

## Functional Categories of Lexical Bundles in Indonesian EFL Textbooks: A Corpus-Based Study

Esti Sapta Lestari <sup>1</sup>, Ikmi Nur Oktavianti <sup>2</sup>, Roslina Abdul Aziz <sup>3</sup>

<sup>1,2</sup> Universitas Ahmad Dahlan, Indonesia, <sup>3</sup> Universiti Teknologi MARA Malaysia  
[2307042010@webmail.uad.ac.id](mailto:2307042010@webmail.uad.ac.id), [ikmi.oktavianti@pbi.uad.ac.id](mailto:ikmi.oktavianti@pbi.uad.ac.id), [leenaziz@uitm.edu.my](mailto:leenaziz@uitm.edu.my)

### Abstract:

*Lexical bundles, which are recurring sequences of multiple words, play a crucial role in both written and spoken discourse. Understanding the presence and patterns of these bundles in educational materials is essential for effective language instruction and material development. This research aims to identify the most commonly utilized functional classifications of 3-word lexical bundles in junior high school EFL textbooks. This study employed a descriptive quantitative research methodology with a corpus-based analysis approach. The present study employed a corpus study combining quantitative and qualitative components to offer a thorough examination of the data. The data of this research comprises a collection of Indonesian junior high school EFL textbooks officially approved by the Ministry of Education. Specifically, the study examined textbooks used in grades 7-9, published in 2022. AntConc 4.3.1 was used to analyze the corpus, and N-Gram Tool was utilized to generate the 3-word lexical bundles in this study. The study shows that the most commonly utilized functional classifications of 3-word bundles in junior high school English textbooks is the referential expressions category; there are 914 bundles (44.76%). Furthermore, the implication of this research is that these bundles should be used by educators, especially those implementing the independent curriculum, to create instructional materials that enhance students' syntactic and pragmatic skills.*

**Keywords:** corpus linguistics, EFL textbooks, lexical bundle, three-word

---

### 1. INTRODUCTION

Textbooks are important in English language teaching, especially in the context of EFL, because they are the main source of teaching resources. Nevertheless, a number of studies indicate that many EFL textbooks fail to align with the real use of English, especially in terms of formulaic

language, despite their crucial role. Mustafa (2018) indicated that many conversational texts in EFL textbooks contain formulaic expressions that are inaccurately or inappropriately used, leading to unnatural dialogue. Foll (2024) found that EFL textbooks often present a distinct variety of English, termed "Textbook English," which lacks the naturalness found in authentic language use.

Lexical bundles are extended collocations appearing more typically and commonly than expected by chance, contributing to coherence and distinctiveness in a text; they help make a text more coherent and unique (Hyland & Jiang, 2018). Recurrent word sequences, or lexical bundles, are crucial discourse building blocks in both written and spoken formats (Biber & Barbieri, 2007). Lexical bundles are important language units in English as a Foreign Language (EFL) learning because they offer a foundation for new vocabulary acquisition and help in the development of fluency in both spoken and written communication. Albelihi (2022) found that exposure to large lexical parts of reading and learning improved speaking fluency for students at intermediate or lower-intermediate level of EFL. Biber et al. (2004) reported that lexical bundles include noun phrases, verb phrases, and prepositional phrases are crucial components of both academic and non-academic speech. These bundles give learners ready-made pieces of English that they may incorporate into their spoken and written communication, improving their language ability (Biber et al., 2004).

Lexical bundles in EFL textbooks have been the subject of numerous research. Lexical bundles are common in Japanese junior high school textbooks and, at shorter lengths (3-word bundles), they correspond to English patterns quite well, according to Northbrook & Conklin (2018). Three- and four-word lexical bundles are less common in Saudi high school textbooks, according to Alzahrani (2020). More lexical bundles that are representative of the structure and function of the bundles used by native speakers should be included by material designers, considering the relatively low frequency and restricted structure and function of lexical bundles in a major English textbook series used in public Saudi high schools.

In Indonesia, previous research has found that lexical bundles play an important role in promoting language learning. Ardi et al. (2023) conducted a corpus-based study on the patterns and structural classifications of lexical bundles found in EFL course textbooks for Indonesian senior high school students. The study found that the most common lexical bundles were three-word bundles with 32,527 instances followed by four-word bundles with 11,620 instances (Ardi et al., 2023). This suggests that small bundles may be used to assist pupils in retaining terminology and recalling the bundles when they are used. Meanwhile, Inaroh et al. (2020) investigated the structures and functions of lexical bundles of conversation texts in English language textbooks published by Kemendikbud Indonesia. This study discovered that personal pronoun + lexical VP was the most common bundle structure in the conversation texts, and there was a significant percentage of the bundle used in the stance statement.

Furthermore, the structural classifications of lexical bundles in Indonesian EFL textbooks have been examined. Ardi et al. (2023) divided the bundles into eleven structural categories, including "noun phrase + of phrase fragment" (173 occurrences) and "other prepositional phrases" (243 occurrences). These data indicate that the most prevalent structural categorization is the prepositional phrase, which is necessary for constructing complex sentences and improving fluency.

Given the importance of lexical bundles in improving language competency, there are some previous studies on lexical bundles in Indonesian context (Budiwiyanto & Suhardijanto, 2020; Fajri et al., 2020; Kwary et al., 2017; Oktavianti & Sarage, 2021). However, based on previous research, little is known about lexical bundles in *Kurikulum Merdeka* (Merdeka Curriculum) textbooks. Therefore, further study is needed to better understand how lexical bundles are used in the EFL textbooks for junior high school students in this context. This study attempts to fill this gap by completing a corpus-based analysis of the usage of lexical bundles in the EFL textbooks for junior high school students in Indonesia. This study aims to provide useful insights on how lexical bundles are used in EFL textbooks for junior high school students in Indonesia. The findings help curriculum developers refine textbook content by integrating more frequent and natural lexical bundles. This ensures that learners acquire formulaic sequences that improve fluency and communicative competence.

## **2. LITERATURE REVIEW**

### **2.1 Formulaic Languages**

According to Wray (2002), formulaic language refers to expressions that appear to be processed without recourse to their lowest level of composition. These include words, phrases, and longer passages that retain meaning beyond their constituent parts, often exhibiting predictable form and idiomatic usage. Lexical bundles, the focus of this study, are one type of formulaic language. Recent research by Bestgen (2020) shows that mastery of formulaic language has a positive correlation with overall language proficiency in EFL learners. This highlights how crucial it is to include formulaic language in EFL curriculum. Studies show that a large percentage of common language is formulaic. People frequently rely on prefabricated chunks of text due to their ease of processing and communicative effectiveness (Sidtis, 2021; Szudarski, 2017; Wray, 2002).

Formulaic language serves multiple purposes in communication, such as (1) efficiency by leveraging prefabricated structures, speakers reduce cognitive load and enhance fluency, (2) idiomacy (formulaic expressions contribute significantly to the richness and nuance of language, allowing for efficient conveyance of complex meanings), and (3) learning (native speakers utilize formulaic language extensively in their daily interactions, while learners initially rely heavily on these structures to initiate language development (Wray, 2002)). However, formulaic language presents difficulties for second-language learners, particularly those at intermediate levels of proficiency. While novice learners benefit from relying on formulaic expressions, advanced learners struggle to move beyond these prefabricated structures to achieve native-like fluency (Sidtis, 2021; Szudarski, 2017; Wood, 2015; Wray, 2002).

### **2.2 Lexical Bundles**

Combinations of three or more words that are found in a natural language corpus using corpus analysis software programs, identified using a certain frequency cutoff, and present in a specific set of texts within the corpus are known as lexical bundles (Cutler, 2021; Sidtis, 2021; Szudarski, 2017; Wood, 2015). According to Biber et al. (2004), lexical bundles—which are often occurring groups of three or more words—have gained a lot of attention in academic discourse analysis and corpus linguistics in recent decades. These multi-word sequences are essential for coherent communication and as indicators of register change in both written and spoken speech. The phrase "lexical bundles" was first introduced by Biber et al. (2004) to differentiate these commonly occurring word sequences from other forms of formulaic language, and the idea

originated from corpus-based research. Lexical bundles, in contrast to idioms, frequently seem structurally incomplete and do not always have an idiomatic meaning. Nonetheless, they are regarded as essential components of fluent verbal output and fulfill significant discourse roles (Hyland, 2008).

Biber's concept divides lexical bundles into structural and functional types and offers a thorough approach to examining their grammatical structures and communicative functions (Biber et al., 2004). Lexical bundles fall into four main structural categories: bundles based on noun phrases, like "the end of the"; based on prepositional phrases, like "in the middle of"; based on verb phrases, like "is going to"; and bundles based on dependent clauses, like "if you want to." In terms of function, Biber divides lexical bundles into three primary groups: stance expressions, which convey attitudes, judgments, or epistemic stance; referential expressions, which make references to entities or abstract concepts (e.g., "one of the most"); and discourse organizers, which organize discourse and direct the flow of ideas (e.g., "on the other hand").

Hyland's framework, on the other hand, provides a comprehensive functional classification that particularly focuses on academic discussion, which makes it particularly relevant for the analysis of educational material (Hyland, 2005; Hyland & Jiang, 2018; Oktavianti & Prayogi, 2022). Lexical bundles are categorized into three primary functional types by Hyland's theory, which is particularly important to scholarly discourse. Research-oriented bundles, which include elements of place, process, quantification, description, and topic (e.g., "at the beginning of," "the use of the," and "a wide range of"), assist authors in describing their research activities and experiences. Text-oriented bundles, which include transitional, resultative, structural, and framing signals (e.g., "on the other hand," "as a result of," and "in the next section"), are used to arrange the text and present arguments. Lastly, participant-oriented bundles, which are separated into stance bundles and engagement bundles (e.g., "are likely to be," "it should be noted"), center on the writer or reader. The way that various lexical bundles fulfill particular communicative functions in academic writing is well captured by this classification scheme.

### **3. RESEARCH METHODOLOGY**

#### **3.1 Corpus Tool**

In this study, AntConc corpus software version 4.3.1 was used to analyze the corpus (Anthony, 2024). AntConc 4.3.1 is a freeware corpus analysis toolkit developed by Laurence Anthony and released in 2024. AntConc was the primary tool for identifying lexical bundles in the Indonesian EFL textbooks. It was used to generate lists of potential bundles, which are then analyzed further using structural and functional framework by Biber. This study focuses on the description of the bundles of the textbooks without comparing to a native corpus to figure out the variety of the bundles in regard to the functional categories. It helps identify whether textbooks introduce lexical bundles systematically, reinforcing patterns for learners as they progress.

#### **3.2 Subject and Object of the Research**

The data of this research comprises a collection of Indonesian junior high school EFL textbooks officially approved by the Ministry of Education. Specifically, the study examined textbooks used in grades 7-9, published in 2022. The following table summarizes the corpus data used in the study.

**Table 1:** The corpus data

No	Title	Author(s)	Publisher(s)	Tokens
1	English for Nusantara untuk SMP/MTs Kelas VII	Ika Lestari Damayanti Yusnita Febrianti Iyen Nurlaelawati Pipit Prihartanti Suharto Aji Jehan Fellani Raymon Rahmadhan	Pusat Perbukuan Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi	23,219
2	English for Nusantara untuk SMP/MTs Kelas VIII	Ika Lestari Damayanti Iyen Nurlaelawati Mahmud Layan Hutasuhut Yusnita Febrianti Raymon Rahmadhani Gilang Asri Devianty	Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi	27,598
3	English for Nusantara untuk SMP/MTs Kelas IX	Ika Lestari Damayanti Yusnita Febrianti Pipit Prihartanti Suharto Iyen Nurlaelawati Aji Jehan Fellani Hendra	Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi	30,305
<b>Total tokens in corpus</b>				<b>81,122</b>

### 3.3 Data Collection and Lexical Bundle Identification

EFL textbooks used in Indonesian junior high schools under the *Merdeka* Curriculum were collected and digitized. The selection of textbooks was based on their current use and approval by the Ministry of Education and Culture (Kementerian Pendidikan dan Kebudayaan). The digitized textbooks were cleaned and formatted to ensure compatibility with the corpus analysis tool. This process involved removing non-textual elements, standardizing formatting, and converting files to *txt* format. Using AntConc, lexical bundles will be extracted from the prepared corpus.

The N-gram size was adjusted to the required length of lexical bundles, which is usually three to five words, in order to calculate the frequency of lexical bundles. A minimum frequency threshold was also defined. Although the precise threshold may differ; 10–20 occurrences per million words is a typical place to start (Biber et al., 2004).

To ensure that the discovered bundles were dispersed throughout the corpus and not merely frequent in one text, dispersion is essential. The range cut-off set was determined. According to Biber & Barbieri (2007), a lexical bundle occurring in at least 5% of the corpus's texts would be considered as adequately dispersed. This study, thus, complied with the dispersion criterion of 15 as posited by Biber & Barbieri (2007).

### 3.4 Data Verification

An Intercoder Reliability (ICR) was carried out to verify the data. ICR is defined as a numerical measure that quantifies the level of agreement between multiple coders regarding how data should be classified or coded (O'Connor & Joffe, 2020). The study uses Intercoder Reliability (ICR) as a data verification technique to guarantee the validity and dependability of the results.

### 3.5 Data Analysis

Data analysis was done for determining the frequency, function, and structure of lexical bundles in Indonesian junior high school EFL textbooks. Using AntConc, lexical bundles were extracted from a sample of three EFL textbooks. A frequency analysis was administered after the data was gathered in order to find the frequency of three-word bundles most common in the texts. This phase involved quantifying the frequency of each pattern. Next, the extracted lexical bundles were categorized into their structural categories, which consist of noun phrases and prepositional phrases, in accordance using the framework developed by Biber et al. (2004). This classification exposed the structural diversity of the bundles found in the textbooks. The purposes of these lexical bundles in the educational context were then ascertained by a functional categorization study. The results were evaluated and their pedagogical implications examined in order to shed light on how these lexical bundles can enhance instructional strategies and students' communicative skills in English language learning environments.

## 4. RESULTS

**Table 2:** Presents the functional categories of the bundles found in the analyzed EFL textbooks.

**Table 2:** The functional categories of lexical bundles (stance expression)

Grade VII			Grade VIII			Grade IX		
Stance expressions								
No	Type	Freq	No	Type	Freq	No	Type	Freq
1	do you have	28	1	what do you	19	1	been done for	40
2	say what you	22	2	answer the questions	17	2	done for you	40
3	been done for	22	3	do you think	16	3	has been done	40
4	done for you	22	4	what you know	15	4	answer the questions	34
5	has been done	22	5	say what you	15	5	what do you	17
6	what do you	20	6	what did the	14	6	what you know	16
7	what you know	18	7	been done for	13	7	did you know	15
8	do you like	11	8	done for you	13	8	do you know	15
9	i can describe	11	9	has been done	13	9	your teacher will	12
			10	one has been	12	10	you want to	11
			11	can do it	12	11	how to make	11
			12	i know i	11	12	do you think	11
			13	know i can	10	13	how do you	11
			14	i can do	10	14	what an experience	11
						15	going to a	10
<b>Total</b>		<b>176</b>	<b>Total</b>		<b>190</b>	<b>Total</b>		<b>294</b>

Besides stance expressions, some other lexical bundles classified in discourse organizer category were also found in the EFL textbooks. Table 3 below presents all the bundles within this category.

**Table 3:** The functional categories of lexical bundles (discourse organizer)

<b>Discourse organizer</b>								
1	look at the	25	1	based on the	36	1	look at the	39
2	see the wordbox	49	2	in the story	34	2	listen to audio	23
3	based on the	39	3	of the story	24	3	questions that follow	18
4	listen to audio	25	4	look at the	22	4	the questions that	18
5	read the text	18	5	from the story	15	5	then answer the	17
6	for specific information	12	6	listen to audio	15	6	based on the	16
7	information based on	11	7	the following table	15	7	and answer the	15
8	listen again to	10	8	language focus unit	13	8	language focus unit	13
9	work with a	10	9	on the story	13	9	go to the	12
10	on the text	10	10	the story part	13	10	make a pocket	10
			11	the following questions	12	11	let s donate	10
			12	questions based on	11	12	look at my	10
			13	write the sentences	11			
			14	with a classmate	11			
			15	the story here	11			
			16	complete the following	11			
			17	in the following	11			
			18	look around you	10			
			19	that tell you	10			
			20	sentences that tell	10			
			21	the sentences that	10			
<b>Total</b>		<b>209</b>	<b>Total</b>		<b>318</b>	<b>Total</b>		<b>201</b>

Besides discourse organizer, some other lexical bundles classified in referential expressions were also identified in the textbooks. Table 4 below display all the bundles within this category.

**Table 4:** The functional categories of lexical bundles (referential expressions)

<b>Referential expressions</b>								
1	number one has	21	1	the ugly duckling	107	1	number one has	36
2	one has been	21	2	independence day celebration	22	2	one has been	36
3	in front of	19	3	one of the	19	3	back to the	24
4	there is a	16	4	the yellow ducklings	19	4	merdeka friday shop	19
5	interrogative short answers	14	5	in the river	17	5	a pocket book	15
6	adverbs of frequency	13	6	the independence day	17	6	and princess suripit	15
7	food and drinks	12	7	a lot of	15	7	andre and princess	15

8	in the school	12	8	you know unit	15	8	lestari national park	15
9	of the room	12	9	happened to the	14	9	of timun mas	15
10	of the activity	11	10	the word box	14	10	say what you	15
11	on the picture	10	11	what happened to	14	11	to make a	15
12	once a week	10	12	a group of	13	12	a national park	14
13	rica rica chicken	10	13	in spoon race	13	13	one of the	14
14	where is the	10	14	marble in spoon	13	14	the charity shop	14
			15	section enrichment unit	13	15	the security officer	14
			16	see the word	13	16	to the hotel	14
			17	the sea animals	13	17	story of timun	13
			18	to the sea	13	18	fill in the	12
			19	ugly duckling story	13	19	saw green sea	12
			20	what is the	13	20	the help of	12
			21	did it rain	12	21	the story of	12
			22	in the oceans	12	22	to a national	12
			23	the mother duck	12	23	what is the	12
			24	what was the	12	24	with the help	12
			25	yes no questions	12	25	green sea turtles	11
			26	in the ocean	11	26	my underwater adventure	11
			27	number one has	11	27	of the story	11
			28	part of the	11	28	on the internet	11
			29	rain last night	11	29	the national park	11
			30	read the story	11	30	turn back hoax	11
			31	tell you about	11	31	a little further	10
			32	you about the	11	32	at my creation	10
			33	champion of panjat	10	33	difficulty with the	10
			34	characters in the	10	34	help of the	10
			35	going to a	10	35	i saw green	10
			36	help of the	10	36	i was swimming	10
			37	it rain last	10	37	in front of	10
			38	kindness towards differences	10	38	in the box	10
			39	of panjat pinang	10	39	orangutan and gorilla	10
			40	the champion of	10	40	re on sale	10
			41	the elements of	10	41	staying safe in	10
			42	the help of	10	42	study the following	10
			43	to a parade	10	43	they re on	10
						44	to the s	10
						45	while i was	10
						46	worksheet did you	10
<b>Total</b>		<b>191</b>	<b>Total</b>		<b>637</b>	<b>Total</b>		<b>613</b>

As presented in table 4.13 the distribution of the functional categories of lexical bundles in Grade VII is rather balanced. There are 176 occurrences of stance expressions, with the most common bundles being "do you have" (28 occurrences) and "what do you" (20 occurrences). There are 209 occurrences of discourse organizers, with the most common ones being "see the wordbox" (49 occurrences) and "based on the" (39 occurrences). With 191 occurrences, referential expressions are less common. The most frequent bundles are "one has been" (21 occurrences) and "number one has" (21 occurrences).

Lexical bundle frequency increases significantly in Grade VIII. There are 190 occurrences of stance expressions, with the most common ones being "what do you" (19 occurrences) and "answer the questions" (17 occurrences). With 318 occurrences, discourse organizers are the most frequent in this grade. Two notable examples are "based on the" (36 occurrences) and "in the story" (34 occurrences). With 637 occurrences, referential expressions are most common, exhibiting bundles such as "the ugly duckling" (107 occurrences) and "independence day celebration" (22 occurrences).

The use of lexical bundles is most varied and complex in grade IX. With 294 instances, stance expressions are the most common in this grade. Common bundles include "done for you" (40 occurrences) and "been done for" (40 occurrences). With 201 occurrences, discourse organizers are less common than in Grade VIII; yet, bundles like "look at the" (39 occurrences) and "listen to audio" (23 occurrences) are still important for task organization. With 613 occurrences, referential statements are likewise very common. Typical bundles that include specific content references are "number one has" (36 occurrences) and "one has been" (36 occurrences).

As for the functional categories in the EFL textbooks, the following table 5 shows the data for frequency of stance expression bundles.

**Table 5:** Frequency of stance expression bundles

<b>Stance expressions</b>	<b>Freq</b>	<b>Percentage</b>
<b>A. Epistemic Stance</b>		
1. Personal		
a. i shared with	13	0,637%
b. shared with my	13	0,637%
<b>Total</b>	<b>26</b>	<b>1,274%</b>
2. Impersonal		
a. what do you	56	2,742%
b. say what you	52	2,547%
c. what you know	49	2,400%
d. do you think	34	1,665%
e. did you know	33	1,616%
f. do you have	32	1,567%
g. what is the	29	1,420%
h. do you know	23	1,126%
i. how do you	18	0,881%
<b>Total</b>	<b>326</b>	<b>15,964%</b>
<b>B. Attitude/Modality Stance</b>		
1. Desire		
a. you want to	28	1,371%
b. do you like	16	0,784%
<b>Total</b>	<b>44</b>	<b>2,155%</b>
2. Obligation/ Directive Bundles		
a. you need to	20	0,979%
b. you should be	16	0,784%
c. to help you	16	0,784%
<b>Total</b>	<b>52</b>	<b>2,547%</b>
3. Intention/ Prediction Bundles		
are going to	17	0,833%

4. Ability Bundles		
a. we can use	16	0,784%
b. be able to	15	0,735%
c. should be able	15	0,735%
<b>Total</b>	<b>46</b>	<b>2,254%</b>
<b>Total for stance expressions</b>	<b>511</b>	<b>25,027%</b>

The first category is stance expression, with a total of 511 occurrences, stance expressions contribute to 25.027% of the lexical bundles and represent the attitudes, beliefs, and intentions of the speaker or writer. There are two primary types of this category: epistemic stance and attitude/modality stance. Attitude/modality stance, which is further subdivided into four subcategories: ability bundles (46 occurrences, 2.254%), obligation/directive bundles (52 occurrences, 2,547%), intention/prediction bundles (17 occurrences, 0,833%), and desire (44 occurrences, 2,155%). Epistemic stance includes both personal and impersonal expressions (26 occurrences, 1.274%, and 326 occurrences, 15.964%, respectively).

Along with stance expression, there is also discourse organizer bundle being identified more specifically into the subcategories as in the following (Table 6).

**Table 6:** Frequency of discourse organizer bundles

Discourse organizer	Freq	Percentage
<b>A. Topic Introduction</b>		
1. look at the	86	4,212%
2. listen to audio	63	3,085%
3. read the text	27	1,322%
4. the following questions	25	1,224%
5. go to the	21	1,028%
6. see the word	21	1,028%
7. at the following	19	0,930%
8. this is the	18	0,881%
9. in the following	17	0,833%
10. the goal of	14	0,686%
<b>Total</b>	<b>311</b>	<b>15,229%</b>
<b>B. Topic Elaboration</b>		
1. answer the questions	60	2,938%
2. to make a	29	1,420%
3. and answer the	26	1,273%
4. are talking about	20	0,979%
5. listen again to	20	0,979%
6. again to audio	19	0,930%
7. complete the following	19	0,930%
8. fill in the	19	0,930%
9. questions based on	18	0,881%
10. answer the following	17	0,833%
11. to ask for	15	0,735%
12. upon completion of	15	0,735%
13. how to make	15	0,735%
14. ask and give	14	0,686%
<b>Total</b>	<b>306</b>	<b>14,984%</b>
<b>Total for Discourse organizer</b>	<b>617</b>	<b>30,213%</b>

Discourse organizers, which control the discourse's flow and structure, account for 617 occurrences, or 30,213% of the lexical bundles. The two subcategories that they fall under are topic elaboration (306 occurrences, 14.984%) and topic introduction (311 occurrences, 15.229%).

Besides discourse organizers, the detailed frequency of referential expression is displayed below.

**Table 7:** Referential expression bundles

<b>Referential expressions</b>	<b>Freq</b>	<b>Percentage</b>
<b>A. Identification</b>		
1. and his friends	22	1,077%
2. galang and his	22	1,077%
3. there is a	21	1,028%
4. it is a	14	0,686%
5. with my family	14	0,686%
6. in a group	13	0,637%
<b>Total</b>	<b>106</b>	<b>5,191%</b>
<b>B. Imprecision Bundle</b>		
a lot of	25	1,224%
<b>C. Specification of Attributes Bundle</b>		
1. Quantity Specification Bundles		
a. one of the	36	1,763%
b. a group of	24	1,175%
c. group of four	13	0,637%
d. a piece of	16	0,784%
<b>Total</b>	<b>89</b>	<b>4,359%</b>
2. Tangible Framing Attribute		
a. one has been	69	3,379%
b. number one has	68	3,330%
c. the word box	20	0,979%
d. a picture of	16	0,784%
e. in the box	16	0,784%
f. on the picture	14	0,686%
g. part of the	14	0,686%
<b>Total</b>	<b>217</b>	<b>10,628%</b>
3. Intangible Framing Attributes		
a. based on the	91	4,456%
b. been done for	75	3,673%
c. has been done	75	3,673%
d. done for you	75	3,673%
e. the name of	18	0,881%
<b>Total</b>	<b>334</b>	<b>16,356%</b>
<b>D. Time/Place/Text Reference Bundles</b>		
1. Place and Time Reference Bundles		
a. in front of	33	1,616%
b. at smp merdeka	17	0,833%
c. front of the	16	0,784%
<b>Total</b>	<b>66</b>	<b>3,233%</b>
2. Time Reference		
a. in the morning	14	0,686%
b. at the end	14	0,686%
<b>Total</b>	<b>28</b>	<b>1,372%</b>

3. Text Deixis Bundles		
a. the words in	18	0,881%
b. words in the	16	0,784%
c. at the pictures	15	0,735%
<b>Total</b>	<b>49</b>	<b>2,400%</b>
<b>Total for Referential expressions</b>	<b>914</b>	<b>44,760%</b>

With 914 occurrences, referential expressions make up 44,76% of the lexical bundles and are employed to allude to things, numbers, or attributes. This category includes the following subcategories: identification (106 occurrences, 5.191%); imprecision bundle (25 occurrences, 1.224%); specification of attributes (including quantity specification bundles with 89 occurrences, 4.359%; tangible framing attributes with 217 occurrences, 10.628%; and intangible framing attributes with 334 occurrences, 16.356%); and time/place/text reference bundles (which are further subdivided into place and time reference bundles (66 occurrences, 3.233%), time reference (28 occurrences, 1.372%); and text deixis bundles (49 occurrences, 2.4%). The total number of lexical bundles in the textbooks is 2,042.

According to the previously indicated information, the referential expressions category has 914 bundles (44,76%), the discourse organizer category has 617 bundles (30,213%), and the stance expressions category has 511 bundles (25,027%). Referential expression is hence the most common use of lexical bundles in textbooks. In contrast to the other functional categories of lexical bundles, it demonstrates that textbook authors frequently employ referential expressions, demonstrating the primary purpose of lexical bundles as reference markers. Discourse organizer, which has to do with how ideas are arranged in a discourse, is the second most common bundle function.

The following part provide examples of each functional classification.

**a. Stance Expressions**

1) Epistemic Stance

(a) Personal

- (1) What I didn't like or found difficult. (Grade VIII page 122)
- (2) I didn't go to the national park a week ago. (Grade IX page 86)

(b) Impersonal

- (1) What do you usually post there? (Grade VIII page 149)
- (2) What do you think about the sky and the water? (Grade IX page 120)

2) Attitude/Modality Stance

(a) Desire

- (1) Mention the tips from the infographic that you want to do. (Grade VII page 177)
- (2) What do you want to know more about throwing trash? (Grade VIII page 151)
- (3) "Do you want to meet the professor?" (Grade IX page 174)

(b) Obligation/ Directive Bundles

- (1) Underline the suitable action words/verbs that you need to write your procedural steps. (Grade VII page 94)
- (2) Underline the suitable action words/verbs that you need to write your procedural steps. (Grade VIII page 140)
- (3) You need to buy some items. (Grade IX page 243)

- (c) Intention/ Prediction Bundle
  - (1) You are going to make a pencil case using a used plastic bottle. (Grade VII page 140)
  - (2) Think of what you are going to mime. (Grade VIII page 246)
- (d) Ability Bundles
  - (1) To make our posters engaging, we can use attractive colors and illustrations. (Grade VIII page 170)
  - (2) To give more information about an animal, we can use a noun group. (Grade IX page 38)
- b. Discourse Organizer
  - 1) Topic Introduction
    - (1) Look at the following pictures. Label the pictures with the correct names of the activities. (Grade VII page 220)
    - (2) Look at the examples in the following table. (Grade VIII page 196)
  - 2) Topic Elaboration
    - (1) Answer the questions based on the text. (Grade VII page 46)
    - (2) Based on the dialogue, answer the questions. (Grade IX page 83)
- c. Expressions
  - 1) Identification
    - (1) There is a television in the living room. (Grade VII page 108)
    - (2) When one question has been guessed then there is a change of role. (Grade VIII page 64)
  - 2) Imprecision Bundle
    - (1) Its library has a lot of books. (Grade VII page 216)
    - (2) Why is there a lot of plastic trash in the river? (Grade VIII page 189)
    - (3) There were a lot of orchids. (Grade IX page 82)
  - 3) Specification of Attributes Bundle
    - (a) Quantity Specification Bundles
      - (1) Now, Shishi high school is one of the top ranked schools in China. (Grade VII page 216)
      - (2) Circle one of the pictures that shows what happened to the turtle in the YouTube video. (Grade VIII page 208)
    - (b) Tangible Framing Attribute
      - (1) Which statements are true according to the story? Number one has been done for you. (Grade VIII page 210)
      - (2) Underline the first actions in these sentences. Number one has been done for you. (Grade IX page 128)
    - (c) Intangible Framing Attributes
      - (1) Answer the questions based on the story. (Grade VIII page 211)
      - (2) Tick the correct picture based on the summary of the story you heard. (Grade IX page 156)

(d) Time/ Place/ Text Reference Bundle

Place and Time Reference Bundles

- (1) There is a whiteboard in front of the class. (Grade VII page 234)
- (2) The teacher asked Radit to sit in the next row in front of Mumtaz. (Grade VIII page 268)
- (3) Monita met Prof. Mahmud in front of the warehouse. (Grade IX page 174)

Time Reference

- (1) She came to me in the morning, (Grade VIII page 276)
- (2) In the morning, the couple quickly prepared the things in a bag. (Grade IX page 190)

(e) Text Deixis Bundles

- (1) Read the words in the bubbles. (Grade VII page 28)
- (2) Every group writes the nine words in the grid. (Grade VIII page 32)

## 5. DISCUSSION

Referential expressions predominate in the data across all levels, although they are most prevalent in Grade VIII, where they occur 637 times. This suggests that comprehension of the subject matter and referencing specific topics are highly valued in the textbooks. This outcome is consistent with the research of Lee (2020), who examined linguistic textbooks and discovered that a significant percentage of lexical bundles (84.9%) had referential functions according to the functional taxonomy of lexical bundles. This outcome differs with the research conducted by Inaroh et al. (2020), who looked into the Ministry of Education and Culture's Bahasa Inggris textbook. They discovered the stance expression contained a significant amount of the bundle's most frequently used function in the textbooks.

Discourse organizers are frequently used for guiding students through exercises and preserving literary coherence, particularly in Grades VII and VIII. Although they are less common overall, stance expressions show an increasing trend from Grade VII to Grade IX, with Grade IX having the highest frequency. This indicates that as the curriculum advances, there is a greater emphasis on instructional language and student interaction. The growing linguistic and cognitive demands on students are reflected in the evolution of lexical bundle usage throughout grades.

Lexical bundles exhibit varying increases in complexity and frequency in EFL textbooks for Grades VII, VIII, and IX. The most common category is referential statements, emphasizing the value of narrative comprehension and content-specific references in the curriculum. This finding is in line with the study of Lee (2020) that revealed referential expressions represent the largest percentage of lexical bundles in linguistics textbooks, at 84.9%.

Discourse organizers are crucial for the lower grades since they help students with exercises and organize tasks. Although they are less common, position expressions are essential for promoting participation and enhancing the clarity of instruction; their usage rises in Grade IX. Overall, by deftly employing lexical bundles that correspond with the learning goals of each grade, the textbooks support students' language and cognitive development in a systematic and gradual manner.

The research findings partially address the identified problems but do not fully resolve them. The analysis confirms that referential expressions dominate across all grades, particularly in

Grade VIII, where they appear 637 times. While this suggests that textbooks emphasize content comprehension and referencing, it also reinforces the problem of structural imbalance, as clausal bundles remain underrepresented. Although discourse organizers and stance expressions increase progressively across grades, their usage is still limited compared to referential expressions, meaning students are not exposed to a wide variety of linguistic structures.

Additionally, discourse organizers and stance expressions are used less frequently, with stance expressions increasing from Grade VII to IX but still remaining secondary. This aligns with the problem of insufficient academic-style lexical bundles, as the absence of clausal bundles—essential for academic and formal writing—remains unaddressed. Without these, students may struggle with complex sentence formation and academic discourse, highlighting the need for further improvements in textbook content. Despite these gaps, the study successfully addresses the research gap by analyzing lexical bundles in junior high school EFL textbooks under *Merdeka* Curriculum, where previous research mainly focused on senior high school textbooks. The findings provide valuable insights into the distribution and function of lexical bundles within these textbooks.

The findings of the study can inform textbook writers and curriculum developers to design textbooks that offer more varieties of lexical bundles. These bundles play a crucial role in academic writing, subject-specific discourse, and language acquisition. By integrating a greater diversity of lexical bundles, curriculum developers and textbook writers can enhance the clarity, coherence, and effectiveness of learning materials. The results of corpus research can inform the materials or curriculum developers to design materials or syllabus to address the learners' competence needs and to prepare them to be communicatively competent (McCarten, 2022; McCarthy & McCarten, 2022; Rodríguez-Fuentes & Swatek, 2022).

Although the study adds new information, it makes no assessment of the pedagogical efficacy of these lexical bundles for EFL learning in contrast to native English corpora. To determine whether the discovered bundles actually improve students' academic and communication abilities, more research is required. In summary, the findings corroborate the identification of the problem but fall short of providing a complete solution by validating the structural and functional imbalances in lexical bundle utilization. In order to address the lack of clausal bundles, enhance structural diversity, and guarantee compatibility with best practices in EFL learning, more research and textbook updates are required, even though the study effectively detects patterns and trends.

## **6. CONCLUSION**

The findings of this study highlight that referential expressions are the most frequently occurring functional classification of 3-word lexical bundles in Indonesian junior high school EFL textbooks, accounting for 44.76% of the total bundles identified. This prevalence suggests that these textbooks prioritize referential expressions, which could significantly impact students' acquisition of discourse coherence and cohesion. The study's use of a corpus-based methodology, employing AntConc 4.3.1 and N-Gram Tool for analysis, allowed for a systematic identification of lexical bundles, reinforcing the importance of such tools in language instruction research. The results indicate that understanding the distribution of lexical bundles can help educators develop more effective instructional materials that align with natural language use and improve students' linguistic competence. Exposure to a variety of lexical bundles allows students

to predict and process meaning more efficiently. Therefore, educators and curriculum developers are encouraged to incorporate findings from corpus-based studies into material design.

This study has several drawbacks in spite of its contributions. Only textbooks released in 2022 were included in the analysis, and differences between textbook publishers and supplemental learning resources were not taken into consideration. Furthermore, although the study concentrated on functional categories, it did not investigate how well these bundles taught in real classroom environments. Future studies should look at how lexical bundles affect students' language development in real-world situations and take into account a larger dataset that includes a variety of educational materials.

## 7. REFERENCES

- Albelihi, H. H. M. (2022). Intermediate English as a Foreign Language learners' formulaic language speaking proficiency: Where does the teaching of lexical chunks figure? *Frontiers in Psychology, 13*, 949675. <https://doi.org/10.3389/fpsyg.2022.949675>
- Alzahrani, A. (2020). The Structure and Function of Lexical Bundles in Communicative Saudi High School EFL Textbooks. *International Journal of Applied Linguistics and English Literature, 9*(5), Article 5. <https://doi.org/10.7575/aiac.ijalel.v.9n.5p.1>
- Anthony, L. (2024). *AntConc* (Version 4.3.1) [English]. Waseda University.
- Ardi, P., Oktafiani, Y. D., Widianingtyas, N., Dekhnich, O. V., & Widiati, U. (2023). Lexical Bundles in Indonesian EFL Textbooks: A Corpus Analysis. *Journal of Language and Education, 9*(2), Article 2. <https://doi.org/10.17323/jle.2023.16305>
- Bestgen, Y. (2020). Comparing Lexical Bundles across Corpora of Different Sizes: The Zipfian Problem. *Journal of Quantitative Linguistics, 27*(3), 272-290. <https://doi.org/10.1080/09296174.2019.1566975>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), Article 3. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), Article 3. <https://doi.org/10.1093/applin/25.3.371>
- Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics, 10*(2), Article 2. <https://doi.org/10.17509/ijal.v10i2.28592>
- 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.
- 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), Article 4. <https://doi.org/10.17323/jle.2020.10719>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), Article 2. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), Article 1. <https://doi.org/10.1016/j.esp.2007.06.001>

- Hyland, K., & Jiang, F. (Kevin). (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407. <https://doi.org/10.1075/ijcl.17080.hyl>
- Inaroh, I., Faridi, A., & Wuli Fitriati, S. (2020). The Use of Structures and Functions of Lexical Bundles in Conversation Texts in Bahasa Inggris Textbook Published By Kemendikbud. *English Education Journal*, 11(1), 105-113. <https://doi.org/10.15294/ej.v11i1.43411>
- Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, 7(1), Article 1. <https://doi.org/10.17509/ijal.v7i1.6866>
- Le Foll, E. (2024). *Textbook English: A multi-dimensional approach* (Vol. 116). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.116>
- Lee, H.-K. (2020). Lexical bundles in linguistics textbooks. *언어연구*, 37(1), Article 1. <https://doi.org/10.17250/KHISLI.37.1.202003.005>
- McCarten, J. (2022). Corpus-informed course design. In A. O’Keeffe & J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371-386). Routledge.
- McCarthy, M., & McCarten, J. (2022). Writing corpus-informed materials. In J. Norton & H. Buchanan (Eds.), *The Routledge handbook of materials development for language teaching* (pp. 170-184). Routledge.
- Mustafa, F. (2018). How much do high schools contribute to improving students’ English proficiency? Seeking alumni’s perception in Indonesia. *ASIA TEFL*, 20(2), Article 2.
- Northbrook, J., & Conklin, K. (2018). “What are you talking about?”: An analysis of lexical bundles in Japanese junior high school textbooks. *International Journal of Corpus Linguistics*, 23, 311-334. <https://doi.org/10.1075/ijcl.16024.nor>
- O’Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, 19, 160940691989922. <https://doi.org/10.1177/1609406919899220>
- Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners’ argumentative essays: A corpus study. *Studies in English Language and Education*, 9(2), Article 2. <https://doi.org/10.24815/siele.v9i2.23995>
- Oktavianti, I. N., & Sarage, J. (2021). Lexical Bundles in Students’ Argumentative Essays: A Study of Learner Corpus. *Indonesian Journal of EFL and Linguistics*, 6(2), Article 2.
- Rodríguez-Fuentes, R. A., & Swatek, A. M. (2022). Exploring the effect of corpus-informed and conventional homework materials on fostering EFL students’ grammatical construction learning. *System*, 104, 102676. <https://doi.org/10.1016/j.system.2021.102676>
- Sidtis, D. (2021). *Foundations of familiar language: Formulaic expressions, lexical bundles, and collocations at work and play*. John Wiley & Sons Inc.
- Szudarski, P. (2017). *Corpus linguistics for vocabulary: A guide for research* (1st ed.). Routledge. <https://doi.org/10.4324/9781315107769>
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.