ENGLISH LEXICAL BUNDLES IN THE GRADUATE THESES: THE FREQUENCY, STRUCTURE, AND DISTRIBUTION

Mohamad Syihabuddin Faqih[1] & Ignatius Harjanto[2]
Universitas Katolik Widya Mandala Surabaya, Indonesia
[1]sihabfaqih@gmail.com; [2]harjanto@ukwms.ac.id.

Abstract: Lexical bundles are one of the important characteristics of academic discourse which tell readers to know whether the writer is professional or novice. Inevitably, studies on lexical bundles in scientific essays are important to do. This study identifies the most frequent, structural characteristics, and the functional categorization of lexical bundles in the Master Theses in Teaching English as a Foreign Language (TEFL), specifically in the Findings and Discussion section. There were 651.083 words from 74 different theses compiled to create the corpus by using Antconc 3.5.8. The results found 117 different lexical bundles and the sequences 'the result of the' and 'on the other hand' dominate the section. Noun phrase + *of* structure which covers one third of overall forms in the corpus were the most lexical bundles' structural types in the findings and discussion section followed by other noun phrase structures (22% out of overall bundles). Functionally, research-oriented bundles (45% of overall bundles) were the most frequent ones followed by text-oriented (40%) and the least frequent bundles were participant-oriented. Reported findings are further discussed with related theories.

Keywords: Lexical bundles, corpus linguistics, findings and discussion section, graduate theses

INTRODUCTION

The development of vocabulary studies has resulted in an extended concern into groups of words instead of individual word. Collocation in the textbooks (Alfiandita & Ardi, 2020) and academic

writing (Thongvitit & Thumawongsa, 2017) had been done. Further similar studies dealt with multi-word sequences called "lexicalized sentence stem" as introduced by Pawley & Syder (1983) and "lexical phrase" studies (Çandarlı, 2020; Li & Schmitt, 2009; Ngadiman, 2013).

To date, several studies on multi-word sequences used another term called lexical bundles (Biber et al., 1999) . Previous studies on lexical bundles were comparative studies used by natives and non-natives of English (Amirian et al.: 2013, Esfandiari & Barbary: 2017, and Novita & Kwary: 2018). Other researchers focused on identifying lexical bundles in several sections of academic theses: abstract, introduction, theoretical framework, data analysis (findings and discussion), and conclusion. Guiling (2015) studied word sequences in the introduction section of International and Chinese students research theses. Samodra & Pratiwi (2018) identified three and four-word combinations in undergraduate thesis abstracts of Indonesian and English.

Wachidah et al. (2020) used non-corpus approach to investigate word combinations that occurred in the Findings and Discussion Section of ten randomized graduate theses. Manually, the data were taken from 10 graduate students' theses. The result showed that other prepositional phrase was the most frequent type in the Findings and Discussion section of students' theses. In addition, text-oriented function of lexical bundles was identified as the biggest portion which focused on the information (meaning) in a text.

Findings on lexical bundles in academic writing were used as learning material in teaching writing. For second and foreign language learners, various kinds of lexical bundles contributed to develop a coherence text (Hyland, 2008). Indeed, learners need to know lexical bundles because of its wide range of discourse meaning and functions (Wright, 2019). Learners need to use correct lexical bundles in their academic essays to maintain the comprehensibility of the texts. In other words, expert academic writers used lexical bundles properly.

In theses, students can use lexical bundles based on its structure and function, which may vary in every part or chapter. Functionally, research-oriented lexical bundles are the biggest portion in the abstracts of student theses (Samodra & Pratiwi, 2018). Meanwhile, in the different part of student theses, Findings and Discussion section, Wachidah et al. (2020) concluded that text-oriented function is the most common lexical bundles based on physical data collection.

Identifying lexical bundles with manual data collection could be difficult and possibly "idiosyncratic" (Biber & Barbieri, 2007, p. 268). Therefore, specific criteria for the frequency and range of dispersion of lexical bundles need to be adopted. This study used Biber et al. (1999) criteria of lexical bundles to give further understanding about the distribution of lexical bundles on the Findings and Discussion section of graduate students' theses, which may vary in terms of forms, structures and functions by using corpus data. The data were taken from master program theses of English Education Department. An online resource provided the data to compile a corpora through an institutional repository. As different parts of academic essays rely on different lexical bundles, this corpus study was presumed to give insights about the use of lexical bundles in the finding and discussion part of academic essays.

LITERATURE REVIEW
Formulaic Language

Formulaic language has been an interesting topic of investigation in the last century. In a conservative perspective, the term formulaic language has been restricted into vocabulary units which have grammatical structure or phrasal lexical chunks in oral and written form. The formulaic sequences include things such as idioms, collocations, phrasal verbs, other multiword expressions, and lexical bundles. Further questions raised are the issues on identification of formulaic language, acquisition process, and the function in speech and writing.

There is no single consensus about the definition of formulaic language. However, Wray & Perkins (2000) provide us about the identification of formulaic language as the following:

> A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 1)

Based on the definition, it can be noticed that formulaic sequences are characterized as its 'wholeness' or 'unity'. In more details, the acquisition, memorization, and retrieval processes of sequences are holistic. The process seems spontaneous and simple because it is "stored and retrieved from memory". The cognitive process of language production does not require additional grammatical processing.

There are two perspectives to understand the term formulaic language by its continuum: open and closed (Liu, 2014). In some literatures, formulaic language is called idiomatic and non-idiomatic. Wray & Perkins (2000) stated that formulaic language is a phenomenon which covers numerous forms of word sequences. On the contrary, closed formulae are indicated by idioms or phrases that originally taken from its semantic and syntactic life (Liu, 2014). Idioms have their fixed meaning and form, while the other multi-word expressions can be varied based on their types.

There are several functions of formulaic expression. Schmitt (2005) identified the importance of formulaic language to particular functions in the field of language use. The first function is related to its functional use, such as in apologies, compliments, directions, and complains. Formulaic language can also be used as social interaction to maintain social relationship. In discourse organization, formulaic phrases are used as connectors in spoken or written discourse, such as 'on the other hand', 'in other words', and 'as a conclusion'. The last function of formulaic language is to precise information transfer.

The use of formulaic language has been investigated under several terminologies, including "`lexical phrases', `formulas', `routines', `fixed expressions', `prefabricated patterns' (or `prefabs'), and `lexical bundles'"(Biber et al., 2004, p. 372). Lexical bundles have become the newest and interesting topic to be discussed because their roles in the field of writing discourse and ESP (English for Specific Purpose) material development are important. Indeed, those multi-word sequences are used for distinctive functions in academic discourse. For example, AlHassan & Wood (2015) found that multi-word sequences are needed to facilitate second language learners to create proper structure and more systematic academic written essays.

Lexical Bundles

As an important building-block to create a cohesive-essays, lexical bundles need to be placed in the spotlights. Lexical bundles are characterized as intermittent sequences of at least three words that appear frequently in the texts (Biber et al., 1999). In addition, Byrd & Coxhead (2010) stated that lexical bundles are the sequence of three words (minimum) or more that are repeated constantly for a specific time. Based on the definition, lexical bundles can be distinguished with other word-sequences based on their frequency and range of dispersion criteria. To be qualified as lexical bundles, the word-sequence needs to occur "at least 10 times per million words" (Biber et al., 1999, p. 990) and it needs to appear in (at least) five texts.

Both two characteristics of lexical bundles can be identified automatically by using computer software called concordance. Concordance program is commonly used to find a variety of options on the menu, find a word, group of words, and other structures. The concordance results from the program can further be analyzed (for example) to identify lexical bundles, part of speech (verbs, adverbs, and adjectives) in product review (Altun, 2019), or to find hedging in the spoken discourse (Nuraniwati & Permatasari, 2021).

Four word sequences (four word bundles) are the most bundles' length to be identified in various research. Word combination can be

in the form of 2-, 3-, 4-, 5-, and 6-word units. Four word units have become the main focus of the research because they "offer a clearer range of structures and functions" (Hyland, 2008, p. 8) than the shorter word units. In other literature, 5- or 6-word units (longer sequences) sometimes are formed from 4-word units (Biber & Barbieri, 2007).

Lexical bundles can be organized by their structural types and functions in discourse. Biber et al. (1999) assumed that lexical bundles have solid grammatical correlation and made categorizations that group them into a few fundamental structural taxonomies. On the basis of structural classification, lexical bundles can be split into three basic structures: noun phrase, prepositional phrase, and verb phrase-based structure. Following Biber et al. (1999), Hyland (2008) made general structural (grammatical) types of lexical bundles in academic essays. Common lexical bundle structures are noun phrase, prepositional phrase, verb phrase, and other bundles.

Lexical bundles can also be analysed by looking for its essential role in the discourse. Hyland's functional distribution pattern was adopted to be specifically valuable for the current study because it is adjusted to the particular consideration of academic writing genres (Hyland, 2008). This classification subsequently compiles lexical bundles into three broad focuses of investigation: "research, text, and participants, and introduces sub-categories which specifically reflect the concerns of research writing" (p. 13). While text-oriented bundles focus on text-organization, research-oriented bundles assist writers in organizing their real-world activities in research. The other functional classification proposed by Hyland (2008) is participant-oriented bundles, the sequences that concern with the reader or writer of the text.

A series of studies explored lexical bundles in speaking and writing (Biber & Barbieri, 2007; Biber & Conrad, 1999; Biber et al., 2002). They found the different bundles' use between speakers and writers to build the specific discourse. Based on Biber et al. (1999) study, lexical bundles in academic writing were less than 5% which

are completely structural. It means that most of the bundles in academic writing were incomplete structural units.

Focused on academic writing, some studies tried to contrast the different use of lexical bundles in the different subjects of academic essays (Byrd & Coxhead, 2010; Hyland, 2008; Kwary et al., 2017). They found that there were some distinctive uses of lexical bundles across different fields of study. The findings showed that the use of stance bundles occurred frequently in social science (soft science) academic writing. The writer of social science article use stance bundles more frequently to express their opinion rather than accredited information.

Some other studies explored the variations of lexical bundles used by the first and second language writers (Chen & Baker, 2010; Gungor & Uysal, 2016). Both literatures suggest that L2 writers used lexical bundles more often than L1 as discourse organizers, organizational signals to maintain the text. However, L1 writers used more variation structure of lexical bundles.

Other literature focused on identifying lexical bundles in several sections of academic theses. Guiling (2015) identified word sequences in the introduction section of International and Chinese students' research theses. The findings showed the greater amount use of lexical bundles by Chinese students with the noun-based structure dominating the form. Students used lexical bundles as discourse organizers and referential expressions in the introduction part. On the other side, Samodra & Pratiwi (2018) identified three and four-word combinations in undergraduate thesis abstracts of Indonesian and English. They found that research-oriented bundles dominated the text. In addition, Wachidah et al. (2020) found that text-oriented was the most frequent lexical bundles in Findings and Discussion section of ten graduate students theses.

METHOD

This study used corpus-based approach, the analysis of collections of texts or corpora (Altun, 2019). One of the main

outcomes from a corpus-based approach is that "descriptions of grammatical variation and use" (Heine et al., 2015, p. 4) in different registers. In this study, collections of texts from the written register (Findings and Discussion section of graduate students' theses) were analyzed to identify lexical bundles. There were 74 collections of Findings and Discussion section of Master Thesis in Teaching English as a Foreign Language (*TEFL*) of a private university in Surabaya. These data sources were chosen for compiling the thesis section. The Corpus of Master Theses in TEFL was created using Antconc 3.5.8 version. After collecting the bundles, structural characteristics and functions of lexical bundles were analysed.

Table 1 Master Theses in Teaching English as a Foreign Language (MTEFL) Corpus

Year	Total Thesis	Word Tokens
2015	8	99084
2016	15	96118
2017	20	156826
2018	9	104540
2019	21	173365
2020	1	21150
Total	74	651083

Following Hyland (2008), four-word sequences of lexical bundles were retrieved because they provide clearer lexical patterns and functions. The selection of lexical bundles process regards the frequency and range. The frequency cut-off followed Hyland's (2008) guidelines, 20 times per million words because of relatively small corpus and distributed across 10% of different texts. MTEFL Corpus consists of 74 texts, thus the range of dispersion can be set in 7 (10% of 74 texts). On the other hand, based on the total words in MTEFL Corpus (around 650.000), the x frequency was (20*650.000)/1.000.000. For 650.000 words compiled in the corpus, the frequency was 13. Hence, in this research, a lexical bundle was identified when it had 13

minimum of frequency and it could be found in seven different theses.

After extracting lexical bundles by using Antconc, the researchers conducted several stages to analyze the data. In the first step, classifying the structures of lexical bundles by Hyland (2008). In the next step, the researchers classified the functions of lexical bundles according to Hyland's taxonomy (2008). The last step was interpreting and explaining the data based on research and discussing the findings by some relevant theories and previous related findings.

FINDINGS

Table 2. The Most Frequent Bundles in MTEFL Corpus

Freq	Range	Lexical bundles	Freq	Range	Lexical bundles
157	41	the result of the	83	10	by the students in
150	39	on the other hand	76	25	it can be concluded (that)
139	16	(teacher) asked the students to	74	18	the end of the
124	30	the results of the (data)	73	23	most of the students
116	13	the total number of	66	7	the mean score of
103	30	it can be seen (that/in/from)	63	25	to be able to
95	11	can be found in	57	21	the meaning of the
94	17	students were able to	55	21	in line with the
90	16	in front of the	51	22	it was found that (the)
87	24	in the form of	51	9	the students were asked (to)

A list of 117 four-word concordances was identified as lexical bundles with the minimum cut-off frequency of 13 and 7 text dispersion ranges. Antconc identified lexical bundles across MTEFL Corpus and presented the list that had been sorted automatically from the most frequent bundles into the lowest one. Table 2 describes the details of the most frequent four-word bundles used by graduate

students of TEFL. It indicates that graduate students' thesis, especially in the Findings and Discussion section, contain various types of bundles and the most frequent lexical bundles.

Following Hyland (2008), structural types of lexical bundles were divided into eight categories: noun-based (noun phrase + *of* and other noun phrases), prepositional-based (preposition phrase + *of* and other prepositional phrases), verb-base (passive verb + prepositional phrase fragment, *it* + verb or adjective, *be* + noun or adjective), and other bundles. Table 3 give information about the overall distribution of lexical bundles in the corpus.

Table 3. Structural Types of Lexical Bundles in MTEFL Corpus

Structural Types	Freq	%	Lexical bundles
Noun phrase + of	1117	29%	the result of the
			the total number of
Other noun phrases	844	22%	most of the students
			the teaching and learning
Prepositional phrase + of	391	10%	in front of the
			in the form of
Prepositional phrases (others)	745	20%	on the other hand
			in line with the
Passive form + prep.- phrase fragment	161	4%	can be found in
			could be seen from
Anticipatory it + verb/adj.	342	9%	it can be seen (that/in/from)
Be + noun/adj. phrase	83	2%	is one of the
Others	123	3%	as shown in the

With regard to lexical bundles' function in discourse, Table 4 displays the distribution of functional bundles in Chapter Four, Findings and Discussion section of master students' theses in TEFL.

Table 4. Functional Distribution of Lexical Bundles in MTEFL Corpus

Function	Types	Freq	%	Lexical Bundles
Research-oriented	Location	280	7%	in front of the the end of the
	Procedure	175	4%	they were asked to the teacher asked the (students)
	Quantification	603	14%	the total number of most of the students
	Description	705	17%	the meaning of the the difference between the
	Topic	116	3%	the teaching and learning (process) the use of English
Text-oriented	Transition signals	245	6%	on the other hand in order to be
	Resultative signals	405	10%	the result of the the researcher found that
	Structuring signals	688	17%	it can be concluded (that) as shown in the
	Framing signals	345	8%	in the form of in line with the
Participant-oriented	Stance	434	10%	students were able to it is important to
	Engagement	169	4%	it can be seen (that/in/from)

DISCUSSION

Within 74 Findings and Discussion section of students' graduate theses, 117 different lexical bundles were retrieved. Most of the bundles were not complete unit systems. As Csomay (2013) stated lexical bundles are not structural units or settled expressions but they can be classified based on discourse capacities. Lexical bundles have their own function in a certain discourse. Some examples such as 'in the middle of' and 'at the beginning of' are prepositional phrase patterns that can be the bridge for the other structural unit. In

addition, the bundles such as 'the purpose of the' and 'the meaning of the' are noun phrase patterns of which the last words are the first component of another structural unit.

'The result of the', with the frequency of 157 in the corpus (around twelve times more than the frequency cut-off at 13) was the highest rank of lexical bundle in the MTEFL Corpus. This bundle was widely used in the Findings and Discussion section in pairs with some other words such as 'interview', 'observation', 'questionnaire', and 'test'. On the other hand, the plural version of the bundle, 'the results of the' appeared in 30 texts with the frequency of 124 times. With the high frequency and range of dispersion, 'the result of the' has been used in both qualitative and quantitative research. This bundle often serves the purpose of introducing the result of the data collection through a qualitative and quantitative approach.

Previous research about lexical bundles in academic prose (Hyland, 2008) showed that the bundle 'the results of the' (with –*s*) was included in the list of most frequent bundles in four fields of studies (Applied Linguistics, Business, Biology, and Electrical Engineering). The bundle occurred in four disciplines with high frequency from the corpus which consists of 3.4 million words.

Another most frequent bundle was 'on the other hand' which appeared in 39 different texts. It occurs 150 times in the corpus and places as the second of the most frequent lexical bundle in MTEFL corpus. In BAWE-CH (British Academic Written English by Chinese L1) Corpus and FLOB (Freiburg-Lancaster-Oslo/Bergen) Corpus, Baker & Chen (2010) also identified 'on the other hand' as the most common bundle. This bundle is important as a discourse organizer to show the contrast between sentences with the previous one. In the Findings and Discussion section, 'on the other hand' was frequently used to show the contrast between research findings. This bundle was used to be a contrastive marker in the beginning of the sentence.

From the most frequent lexical bundle findings, four-word sequences in the list were common in academic writing. This finding supported previous research on lexical bundles in academic discourse

(Bal, 2010; Chen & Baker, 2010; Hyland, 2008). In the Findings and Discussion section of students' graduate theses, the bundles such as 'the results of the', 'on the other hand', 'the total number of', and 'it can be seen' appeared more frequently to indicate several functions in the discourse. However, 'the result of the' (most frequent bundle in MTEFL Corpus) was not mentioned in the literature of lexical bundles written by native academic writers. Some context-dependent-bundles were shared in the list, such as 'asked the students to', 'students were able to', and 'by the students in'.

Structural Types of Lexical Bundles

Based on the structural categories of lexical bundles, as can be seen in table 3, the most frequent structural subcategory of lexical bundles is noun phrase + *of*. Noun phrase + *of* structure such as 'the result of the', 'the total number of', 'the end of the', 'the mean score of', and 'the purpose of the' were widely spread across 29% of the total bundles. Other noun phrases such as 'most of the students', 'English as a foreign', and 'all of the students' were the second most frequently used structural category of lexical bundles (around 22%).

Performing as the most frequent lexical bundle in Findings and Discussion section, the noun phrase +*of* structure covered numbers of function in academic discourse. This finding is similar to Hyland (2008) previous research on lexical bundles across disciplinary variation in academic discourse. From his findings, noun phrase + *of* structural type was used in 24.4% across four academic discourses. This structure of bundle is commonly used to show the quantity of things, place, and size (Hyland, 2008) such as the following example.

> *The mean score of* pretest of the experimental group is 51.72, while the control group is 68.28. (2017_TR)

However, the findings of this research were different from the previous study by Wachidah et al. (2020) who identified other

prepositional phrase fragments (following Biber's taxonomy) such as 'from the result of', 'in the same time', and 'by looking at the' as the most frequent lexical bundles in ten Findings and Discussion section of graduate students theses. Previous research which used a manual approach to identify lexical bundles in ten graduate students' theses might account for the differences.

Prepositional phrase + *of* was distributed in total 10% out of overall structural categories of lexical bundles. The bundles beginning with prepositional phrases such as 'in the case of', 'in the form of', and 'in the context of' were used to show how prepositional items are linked in logical relations (Hyland, 2008). In addition, MTEFL Corpus shows the other structural subcategory of lexical bundles in different frequencies. Other prepositional phrase patterns accounted for about 20% of the lexical bundles used in the section, such as 'at the same time', 'on the other hand', and 'in line with the'. One specific prepositional phrase bundle, 'on the other hand' showed the greatest frequency in the corpus from this category.

Functional Distribution of Lexical Bundles

In terms of its functional categories, the first function, research-oriented bundles support the writer to describe and contextualize the structure of research (Hyland, 2008). There were five subclasses of research-oriented bundles that occurred in the MTEFL Corpus: location, procedure, quantification, description, and topic. Location subcategory was used to indicate the time or place, such as 'in front of the', 'in the middle of', and 'at the beginning of' as shown in the following example.

Firstly, the teacher showed a picture of two children sitting *in front of the* TV (2018_AL)

The second subcategory of research-oriented bundles is procedure, for example 'the use of the' and 'the students were asked (to)'. Students used procedure bundles to indicate the method or

purpose of the research (Jalali, 2018). In the corpus, 'the students were asked to' was usually followed by verbs such as 'answer', 'read', and 'guess'. This bundle showed the ways that the research was conducted.

> After that, *the students were asked* to guess the unknown words found in the text. (2016_FI)

Quantification is the third sub-class of research-oriented lexical bundles. It was used to describe the amount of the object being discussed in the text. The examples of this sub-category were found in these two examples.

> As to the pronunciation, *most of the students* usually exhibit clear and accurate pronunciation of words despite being anxious (2016_AG)
> *The percentage of the* vowels errors were 28.05%. (2017_IS)

The word 'most' in the first example indicates determiners used to indicate countable nouns. In addition, the second example shows a direct number of vowel errors in the form of percentage. Both bundles supported the writer to describe the number of things in the research.

While quantification tries to focus on the quantity of the object, the fourth sub-class of research-oriented bundles, description tells the quality or the properties of the object. Hyland (2008) identified several bundles in this sub-category, such as 'the size of the', 'the surface of the', and 'the structure of the'. In MTEFL Corpus, the bundles such as 'the meaning of the', 'the implementation of the', and 'the difference between the' were examples of the description sub-class of lexical bundles. Those bundles were contributed to the description of the research context.

The last sub-class of research-oriented lexical bundles is topic. It is directly related with the subject of discussion in research (Hyland, 2008). Since MTEFL Corpus was compiled from graduate

student theses in the field of Teaching English as a Foreign Language (TEFL), some topic bundles such as 'the teaching and learning (process)', 'English as a foreign (language)', and 'the use of English' were identified in a relatively high number of theses.

The second functional classification of lexical bundles is text-oriented bundles. Hyland (2008) mentioned that this function related with the "organisation of text and its meaning as a message or argument" (p. 13). In the other source, Biber et al. (1999) proposed the term, discourse organizer to identify lexical bundles that are related with the structure of certain discourse. There are four main subcategories (signals) for this function: transition, resultative, structuring, and framing signals. All of the four subcategories of the second function (text-oriented) bundles were discovered in MTEFL Corpus.

Transition signals are used to create additional information or show the contrastive correlation between different ideas in the discourse (Hyland, 2008). In MTEFL Corpus, one of the examples of transition signal was 'on the other hand', which is frequently used in the thesis to contrast different ideas. An example of this subcategory can be seen in the following sentence.

> The preposition phrase shows a high dominance in this study. *On the other hand*, adjective phrase and present participle clause both appear only once across the 36 compositions. (2017_AW)

From the example above, the writer used lexical bundle, 'on the other hand' to contrast between two different elements in the text: preposition phrase and adjective phrase. In addition, transition signals can also be used as additive links. In MTEFL Corpus, 'as well as the' performed as the additive link.

> The author decided to have designer to design the cover so it would match the content of the textbook *as well as the* color (2018_KW)

The second subcategory of text-oriented bundles is resultative signals, causative relations between elements (Hyland, 2008). The example of this subcategory, 'the result of the' is the most frequent bundle. Almost similar, another resultative signal can be identified in the bundle 'the results of the' (with –*s*) which occurred in 30 different texts in the corpus. Resultative signals were frequently used in Findings and Discussion because the section is one part of the thesis focusing on the results of the research. The example of resultative signals can be seen in these following two examples.

> *The result of the* post-test was in line with the class observation result (2018_WY)
> *The results of the* closed-ended and the open-ended questionnaire indicate that most of grade 9 subjects more frequently used Indonesian to communicate with their parents rather than English (2017_JA)

Another subcategory of text-oriented lexical bundles is structuring signals. This subcategory is used as the "text-reflective markers" that can direct readers elsewhere in the text. Some lexical bundles in MTEFL Corpus can be grouped into this subcategory, such as 'it can be concluded (that)', 'as shown in the', 'as stated in the', and 'from the table above'. The writer used the bundle, 'as stated in the' to direct the readers into another part of the text (for example, on a different page). The following example shows the example of structuring signal to direct the reader to go to the previous chapter.

> *As stated in the* previous chapter, Halliday and Hasan (1976:77-78) distinguished comparison into two parts: (2015_GN)

The last subcategory of text-oriented lexical bundles is framing signals, which are used to frame an argument "by specifying limiting conditions" (Hyland, 2008, p.14). Some bundles in the Findings and Discussion section of graduate students' theses could be classified in

the framing signals, for example, 'in the form of', 'in line with the', 'in accordance with the', and 'in terms of the'. The following example showed that the writer situated the argument by specifying the term.

> Teachers found the coursebook satisfactory *in terms of the* reading materials. (2019_SB)

Participant-oriented bundles, also called stance bundles (Biber et al., 1999), focus on the person (writer or reader) of the writing (Hyland, 2008). This function is divided into two main subcategories: stance features and engagement features. While stance features state the writer's personal feelings, attitude, and evaluation, engagement features try to address the reader directly. The lexical bundles in the MTEFL Corpus showing stance features were expressions such as 'students were able to', 'it is important to', and 'there was a significant'. Examples of this bundle can be found in the following sentence.

> *It is important to* take some notes on why parents enrolled their children to the school. (2017_RT)

The second subcategory of participant-oriented bundles is engagement features. As the names suggest, this subcategory refers to the way the writer is directly involved in addressing the readers in the text. The bundles such as 'it can be seen', 'it could be seen', and 'could be seen from' were included in the engagement features subcategory as shown in the following example.

> From the discussion above, *it can be seen* that there were only five out of six types of social power revealed in the speeches. (2019_HD)

Based on the findings, with the total percentage of 45%, research-oriented function of lexical bundles was the highest percentage in the corpus. It indicates that students were focused on

the structure or research activity. Indeed, Hyland (2008) argued that master's thesis is a pedagogic genre which forces students to handle research methods carefully because of its assessment purpose. However, the lower percentage of text-oriented bundles reflects that the students did not use adequate bundles to organize the text. In social science, text-oriented functions of lexical bundles are important to engage with the reference, connect the idea, direct the readers, and specify limitations.

CONCLUSION

It has been displayed that the most structural type of lexical bundle in the findings is noun phrase + *of* structure (e.g. 'the result of the' and 'the total number of'), which covers one third of the overall forms in the corpus. While native writers commonly use 'the results of the', EFL students in this study tended to use 'the result of the' (singular). In addition, other noun phrases were distributed in the total of 22% out of overall bundles. On the other side, the use of passive structure bundles and anticipatory-*it* pattern are rarely used in the Findings and Discussion section. Anticipatory *it* patterns (around 8% in the corpus) such as 'it can be seen' and passive structure bundles (2% in the corpus) such as 'can be found in' were rarely used to accommodate the writers' point of view and address the reader in the text.

This study revealed that graduate students in Teaching English as a Foreign Language (TEFL) used more research-oriented bundles (45% of overall bundles), which meant that the students were focused on the structure of their research. The lower percentage of text-oriented bundles (40%) in the corpus reflects that the students did not use adequate bundles to organize the text. On the other side, participant-oriented bundles were the least frequently used form. Students sometimes used stance features as attitude-markers that expressed their beliefs.

Finally, the findings of the present study could be compared to academic essay writing to reveal the possible similar and different

lexical bundles used in scientific writing (i.e., thesis writing) and academic writing. Such comparative investigations may offer a more comprehensive picture of formulaic expressions and lexical bundles used in university context for EFL students.

REFERENCES

Alfiandita, O. L., & Ardi, P. (2020). Collocations in the English LKS Books for Senior High School Students. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, *4*(2), 213. https://doi.org/10.21093/ijeltal.v4i2.463

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, *17*, 51–62. https://doi.org/10.1016/j.jeap.2015.02.001

Altun, L. (2019). A Corpus Based Study: Analysis of the Positive Reviews of Amazon.com Users. *Advances in Language and Literary Studies*, *10*(1), 123. https://doi.org/10.7575/aiac.alls.v.10n.1p.123

Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The Use of Lexical Bundles in Native and Non-native Post-graduate Writing: The Case of Applied Linguistics MA Theses. *Language-Journal.Tabrizu.Ac.Ir*, (11). Retrieved from http://language-journal.tabrizu.ac.ir/Files/Journal/2013-10-21_10.02.46_The Use of Lexical Bundles in Native and Non-native Post-graduate Writing The Case of Applied Linguistics MA Theses.pdf

Bal, B. (2010). Analysis of Four-word Lexical Bundles in Published Resesarch Articles Written by Turkish Scholars. *Analysis*, *11*, 30–2010.

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

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers*, *26*, 181–190.

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., Conrad, S., & Leech, G. (2002). Longman Grammar of Spoken and Written English. *Pearson Education*. https://doi.org/10.1177/0075424202250290

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*(5), 31–64. Retrieved from http://faculty.edfac.usyd.edu.au/projects/usp_in_tesol/pdf/volume05/Article02.pdf

Çandarlı, D. (2020). Changes in L2 writers' self-reported metalinguistic knowledge of lexical phrases over one academic year. *Language Learning Journal*, *48*(6), 768–784. https://doi.org/10.1080/09571736.2018.1520914

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology*, *14*(2), 30–49. Retrieved from http://llt.msu.edu/vol14num2/chenbaker.pdf

Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, *34*(3), 369–388. https://doi.org/10.1093/applin/ams045

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, *29*, 21–42. https://doi.org/10.1016/j.jeap.2017.09.002

Guiling, N. (2015). A corpus-based analysis of lexical bundles in english introductions of chinese and international students' theses. *Lecture Notes in Computer Science (Including Subseries*

Faqih, M. S. & Harjanto, I. (2022). English lexical bundles in the graduate theses: the frequency, structure and distribution.

    Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), *9332*(May 2015), 486–493. https://doi.org/10.1007/978-3-319-27194-1_49

Gungor, F., & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, *9*(6), 176. https://doi.org/10.5539/elt.v9n6p176

Heine, B., Narrog, H., & Biber, D. (2015). Corpus-Based and Corpus-Driven Analyses of Language Variation and Use. In *The Oxford Handbook of Linguistic Analysis*. https://doi.org/10.1093/oxfordhb/9780199677078.013.0008

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

Jalali, Z. S. (2018). A Corpus-Based Study of Lexical Bundles Discussion Section of Medical Research Articles. *Iranian Journal of Applied Language Studies*, (1), 95–124. https://doi.org/https://dx.doi.org/10.22111/ijals.2018.4265

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 132–140. https://doi.org/10.17509/ijal.v7i1.6866

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85–102. https://doi.org/10.1016/j.jslw.2009.02.001

Liu, S. (2014). Formulaic Language: An Analytical Review. *International Journal of Linguistics*, *2*(2), 1–11.

Ngadiman, A. (2013). the Lexical Phrases Employed in the Thesis Abstracts of English. *Magister Scientiae*, (33), 93–106. https://doi.org/https://dx.doi.org/10.33508/mgs.v0i33.613

Novita, H., & Kwary, D. A. (2018). Comparing the use of lexical bundles in Indonesian-English translation by student translators and professional translators. *Translation and*

Interpreting, *10*(1), 53–74.
https://doi.org/10.12807/ti.110201.2018.a04

Nuraniwati, T., & Permatasari, A. N. (2021). Hedging in TED TALKS: A Corpus-based Pragmatic Study. *JEELS (Journal of English Education and Linguistics Studies)*, *8*(2), 29–50.

Oktavianti, I. N., & Sarage, J. (2021). Collocates of 'Great' and 'Good' in the Corpus of Contemporary American English and Indonesian EFL Textbooks. *Studies in English Language and Education*, *8*(2), 457–478.
https://doi.org/10.24815/siele.v8i2.18594

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, *191*, 225.

Samodra, M. C., & Pratiwi, V. D. R. (2018). Lexical Bundles in Indonesian and English Undergraduate Thesis Abstracts. *Atlantis Press*, *166*, 183–191. https://doi.org/10.2991/prasasti-18.2018.35

Thongvitit, S., & Thumawongsa, N. (2017). *A Corpus-Based Study of English Collocations Found in the Abstracts of Research Articles Written by Thai EFL Writers*. *7*(12).
https://doi.org/10.18178/ijssh.2017.7.12.920

Wachidah, W. D. N. A., Fitriati, S. W., & Widhiyanto. (2020). Structures and Functions of Lexical Bundles in Findings and Discussion Sections of Graduate Students ' Thesis. *English Education Journal*, *10*(2), 131–142.
https://doi.org/https://doi.org/10.15294/eej.v10i1.33994

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, *20*(1), 1–28. https://doi.org/10.1016/s0271-5309(99)00015-4

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