of bundles in the corpus, modified from (Biber et al., 2004; Chen & Baker, 2010; Hye-Kyung Lee, 2020; Navarro Gil & Martínez Caro, 2019).

Table 6: Structural classification of lexical bundles

Structural Classification	Subcategory	Example
NP-based	NP + of-phrase fragment	the use of, the quality of, the spread of, the development of, the importance of, the existence of, the number of, various kinds of, a lot of, one of them
	NP + to-infinitive fragment	most value to your life
	NP with postnominal fragment	one of the important, a lot of people
	Another NP fragment	the most important
PP-based	PP + NP fragment	in the future, in my opinion, in the world, around the world, at the same time
	PP + embedded of-phrase fragment	in the form of
	PP + embedded to-phrase fragment	in order to, in addition to
VP-based	Passive verb + PP fragment	be known by many people
	Copula be + VP	be able to
	VP + that-clause fragment	
	to + VP	to sum up, to be seen and recognized
	that + VP	that can be done
	Verb + to clause fragment	according to the
	Anticipatory it + VP	it will be
	Verb + Noun + of	take care of, take advantage of
	Verb + Participle Adjective + Preposition	stay focused on
	Pronoun / NP / DET + be	there are also, there is no, there are some, there are still many people
Others	Adverbial clause fragment	as we know, if you want to
	S + V fragment	I agree with, we must be able to
	Other structures	as long as, not only that, due to the, and so on, as efficiently as possible, as a result

As seen in table 6, NP + of-phrase fragment has more variants compared to other structural categories. This structural pattern, along with "in + the + Noun + of" (PP-based), are called fixed-frames (Biber et al., 2003).

4.2 The Frequency of Bundles

This section deals with the use of the lexical bundles identified in the learner corpus. The focus of analysis was divided into the frequency of each bundle, structure, and n-gram. Concerning the individual bundles, certain bundles are more frequently used in the students' argumentative essays. Those bundles are listed in table 7.

Table 7: Frequency of lexical bundles

Position	Lexical bundles	Absolute frequency	Percentage	Dispersion (*DP*)
1	the use of	105	9.34	0.62
2	a lot of	96	8.54	0.55
3	in the future	65	5.78	0.73
4	be able to	61	5.42	0.68
5	the quality of	43	3.82	0.78
6	in my opinion	39	3.46	0.80
7	the development of	33	2.93	0.79
8	one of the important	30	2.66	0.85
9	in addition to	28	2.49	0.84
10	it will be	25	2.22	0.85

Table 7 shows that the most frequent bundle is the NP-based bundle 'the use of' (105), followed by 'a lot of' (96), 'in the future' (65), 'be able to' (61), 'the quality of' (43), and some other bundles, such as 'in my opinion' (39) and 'in addition to' (28).

It is not surprising that NP-based bundle 'the use of' is frequent since it can be used as a pre-modifier in an NP, and it states something common in daily life (i.e., *the use of* x). the most frequent bundle 'the use of' can also be found in AFL. It is listed in core bundles in spoken and written academic English (Simpson-Vlach & Ellis, 2010). Meanwhile, 'a lot of' is a quantifier of things used frequently, and 'in the future' can serve as the time marker for predictive expressions. In the fourth position, there is 'be able to'. Lexical bundle 'be able' to is a quasi-modal expressing ability or possibility (Collins, 2014; Collins, 2009), and modal verbs are ubiquitous in language use across text genres and registers. In AFL, 'be able to' is also included as one of the academic bundles, although it is under the spoken bundle category (Simpson-Vlach & Ellis, 2010). The following frequent bundle, the fifth position, is 'the quality of', and its function is a pre-modifier of NPs. Hence, we can notice that the most frequent bundles serve general purposes, e.g., as pre-modifiers of NPs or as predictive expressions.

Based on the findings, the five most frequent bundles are fairly general bundles and seem to have nothing to do with the text type. Those bundles are not strictly used to fulfill the communicative function of the text. However, the sixth and the ninth frequent bundles are related to argument delivery, e.g., 'in my opinion' and 'in addition to'. For this context, it is evident that the communicative function of the

text plays a significant role in language feature selection (Biber & Conrad, 2009). An argumentative essay is an essay in which you agree or disagree on an issue, using reasons (and evidence) to support your opinions. Your goal is to convince your reader that your opinion is correct (Oshima & Hogue, 2014). This text type is transactional, meaning the main aim is to convey some information (Jackson & Stockwell, 2011). The bundle 'in my opinion' marks the writers' argument, while 'in addition to' gives additional information that may benefit the opinions.

In terms of structural taxonomy, the most frequent bundle is NP-based (39,8%), while VP-based is in the second place with 24% and PP-based is 18,8%. In the *others* category, there are 17,4% bundles of the corpus.

Table 8: Frequency of bundles based on the structural taxonomy

Position	Structural taxonomy	Percentage
1	NP-based	39,8
2	VP-based	24
3	PP-based	18,8
4	Others	17,4

NP-based bundles are frequent because they are pre-modifiers of NPs needed in writing, such as 'the use of', 'the quality of', 'the development of'. In terms of lexical categories, nouns are essential in building up linguistic constructions and developing discourse.

In terms of word combination (n-gram), each n-gram has its frequency of use. As seen in table 10, 3-word bundles occupy 86% of the lexical bundles in the corpus, followed by 4-word bundles (11%), and the least frequent one is 5-word bundles.

Table 9: Frequency of bundles based on the n-gram

Position	n-gram	Percentage (%)
1	3-word bundles	86
2	4-word bundles	11
3	5-word bundles	3

Table 9 shows that the longer the word combination, the less frequent the bundles found in the corpus. The lesser frequency is due to the compactness and efficiency since it is less effective for language users to construct long word combinations. This phenomenon refers to *the principle of least effort* as human nature (Zipf, 1949).

To provide a comprehensive analysis, the bundles identified in this study (in the learner corpus) were compared to those bundles in AFL. The criteria for comparison to be applied here are (i) whether the bundles of the students' essays are listed in the AFL and (ii) whether those bundles are categorized for spoken or written text). The former is important to investigate the urgency of the bundle, and the latter is for the accuracy of usage context.

Table 10: Shared lexical bundles in the learner corpus and AFL

No	LB in learner corpus	LB in AFL	Register of use (AFL)
1	the use of	the use of	spoken & written
2	be able to	be able to	spoken
3	at the same time	at the same time	spoken & written
4	and so on	and so on	spoken & written
5	according to the	according to the	spoken & written
6	if you want to	if you want to	spoken
7	we have to	we have to	spoken
8	there is no	there is no	spoken & written
9	the development of	the development of	spoken & written
10	the importance of	the importance of	spoken & written
11	(that) can be used	(that) can be used	spoken & written
12	the number of	the number of	spoken & written
13	the existence of	the existence of	spoken & written

Table 10 presents the shared bundles of the learner corpus and AFL. In terms of the registers, some bundles are correctly used by the learners. Meanwhile, several bundles found in the students' essays are categorized as the characteristics of spoken registers in AFL, e.g., *be able to, if you want to,* and *we have to.* These shared bundles, along with the register of use, prove that, despite the various structures used by the learners, they still misuse the bundles.

5. DISCUSSION

Based on the findings of this study, it is clear that the corpus comprises various structural types of lexical bundles, recurring from 3- to 5-word sequences. The results also demonstrate that NP-based bundles are the most frequent structural category, followed by other structural types (VP-based, PP-based, and *others*). Furthermore, the use of the bundles decreases with the increase of the word combination of the bundles. This study shows that the most frequent bundle structure is the NP-based bundles, which confirms the previous research conducted by Fitrianasari et al. (2018). Their study proves the students or learners the most frequently use NP-based bundles. In addition, Navarro Gil & Martínez Caro (2019) state that NP-based bundles have the highest frequency of all bundles, either in learners' writing or professional writing. It shows that NP-based bundles are prominent in word sequences, which this study has demonstrated. Besides students' writing, another study that confirms a similar result is Hye-Kyung Lee (2020), claiming that NP-based bundles are the most frequent ones in Linguistics textbooks. The results of these studies emphasized the importance of NP-based bundles in writing as well as in textbooks.

On the contrary, Yang (2017) shows that the most frequent lexical bundles in students' writing are the VP-based bundles. This result confirms Shin's (2018) study,

proving that VP-based bundles are prominent in students' writing. Both studies offer some proof related to the structure of bundles in students' writing. Moreover, Shin (2018) reveals that the results for native writers are the same in which VP-based bundles are the most frequent in their writing. The results of those studies and this study show that each corpus might result in distinct findings, affected by the learners' proficiency, the text types, and the design of the learner corpus compilation. Nonetheless, it is noticeable that both NP-based bundles and VP-based bundles are highly frequent in students' writing.

Regarding the number of the bundles, the n-gram, this study emphasizes that 3-word bundles are the most frequent than other word combinations. It confirms the findings of the previous studies, such as that done by Lou (2012), Navarro Gil & Martínez Caro (2019), Ulfa & Muthalib (2020). Lou (2012) points out that along with the increase of word combinations, the frequency of the bundles is decreasing. Navarro Gil & Martínez Caro (2019), who investigate the 3-, 4- and 5-word bundles, also figure out that the most frequent bundles are those with 3-word combinations. Ulfa & Muthalib (2020) also examine 3- and 4-word bundles, claiming that 3-word bundles are more frequent than 4-word bundles. These study results are in line with the statement from Cortes (2013), arguing that the frequency of lexical bundles decreases as they contain more words.

This study also shows that some bundles are shared by learners of English and the native corpus, which corresponds to the studies of (Hye-Kyung Lee & Hyeon-Okh Kim, 2017; Lou, 2012; Shin, 2018). Those studies demonstrate that, despite the differences in lexical bundles (types and use), EFL learners and native writers share the lexical bundles in the written register. Lou (2012) identifies the shared patterns of lexical bundles between native and non-native students, that is, *personal pronoun + lexical verb phrase (+complement clause) and (auxiliary +) active verb*. This bundle, however, is used in the spoken register by the native students. Hye-Kyung Lee & Hyeon-Okh Kim (2017) also investigate the shared bundles between learner corpus and native corpus, including some of the following bundles, such as *is one of them, at the same time, on the other hand,* and some other lexical bundles. In terms of argumentative essays, Shin (2018) also proves that there are some shared bundles between native and non-native learners, e.g., *on the other hand, when it comes to, disagree with the statement*. These shared bundles prove that EFL learners have adequate lexicon storage of lexical bundles, regardless of the differences claimed by other studies.

This study demonstrates that there are enormous structural classifications of lexical bundles in the corpus. The various patterns in the classification prove that EFL learners at the university level use various structures of word sequences in their writing. This finding corresponds to the study results conducted by Wachidah et al. (2020), who also found 12 structural forms of lexical bundles proposed by Biber et al. (1999) in their study. Their study and this study confirm that the learners are

practically able to use various structural types of lexical bundles. In this study, specifically, the learners predominantly use fixed-frames, i.e., NP + of-phrase fragments (NP-based) and Prep + NP fragments (PP-based), which indicate the invariable structures in the individual structural type. Considering that NP-based bundles are the most frequent, there should be more types of bundles, e.g., the NP + to-infinitive fragment and other NP fragments.

In addition, this study shows the various distribution of VP-based bundles with a small proportion, different from what has been found in NP-based and PP-based bundles. Thus, apart from multiple structures of lexical bundles, it seems that the learners still tend to use specific structures or patterns. It means that no matter how many patterns of lexical bundles are found, the use can be categorized as invariable. This is similar to the findings of the previous studies (e.g., Ädel & Erman, 2012; Fajri et al., 2020; Samodra & Pratiwi, 2018). Samodra & Pratiwi (2018) claim that the discrepancies of lexical bundle usage between native and non-native writers are related to the limited vocabulary of the non-native writers. Meanwhile, the result of Ädel & Erman (2012)'s study indirectly indicates that the lexical bundles of the non-native writers are less varied. Native writers have a more significant number of types of lexical bundles in comparison to the use of bundles among non-native writers (Ädel & Erman, 2012). Fajri et al. (2020) argue that bundles are more frequently used by L2 writers than L1 writers, although the L2 learners' bundles are sometimes underused. The findings of these studies follow the statements from Meunier (2012), Granger (2018), Paquot & Granger (2012), showing that there are discrepancies of lexical bundle usage between English native speakers and EFL learners.

Although there are shared bundles found in the learner corpus and AFL, what is interesting is that some lexical bundles are identified as spoken bundles in AFL but are used in the students' essays (e.g., *we have to, if you want to, be able to*). This finding corresponds to what has been found by Lou (2012), Hye-Kyung Lee & Hyeon-Okh Kim (2017), and Samodra & Pratiwi (2018). Lou (2012) states that the most frequent structure of bundles in students writing is used in the spoken register of the native learners. Hye-Kyung Lee & Hyeon-Okh Kim (2017) point out that the writings lack academic writing features, indicating the incorrect use of registers, especially lexical bundles. Samodra & Pratiwi (2018) argue that the only difference in lexical bundles used by native and non-native writers is related to the choice of bundles for academic English. These validate the discrepancies of bundles used by native and non-native writers of English, which need to be revisited. In terms of language use, the different modes of communication and other texts might influence the selection of words and structures (Biber & Conrad, 2009).

6. PEDAGOGICAL IMPLICATION

The results of this study can contribute to the teaching of lexical bundles in writing courses. There are several remarkable points that we should take into consideration for pedagogical aspects. Regarding the structure, these findings show that various

and numerous patterns are identified; however, the dominance of a particular structure is vigorous, causing other structural variants to be less significant. The learners seem to rely much on fixed-frame bundles indicating the need to introduce different patterns which are less frequent and the patterns that are used monotonously. Therefore, it is crucial for the learners also to be familiar with all structures proportionally.

Dealing with language use is not only about patterns, but it is also necessary to consider the appropriate use based on the register or text type. The learners should be able to use bundles that are commonly used for written registers more frequently. To this extent, the learners should be well informed about some word sequences that tend to be used in writing and differentiate them from those used in spoken contexts. There are some ways to achieve this, such as getting the learners familiar with different word sequences for different registers (i.e., spoken and written) and text types (e.g., academic vs. narrative). The ability to differentiate the use of bundles in their appropriate context can show the learners' level of proficiency. Therefore, both the writing course instructors and the materials should be register-friendly, providing examples of appropriate usage of lexical bundles.

Regarding the frequency of the structure, this study demonstrates that the learners predominantly use the NP-based bundles in writing their essays, followed by the VP-based and PP-based bundles. Although NP-based bundles are highly frequent, the structures tend to be homogenous with many fixed-frame structures. The result is the same as in the PP-based bundles in which the fixed-frames outnumber other PP-based structures. Thus, it is important to design teaching materials providing enormous exemplary patterns with various patterns. Moreover, the previous studies show that the most frequent bundles are VP-based bundles which contradict the findings of this study and should also be highlighted when designing writing course materials. The use of lexical bundles should be more varied and less repetitive; thus, introducing as many lexical bundles as possible to learners is beneficial.

Apart from the structure and the frequency, this study examines lexical bundles in argumentative essays, which means the bundles are expected to serve as argument building or delivery. Nevertheless, the results show that only three bundles are strongly related to the delivery of argument, e.g., *in my opinion, I agree with, according to (the).* The other two bundles are the combinations with ability/necessity/obligation modals, such as *we have to, must be able to* which are related to suggestion or advice. The rest of the bundles are more general and less discourse-sensitive, indicating that the learners cannot differentiate language use for different texts. Teaching materials for spoken and written language should consider this aspect so the learners can be communicatively competent in any kind of real-life setting.

These pedagogical implications should be supported by appropriate methods to internalize word sequences among learners. Cutler (2021) suggests that there are two

ways of new sequences internalization, (1) learned as a whole and (2) learned formulaic by regular usage to join the components into a single whole. The internalization of lexical bundles among learners should be cultivated earlier in their first year of university by incorporating lexical bundles in the teaching materials and also by considering several things about lexical bundles, such as (1) introducing various lexical bundles, (2) designing materials with register-based lexical bundles and text-based lexical bundles.

7. CONCLUSION

This study has proven that EFL learners use lexical bundles in their writing, and some of the bundles are shared in the native corpus (AFL). Regarding the structure, the bundles vary, although certain structures outnumber other structures. In addition, some lacking aspects are also noticeable, especially concerning the variants and the register of use. These shortcomings, then, can be revisited and highlighted in the designing and developing of writing course materials. Writing instructors should familiarize the learners with the written register (e.g., academic writing) and the register's lexical bundles. Another point to consider is the importance of other structure types of lexical bundles, such as VP-based and PP-based bundles. In some studies, VP-based bundles are the pivot structures; thus, it is also necessary to facilitate the learners with various VP-based bundles in the writing materials. Writing course materials should equip learners with formulaic sequences to enhance their writing quality. However, this study still focuses on the learner corpus itself and compares it briefly to the AFL. This study, unfortunately, does not make any comparison with native corpus appropriately (as control corpus), which is why this study has to be followed up by future research. Therefore, to achieve a more rigorous understanding of this issue, future studies can compare with a native corpus (e.g., *British Academic Written English Corpus* [BAWE]). Future studies can also compile a bigger learner corpus comprising more than a thousand words of tokens and investigate the argumentative essays and other types of essays, such as narrative essays.

8. 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

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. J. Benjamins.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*(3), 263–286.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–190). Rodopi.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, G. Leech, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the Lune: A festschrift for Geoffrey Leech* (pp. 71–93). Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371–405.

Biber, D., & Egbert, J. (2018). *Register Variation Online* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316388228

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.

Brezina, V., Timperley, M., & McEnery, T. (2018). *#LancsBox* (v.4.x.) [Computer software]. http://corpora.lancs.ac.uk/lancsbox.

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.

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

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.

Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press.

Chomsky, N. (2015). *Aspects of the theory of syntax* (50th Anniversary Edition). The MIT Press.

Collins, P. (2014). Quasi-modals and modals in Australian English fiction 1800-1999, with comparisons across British and American English. *Journal of English Linguistics*, *42*(1), 7–30. https://doi.org/10.1177/0075424213512857

Collins, P. C. (2009). *Modals and quasi-modals in English*. Rodopi.

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

Cortes, V., Jones, J., & Stoller, F. (2002). *Lexical bundles in ESP reading & writing*. TESOL Conference.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, amt056. https://doi.org/10.1093/applin/amt056

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, *34*(3), 369–388.

Cutler, S. F. (2021). Path to formulaicity: How do L2 speakers internalise new formulaic material? In A. Trklja & Ł. Grabowski (Eds.), *Formulaic language: Theories and methods* (pp. 81–112). Language Science Press.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165–193.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, *20*(1).

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.

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.

Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230355569

Flowerdew, L. (2014). Learner corpus research in EAP: Some key issues and future pathways. 영어학연구, *20*(2), 43–60.

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.

Granger, S. (2018). Formulaic sequences 1 in learner corpora. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language* (1st ed., pp. 228–247). Routledge. https://doi.org/10.4324/9781315206615-13

Gray, B., & Biber, D. (2015). Phraseology. In *The Cambridge handbook of English corpus linguistics* (pp. 125–145). Cambridge University Press.

Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*(4), 403–437.

Groom, N. (2009). Immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 21–33). Palgrave Macmillan.

Hewings, M. (2010). Materials for university essay writing. In N. Harwood (Ed.), *English language teaching materials*. Cambridge University Press.

Hye-Kyung Lee. (2020). Lexical bundles in linguistics textbooks. 언어연구, *37*(1), 121–145. https://doi.org/10.17250/KHISLI.37.1.202003.005

Hye-Kyung Lee, & Hyeon-Okh Kim. (2017). Features of lexical bundles in academic writing by non-native vis-à-vie native speakers of English. 영어학연구, *23*(2), 67–88. https://doi.org/10.17960/ELL.2017.23.2.003

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4–21.

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.

Jackson, H., & Stockwell, P. (2011). *An introduction to the nature and functions of language*. Continuum.

Jalali, Z. S., & Moini, M. R. (2014). Structure of lexical bundles in introduction section of medical research articles. *Procedia - Social and Behavioral Sciences*, *98*, 719–726. https://doi.org/10.1016/j.sbspro.2014.03.473

Kaur, A., Singh, S., & Ang, L. H. (2019). Lexical bundles in selected children's fiction: A corpus-based analysis. *Journal of Creative Practices in Language Learning and Teaching*, *7*(2), 12–28.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2015). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, amv070. https://doi.org/10.1093/applin/amv070

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1).

Lou, X. (2012). Structural analysis of lexical bundles in EFL English majors' theses of an ordinary normal university in China. *International Journal of Applied Linguistics & English Literature*, *1*(6), 142–153.

Mei, W. S. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students' argumentation. *RELC Journal*, *37*(3), 329–353.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, *32*, 111–129. https://doi.org/10.1017/S0267190512000128

Nasrabady, P., Elahi Shirvan, M., & Ehsan Golparvar, S. (2020). Exploring lexical bundles in recent published papers in the field of applied linguistics. *Journal of World Languages*, *6*(3), 175–197.

Navarro Gil, N., & Martínez Caro, E. (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., & Adnan, A. (2020). A corpus study of verbs in opinion articles of The Jakarta Post and the relation with text characteristics. *English Language Teaching Educational Journal*, *3*(2), 108.

Oktavianti, I. N., & Ardianti, N. R. (2019). A corpus-based analysis of verbs in news section of The Jakarta Post: How frequency is related to text characteristics. *JOALL (Journal of Applied Linguistics & Literature)*, *4*(2), 203–214.

Oshima, A., & Hogue, A. (2014). *Longman academic writing series: Essays*. Pearson Education.

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.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149.

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

Samodra, M. C., & Pratiwi, V. D. R. (2018). Lexical bundles in Indonesian and English undergraduate thesis abstracts. *Proceedings of the Fourth Prasasti International Seminar on Linguistics (Prasasti 2018)*. Fourth Prasasti International Seminar on Linguistics (Prasasti 2018), Surakarta, Indonesia.

Shin, Y. K. (2018). *Lexical bundles in argumentative essays by native and nonnative English-speaking novice academic writers* [Ph.D. Dissertation]. Georgia State University.

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214–225.

Szudarski, P. (2017). *Corpus linguistics for vocabulary: A guide for research* (1st ed.). Routledge. https://doi.org/10.4324/9781315107769

Ulfa, N., & Muthalib, K. A. (2020). Lexical bundles in students' essay writing. *English Education Journal*, *11*(3), 367–379.

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.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, *20*(1), 1–28.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, *54*, 1–14.

Yang, Y. (2017). Lexical bundles in argumentative and narrative writings by Chinese EFL learners. *International Journal of English Linguistics*, *7*(3), 58.

Zipf, G. (1949). *Human behaviour and the principle of least effort*. Addison-Wesley Press, Inc.