

# **Structural classification of lexical bundles in Indonesian EFL textbooks for senior high school: A corpus-based study**

**Selza Azzahra Garini, Ikmi Nur Oktavianti\***

Universitas Ahmad Dahlan, Indonesia

Manuscript received September 4, 2025, accepted November 20, 2025, and published online November 30, 2025.

## **Recommended APA Citation**

Garini, S. A., & Oktavianti, I. N. (2025). Structural classification of lexical bundles in Indonesian EFL textbooks for senior high school: A corpus-based study. *Englisia: Journal of Language, Education, and Humanities*, 13(1), 163-184. <https://doi.org/10.22373/ej.v13i1.32131>

## **ABSTRACT**

Lexical bundles are fundamental to fostering language fluency and facilitating proficient academic writing, particularly within English as a Foreign Language (EFL) context. However, their prevalence and pedagogical utility in Indonesian EFL textbooks at the secondary level remain underexplored. This study examines the deployment of three-word lexical bundles in Indonesian senior high school EFL textbooks (grades 10–12) aligned with the *Kurikulum Merdeka*. Adopting a corpus-based methodology, the research employs AntConc software to analyze the frequency, structural composition, and functional distribution of identified bundles. The corpus consists of texts drawn from officially sanctioned textbooks corresponding to the new curriculum. Analysis revealed 53 distinct lexical bundles across the three grade levels. Structurally, noun phrase (NP)-based bundles predominated, followed by verb phrase (VP)-based and prepositional phrase (PP)-based bundles. These findings indicate a deliberate, graded progression in the linguistic complexity of the textbooks, which appears designed to scaffold the development of students' academic language proficiency. This research contributes to a deeper understanding of lexical bundle integration in EFL instructional materials, underscores their pedagogical significance, and provides insights for future curriculum development and language teaching methodologies.

**Keywords:** *Corpus study; EFL; Lexical bundles; Textbooks; Structural classification*

---

\* *Corresponding author:*

Ikmi Nur Oktavianti  
Universitas Ahmad Dahlan, Pramuka Street Number 42, Umbulharjo, Yogyakarta, Special Region of Yogyakarta 55166, Indonesia  
Email: [ikmi.oktavianti@pbi.uad.ac.id](mailto:ikmi.oktavianti@pbi.uad.ac.id)

## 1. Introduction

English is acquired in an academic or formal setting in Indonesia, enabling students to engage with the language as EFL learners. In this country, English is taught from primary to secondary education. The curriculum presently adopted is *Kurikulum Merdeka* that embodies the principle of student autonomy in learning, allowing individuals to pursue their interests and fostering creativity and knowledge acquisition (Aini & Adiyono, 2023; Nisa, 2024; Wulandari et al., 2024). This curriculum focuses more on students' characteristics which should be supported with the appropriate materials (Wahyudin, 2024). The textbooks under *Kurikulum Merdeka* are adaptive learning resources intended to facilitate a more flexible and relevant educational transformation. It offers instructors the opportunity to customize educational materials according to the requirements and attributes of students, thus promoting the cultivation of comprehensive competences. Thus, EFL textbooks in Indonesia are crucial to the school system and critical for fostering communicative skills.

During the learning of English, students frequently require assistance in comprehending the significance and use of vocabulary within a broader context. Nevertheless, the majority of research on EFL textbooks indicates that their language does not accurately represent authentic language use, hence inadequately preparing students with the essential communicative abilities for real-life tasks (Adawiyah & Oktavianti, 2023; Akhofullah & Oktavianti, 2023; Binti et al., 2024; Intamanjana & Sada, 2021; L. X. Li, 2022; Oktavianti et al., 2025; Oktavianti & Fajria, 2021). The students may know the meanings of individual words but often not understand how people conventionally use them within phrases or sentences. For instance, Alzahrani (2020) asserts that numerous students encounter difficulties in formulating sentences that are natural and coherent in the target language. English textbooks, being the primary resource in the educational process, frequently use lexical bundles to assist students in developing more sophisticated and organized language competencies. The incorporation of lexical bundles in EFL textbooks is crucial for enhancing students' vocabulary, as these textbooks prioritize the target language (Ardi et al., 2023; Binti et al., 2024; Lestari et al., 2025).

Lexical bundles are frequently encountered in textbooks and serve to arrange content and furnish structure to writing or discourse. The concept of "lexical bundle" was first presented by Biber et al. (1999). Lexical bundles are recurrent sequences of words in texts or speech that operate as cohesive linguistic units exceeding the scope of individual words. Biber et al. (1999) assert that lexical bundles are clusters of three or more words that commonly occur inside a particular syntactic structure (e.g., *in terms of*, *at the same time*, *a number of*). Lexical bundles are crucial for language instruction and acquisition, serving both as an indicator of students' English proficiency and as a means to enhance language learning (Lasmita et al., 2023). This idea pertains to phrases or word combinations that frequently occur in particular settings, offering structure and consistency in language. A frequency-based methodology finds the bundles. This indicates that the text possesses a frequency and range threshold (Budiwiyanto &

Suhardijanto, 2020). This threshold is determined by the frequency of these word combinations within a specific corpus. Frequency is the essential attribute of lexical bundles, as a multi-word sequence must be above a specific frequency threshold to be categorized as a bundle (Hassanzadeh & Tamleh, 2023). The frequency criterion may vary from 10 occurrences to 40 occurrences per million words, contingent upon the corpus size (Bychkovska & Lee, 2017; Lee, 2020; Pan et al., 2016). Consequently, examining lexical bundles and their prevalence facilitates the connection between textbook language and practical application, thereby assisting students in cultivating more authentic and proficient communication abilities.

Numerous research has been conducted on lexical bundles. A study by Inaroh et al. (2020) examined the organization of lexical bundles in conversational text from an EFL textbook published by the Ministry of Education through discourse analysis. This analysis identified personal pronouns and lexical verb phrases as the predominant bundles in conversational texts. Moreover, researchers determined that displays of stance constituted the predominant functional application of these bundles. Lasmita et al. (2023) investigated the classifications of lexical bundles included in English textbooks for senior high school students in Indonesia. Their research identified 146 lexical bundles in grade 10, 203 in grade 11, and 199 in grade 12. Furthermore, they recognized 299 categories of noun phrase lexical bundles, 159 categories of verb phrase lexical bundles, and 91 categories of prepositional phrase lexical bundles. Subsequently, a study conducted by Lee (2020) identifies a corpus of linguistics textbooks of 1.14 million words. The research findings indicate that, according to the structural analysis, noun phrase-based and prepositional phrase-based bundles comprised over 80% of all bundles. Meanwhile, the functional taxonomy of lexical bundles demonstrated that a significant number of these bundles served referential roles (84.9%). Conversely, far fewer were discourse organizers (8.6%) and stance expressions (6.5%).

Despite the examination of lexical bundles in Indonesian EFL textbooks (Ardi et al., 2023; Inaroh et al., 2020; Lasmita et al., 2023), the context of *Kurikulum Merdeka* remains inadequately investigated in these prior works. This study addresses this gap by providing insights into the distinctive features of *Kurikulum Merdeka*. Prior research frequently employed restricted samples, whereas our corpus-based approach examines a wider array of texts and educational resources, hence augmenting the validity of our results. This study concentrates on three-word lexical bundles due to their prevalent presence in Indonesian senior high school English textbooks. It aims to identify the usage of these bundles through a corpus-based analysis and to examine their structural classification in EFL textbooks aligned with the *Kurikulum Merdeka* for grades 10, 11, and 12. More specifically, this study addresses the following questions:

- (1) What are the most frequent 3-gram bundles in the textbooks? and
- (2) How are they structurally classified?

## 2. Literature review

### 2.1. Formulaic language

Formulaic language comprises a series of words that can be either contiguous, exemplified by "on the other hand," or non-contiguous, as in "the \* of the" (Cutler, 2021; Sidtis, 2021; Wood, 2015; Wray, 2002). Formulaic language includes idioms, phrasal verbs, expressions, collocations, lexical phrases, and lexical bundles, all of which are essential for producing natural and fluent speech (Sidtis, 2021; Szudarski, 2017; Wood, 2015). Comprehending and employing formulaic language is crucial for EFL learners as it enhances both accuracy and fluency in communication (Meng et al., 2022). Formulaic language is essential for second language (L2) learners as it reduces learning effort, improves communicative competency, and provides pre-constructed linguistic parts for application (Bao & Liu, 2024). By mastering formulaic language, students can enhance their fluency in speaking and writing, as well as improve their comprehension of everyday dialogues and intricate written materials.

Formulaic language, especially lexical bundles, is commonly present in EFL textbooks and is helpful in aiding students to comprehend and utilize language more readily. In EFL textbooks, formulaic language is frequently instructed within pertinent settings and authentic communication scenarios, including reading passages, instructional material, conversations, and practice questions in textbooks that exemplify the application of common idioms in quotidian contexts (Alasmary, 2025; Lee, 2020). This enables students to observe the practical application of formulaic language, facilitating its internalization and utilization in their communication.

In addition to its role in promoting fluency and comprehension, formulaic language also contributes to learners' pragmatic competence, which refers to the ability to use language appropriately in various social and cultural contexts. For EFL learners, exposure to formulaic language in textbooks provides not only linguistic input but also cultural and situational cues that are critical for successful communication (Jakupčević & Čavar Portolan, 2024; Puimège, 2024). Therefore, incorporating formulaic language into instructional materials plays a vital role in developing learners' overall communicative competence beyond grammatical accuracy.

### 2.2. Lexical bundles

Research in corpus linguistics has evolved to focus on the fundamental aspects of language usage. Lexical bundles have garnered significant attention in corpus linguistics studies due to their capacity to unveil patterns of word pairings that illustrate both language structure and communicative function (Ardi et al., 2023). Biber et al. (1999, 2021) described lexical bundles as sequences of words that regularly appear in texts or speech, functioning as bigger linguistic units than individual words. Lexical bundles denote continuous sequences of words identified by a corpus-driven methodology, utilizing defined frequency and distribution criteria (Chen & Baker, 2010; Lee, 2020; Y. Li & Lei, 2025; Oktavianti & Prayogi, 2022; Oktavianti & Sarage, 2021). For example,

three-word bundles encompass phrases such as "based on the," "in terms of," and "at the moment," frequently employed across many contexts. Four-word phrases like "in the case of" and "as a result of" are also prevalent. Moreover, extended phrases such as "at the end of the" or "as far as I can" are found in specific discourse kinds. This study especially examines three-word lexical bundles, as they are the most frequently utilized in Indonesian EFL textbooks.

Lexical bundles differ from other multiword phrases, including idioms and collocations, because of their specific attributes. Unlike idioms, lexical bundles are not fixed expressions with figurative meanings; instead, they are typically semantically transparent, meaning their meaning can be interpreted directly from the individual words (e.g., in the middle of, one of the) (Sidtis, 2021; Wood, 2015). They also differ from collocations, which are defined based on the strength of association between two or more words, regardless of how frequently they appear in extended sequences (Cutler, 2021; Sidtis, 2021). Lexical bundles, in contrast, are identified purely through frequency and dispersion criteria. They are recurring strings of three or more words that appear repeatedly across texts and perform important discourse or organizational functions (Wood, 2015). Because of this, lexical bundles are considered formulaic, highly common, and structurally restricted, but they are generally not idiomatic or creative expressions. Instead, they represent the building blocks of fluent, natural academic and conversational language.

Lexical bundles are recognized by particular frequency and range/dispersion criteria (Budiwiyanto & Suhardijanto, 2020; Chen & Baker, 2010; Lee, 2020; Y. Li & Lei, 2025). A specific combination of words is deemed a lexical bundle only if it appears with sufficient frequency in a particular corpus. Frequency is an essential attribute of lexical bundles; it determines the minimum occurrence required for a multi-word sequence to be identified as a bundle. However, there is no universal agreement on what counts as a "sufficient" frequency threshold. Different studies adopt different cut-off points depending on corpus size, genre, and research purposes. For instance, (Biber et al. (2021) used a threshold of 40 occurrences per million words, whereas Cortes (2013) and Hyland (2008) allowed much lower frequencies in smaller, specialized corpora. Some research even employs raw frequencies rather than normalized counts, which makes cross-study comparison challenging. This variation demonstrates that frequency, although central to bundle identification, is highly context-dependent and should be interpreted in relation to corpus size and dispersion constraints. Therefore, selecting a frequency threshold requires justification rather than simple adoption of previous studies.

Researchers frequently categorize them based on structural and functional criteria to enhance comprehension of their role in language utilization. The bundles are classified primarily in two ways: by their structural attributes and by their functions (Acorinti & Bocorny, 2024; Alasmay, 2025; Ardi et al., 2023; Bestgen, 2020; Oktavianti & Prayogi, 2022; Oktavianti & Sarage, 2021). The following table summarizes the structural

categories of lexical bundles based on Biber et al.'s (1999, 2021) classification and highlights the grammatical forms that these bundles typically take.

**Table 1**

Structural classification of lexical bundles.

Category	Pattern	Examples
Noun Phrase (NP)-based	1. NP + <i>of</i> - phrase fragment	1. <i>the end of, a part of, the result of, the use of</i>
	2. NP phrase without <i>of</i>	2. <i>a good idea, the next step, the main point, the main idea</i>
Prepositional Phrase (PP)-based	3. Prepositional phrase with embedded <i>of</i> - Phrase Fragment	3. <i>in front of, on top of, at the end of, as part of</i>
	4. Other Prepositional Phrase (fragment)	4. <i>for the purpose, On the part, in the case, by the way</i>
Verb Phrase (VP)-based	5. anticipatory it + verb phrase/adjective phrase	5. <i>It has been, It is important, It will be, It seems to</i>
	6. passive verb + prepositional phrase fragment	6. <i>is based on, is described in, is found in, is described as</i>
	7. copula be + noun phrase/adjective phrase	7. <i>is a part, is the best, are the main, are the key</i>
	8. (verb phrase +) that-clause fragment	8. <i>explains that the, suggests that the</i>
	9. (verb/adjective +) to-clause fragment	9. <i>need to address, try to understand, are used to, more likely to, was able to</i>
	10. adverbial clause fragment	10. <i>when he arrives, before we leave, after they finished</i>

	11. pronoun/noun phrase + be	11. <i>these are examples, this is correct, they are students</i> <i>it is clear</i>
Other	12. other expression	12. <i>for a moment</i> <i>in the meantime</i> <i>at the same</i>

---

Table 1 illustrates the structural classification of lexical bundles based on their grammatical composition. The first major group comprises noun phrase (NP)-based bundles, which typically occur either as NP + *of* fragments—such as *the end of*, *a part of*, and *the result of*—or as NP constructions that do not include *of*, for example *a good idea*, *the next step*, and *the main point*. A second group involves prepositional phrase (PP)-based bundles, which include prepositions followed by embedded structures, as seen in *front of*, *on top of*, *at the end of*, and *as part of*. Other PP-based bundles do not contain *of*, including expressions such as *for the purpose*, *in the case*, and *by the way*.

The largest and most varied category is formed by verb phrase (VP)-based bundles. This group encompasses several sub-patterns, including anticipatory *it* constructions (e.g., *It has been*, *It is important*), passive verb forms followed by prepositional phrase fragments (e.g., *is based on*, *is found in*), and copular structures comprising *be* + noun or adjective phrase (e.g., *is the best*, *are the main*). In addition, VP-based bundles include verb or adjective + *to*-clause fragments such as *need to address* and *try to understand*, as well as *that*-clause fragments introduced by a reporting verb (e.g., *explains that the*, *suggests that the*). Adverbial clause fragments (*when he arrives*, *before we leave*) and pronoun or noun phrase + *be* constructions (*this is correct*, *they are students*, *it is clear*) are also represented within this category. The final group, labeled *Other*, contains lexical bundles that do not conform to the preceding structural types, including expressions such as *for a moment*, *in the meantime*, and *at the same*.

### 3. Method

#### 3.1. Research design

The study utilizes a corpus-based methodology to identify the lexical bundles and assist the structural classification. The corpus-based approach is a methodology that employs a corpus to scientifically study language usage (Csomay & Crawford, 2024; McEnery & Wilson, 2022). McEnery and Wilson (2022) assert that a corpus-based analysis investigates a corpus primarily to evaluate existing linguistic hypotheses or ideas.

#### 3.2. Corpus construction

The corpus consists of EFL textbooks under the *Kurikulum Merdeka* for senior high school published by the Indonesian Ministry of Education. The researcher uses three

textbooks for grade 10, grade 11, and grade 12. The data taken for analysis are instructional content, reading passages, and assessment questions in all chapters of the three books. It excludes the table of contents, acknowledgments, references, authors' biographies, and appendices in the front and back of the books because such elements do not constitute English teaching materials that are directly related to the teaching and learning process.

The first step involved converting the textbooks from PDF format to text format. This step was carried out to facilitate further analysis, as the text format allowed the data to be read and processed by linguistic analysis applications. After the textbooks were successfully converted to text format, the researcher carried out the data cleaning process. The inclusive criteria of this study were the reading passages, instructional content, dialogues, and practice questions in the textbooks. The exclusive criteria included the table of contents, acknowledgments, references, authors' biographies, images, and appendices found in the front and back of the books, which did not contain the lexical bundles that were to be analyzed in this study. By applying these exclusive criteria, the researcher ensured that the analysis remained focused on the relevant linguistic features. The data cleaning process involved removing these exclusive elements that were not relevant to the research.

**Table 2**

Textbook corpus.

No	Textbook Detail	Token
1	Bahasa Inggris: Work in Progress grade 10 (2022) written by Budi Hermawan, Dwi Haryanti, and Nining Suryaningsih	34.211
2	Bahasa Inggris: for Change grade 11 (2022) written by Puji Astuti., et al	36.309
3	Bahasa Inggris: Life Today grade 12 (2022) written by Susanti Retno Hardini, et al	36.607
	Total	107.127

### 3.2. Corpus tool

The corpus tool utilized in the present study is AntConc Version 4.3.1 (Anthony, 2024). This tool was selected due to its easy access and complete features to identify and analyze the lexical bundles. Features that facilitate the identification analysis of lexical bundles are N-Gram and KWIC.

### 3.3. Lexical bundle identification

Lexical bundles are typically identified using frequency-based criteria, which generally include the following: (1) they consist of recurring sequences of three or four words, (2) they must meet a minimum frequency threshold, i.e., appearing at least 40 times per million words in the corpus following Hyland (2008) and Lee (2020), and (3)

they must meet a dispersion requirement, meaning the bundles should appear across multiple sections of the text, commonly in at least three different chapters, to ensure they are not specific to just one part of the material. The required number of chapters for dispersion may vary between studies, usually ranging from three to five (Biber & Barbieri, 2007; Lee, 2020). However, this study employed 3 chapters as the dispersion threshold considering the size of the corpus.

In addition, the corpus tool was set to generate three-word n-grams, allowing the system to automatically extract all recurring three-word sequences from the selected textbooks. This n-gram parameter ensured that only lexical items matching the structural definition of three-word bundles were collected before applying the frequency and dispersion filters.

### *3.4. Data collection*

The data collection process began by preparing the textual content of the selected textbooks in a digital format compatible with corpus analysis. These prepared text files were then imported into AntConc, a corpus analysis software widely used in linguistics to investigate language patterns. By compiling these text files, a corpus was created as the basis for analysis. Within this corpus, the researcher used AntConc's built-in tools to extract all instances of three-word lexical bundles (i.e., recurring sequences of three consecutive words). Each identified three-word combination was automatically detected by the software and meticulously documented.

### *3.5. Data analysis*

After the three-word combinations were identified and confirmed as lexical bundles, the next step in the analysis involved categorizing these bundles based on their structural characteristics. This categorization process was guided by the widely recognized framework developed by Biber et al. (1999, 2021) which provides a comprehensive typology for classifying lexical bundles according to their grammatical structure. Examples of structural classifications in Biber's framework include *noun phrase + of*, *prepositional phrase fragments*, *verb phrase fragments*, and others. Each lexical bundle identified in the corpus was carefully examined and aligned with the appropriate structural category as defined in the framework. This alignment process required detailed grammatical analysis to ensure that the bundles were classified accurately and consistently. The use of Biber's framework not only provided a standardized basis for analysis but also allowed for comparisons with previous corpus-based studies, thereby enhancing the validity of the findings.

### *3.6. Data verification*

This study employed Inter-coder Reliability (ICR) for the data verification technique. ICR is a measure used in research to assess how consistent different researchers (coders) are in assigning codes or labels to the same data (O'Connor & Joffe,

2020). This process involved two lecturers from the English Education Master Program at Ahmad Dahlan University, who were experts in linguistics, serving as validators. The selection of the two lecturers as validators was based on their expertise in linguistics and experience in corpus-based research. Their role as validators involved examining the consistency of the lexical bundle classification, identifying any discrepancies in the coding process, and ensuring that the data encoding was accurate and reliable. Through this process, they provided valuable feedback to refine the categorization and improve the overall accuracy of the analysis.

#### 4. Findings and discussion

This section shows the structural classification of three-word lexical bundles identified in the senior high school EFL textbooks within the *Kurikulum Merdeka*. The lexical bundles are classified into four primary structures according to Biber's framework: NP-based, PP-based, VP-based, and other expressions.

**Table 3**

The result of the structural classification of lexical bundles.

Category	Pattern	Phrases
NP based	1. NP + <i>of</i> – phrase fragment	modelling of text, groups of four, one of the, joint construction of, building knowledge of, of the field, the purpose of, construction of text, the story of, one of the, part of the
	2. NP phrase without <i>of</i>	The following questions, the end of, learning objective by, objective by the, this lesson you, the following questions, related texts activity, a procedure text, analytical exposition text, field activity work, the following picture, the main idea, continuing and ending, lesson you are, your social media, The field activity, The following statements
PP based	3. Prepositional phrase with embedded <i>of</i> - Phrase Fragment	of this lesson, end of this, in groups of, of text activity, knowledge of the, of the text, a group of, of the world
	4. Other Prepositional Phrase (fragment)	based on the, work in pairs, in your notebook, based on the, about how to, at the following, so far in, for

		starting continuing, By the end, on social media, based on the, of social media, according to the, on the internet
VP based	5. anticipatory it + verb phrase/adjective phrase	-
	6. passive verb + prepositional phrase fragment	number has been, has been done, been done for
	7. copula be + noun phrase/adjective phrase	you are able, what is the
	8. (verb phrase +) that-clause fragment	-
	9. (verb/adjective +) to-clause fragment	be able to, you need to, you want to, watch the video, are able to, listen to the, we need to
	10. adverbial clause fragment	work in groups, look at the, read the following
	11. pronoun/noun phrase + be	have you learned, your teacher will, you have learned
Other	12. other expression	publish your writing, answer the following, follow the instructions, visit the link, scan the following, answer the question, to watch the, what is this, do you think, why do you, what is the

The following are the results of the findings for each category, which have been systematically organized and presented according to grade level. Each grade is analyzed separately to provide a clearer understanding of the patterns and variations observed across the categories.

#### 4.1.1 NP-based bundles

NP-based (Noun Phrase-based) bundles are defined as lexical groups that center around a noun phrase, which functions as the core unit in the bundle. The phrases from the three books categorized as NP-based are presented in Table 4.

**Table 4**  
NP-Based bundles.

Grade 10		Grade 11		Grade 12	
Bundle	Freq	Bundle	Freq	Bundle	Freq
the following questions	15	construction of text	61	the story of	13
		the end of	37	one of the	12
		modelling of text	32	Part of the	10
		the following questions	31	the following statements	10
		joint construction of	30		
		learning objective by	30		
		objective by the	30		
		this lesson you	30		
		lesson you are	30		
		building knowledge of	29		
		field activity work	28		
		the field activity	28		
		your social media	24		
		groups of four	21		
		related texts activity	21		
		analytical exposition text	21		
		a procedure text	20		
		one of the	17		
		the main idea	16		
		the purpose of	15		
		continuing and ending	15		
		the following picture	14		

Based on table 4, there is only one bundle categorized as NP-based in grade 10's textbook, i.e., 'the following questions,' with a frequency of 15 occurrences. Some examples are below.

- (1) Listen to the audio recording again and discuss the following questions
- (2) Discuss the answer to the following questions with a partner, then share it with the class
- (3) Plan your challenge by answering the following questions

Meanwhile, 21 NP-based bundles were found in grade 11. The most frequently occurring bundle is "construction of text" with 61 occurrences, followed by "the end of" in second place with 37 occurrences, and in third place, "modelling of text" with 32

occurrences. The bundle with the least occurrences in this category is "the following picture" with 14 occurrences. Some examples are mentioned below.

- (4) Joint construction of text
- (5) By the end of this lesson, you are able to correctly identify the main ideas and author's purpose in an analytical exposition text.
- (6) Choose one of the topics
- (7) What's the purpose of the content creator's blog post?
- (8) Look at the following picture about different fruits from Indonesia.

As for 12 grade, four NP-based bundles were found. The most frequently occurring bundle is "the story of" with 13 occurrences, and the bundles with the least occurrences in this category are "part of the" and "the following statement" with 10 occurrences. Below are some examples from grade 12.

- (9) To entertain readers with the story of Kilion and his Amdui Village
- (10) To share the story of West Papua and its abundant solar energy
- (11) Choose one of the following topics. You may also use your own topic
- (12) Could the following statements represent what is told in the story?

4.1.2. PP-based bundles

PP-based bundles begin with prepositions and function to show relationships such as place, time, or context in a text. The phrases from the three books categorized as PP-based are as follows.

**Table 5**

PP-based bundles.

Grade 10		Grade 11		Grade 12	
Bundle	Freq	Bundle	Freq	Bundle	Freq
based on the	18	in your notebook	41	on social media	24
of the story	10	based on the	37	based on the	18
		at the following	22	according to the	11
		so far in	16	on the internet	11
		by the end	32		
		in groups of	30		
		about how to	21		
		of the field	29		

Table 5 shows that there are two PP-based bundles identified in grade 10. Some examples of PP-based bundles in grade 10 are shown in the following sentences.

- (13) Answer the questions based on the conversation.
- (14) Make necessary changes in your writing based on the checklist.
- (15) Based on the writing prompt or health problems faced by students

- (16) Decide what title and part of the story is read to you
- (17) Part of the story you hear

Additionally, there are seven bundles that are categorized as PP-based in grade 11. Examples of the use of PP-based bundles in grade 11 books are shown in the following sentence.

- (18) In your notebook, write down what you have learned
- (19) Write the results of your observation in your notebook
- (20) Look at the following Instagram photo and caption
- (21) Look at the following graphic organizer about healthy life
- (22) By the end of this lesson
- (23) By the end of this week

Meanwhile, there are four bundles that are categorized as other PP-based in grade 12. Some of PP-based bundles in grade 12 are shown in these sentences.

- (24) Why does hate speech on social media platforms may lead to violence?
- (25) Why You shouldn't overshare personal information on social media
- (26) Answer the questions based on the text
- (27) Your teacher will ask you a question based on the number you've chosen
- (28) Feel free to browse on the internet to get some ideas
- (29) the correct or acceptable way of communicating on the internet.

4.1.3. VP-based bundles

VP-based bundles are lexical bundles that start with verbs and are used to convey actions, processes, or relationships in discourse. The phrases from the three books categorized as PP-based are as follows.

**Table 6**  
VP-based bundles.

Grade 10		Grade 11		Grade 12	
Bundle	Freq	Bundle	Freq	Bundle	Freq
be able to	10	learned so far	43	listen to the	11
you need to	19	look at the	42	be able to	10
you want to	10	read the following	32		
		number has been	31		
		you are able	30		
		are able to	30		
		you learned so	30		
		have you learned	30		
		work in groups	30		

your teacher will	29
You have learned	25
watch the video	24
has been done	21
been done for	21
pay attention to	14
able to correctly	13

---

Table 6 shows that three bundles are categorized as VP-based in grade 10, with the most frequent being the phrase "you need to," occurring 19 times. The phrases "be able to" and "you want to" each occur 10 times. Some examples of the bundles are presented below.

- (30) After learning the materials in this chapter, you are expected to be able to.
- (31) Visit the link to download the worksheet you need to do in each task
- (32) If you have to choose, do you want to stay healthy or do you want to stay happy?

Moreover, there are 16 lexical bundles in grade 11. The most frequently occurring bundle is "learned so far" with 43 occurrences, followed by "look at the" with 42 occurrences. Some examples are written below.

- (33) In the box provided, write down what you have learned so far in this lesson.
- (34) Look at the following picture, and answer this question, "What does the statement mean?"
- (35) Scan the following QR Code or visit the link to watch the video
- (36) Watch the video about the explanation of analytical exposition text that your teacher will play
- (37) By the end of this lesson, you are able to actively demonstrate an analytical exposition text through a group debate.

As for grade 12, there are two bundles categorized in this type, namely "listen to the" and "be able to." Some examples are presented as follows.

- (38) Before you listen to the audio, find out the meaning of these phrases
- (39) You will listen to the conversation between two people
- (40) This would make people be able to do more activities at night without any limitation
- (41) When you're registered, you'll be able to pay for purchases using your new account details.

#### 4.1.4. Other expressions

Other expressions refer to various word combinations used within NP, VP, and PP that are not necessarily recurring as fixed bundles. In the Indonesian EFL textbooks under *Kurikulum Merdeka*, several other expressions have been found, such as *do you think*, *publish your writing*, *answer the following*, *follow the instructions*, *visit the link*, *scan the following*, *answer the question*, *what is this*, *what is the*. Some examples of other expressions are mentioned below.

- (42) What do you think are two most important traits that should be owned by great athletes?
- (43) You can publish your writing in the media of your choice.
- (44) Why do you need to hold this event?

## 5. Discussion

The study's findings indicate that the predominant structural classification is NP-based, comprising 26 phrases. This aligns with the findings of Ardi et al. (2023), Oktavianti and Sarage (2021) and Gil and Caro (2019) which indicate that NP is the most prevalent type of lexical bundle identified in their research. The research conducted by Ardi et al. (2023) focused on five Indonesian English textbooks. The study recorded 374 occurrences of NP-based instances. The most prevalent noun phrase was the NP-based with-of-phrase fragments, which occurred 173 times. This research does not identify the term most frequently utilized within NP-based bundles. Furthermore, Oktavianti & Sarage (2021) identified that the predominant 3-gram lexical bundles are NPs. The phrase "the use of" appears most frequently, with 105 instances. Furthermore, Gil & Caro (2019) analyzed dissertations authored by native Spanish students. The findings indicated that the most used term was NP-based. The most often occurring NP lexical bundle is "the use of," with 85 instances.

In contrast to the three previous studies, the current study revealed that the most prevalent phrases are "construction of text" with 61 instances, "the end of" with 37 instances, "modeling of text" with 32 instances, and "the following questions" with 31 instances. This indicates that NP-based structures predominate in the utilization of lexical bundles, both within the setting of English textbooks in Indonesia and in students' argumentative writing, as well as among Spanish students. This dominance affirms that NP-based lexical bundles play a crucial role in facilitating sentence structuring and conveying coherent meaning within the text.

The similarity of the results with Oktavianti and Sarage's (2021) research also strengthens this finding, confirming that NP-based bundles tend to characterize texts with informative or descriptive functions, both in education and academic writing. This shows the relevance and consistency of the findings across different text types in Indonesia. The differences in the most frequent bundles may reflect differences in learning focus, topic, or writing style presented in textbooks or student essays.

The examination of structural classification indicates that NP-based bundles are the most frequent, especially in grade 11 textbooks. This indicates a focus on academic language and formal text structure at this grade level, with expressions such as "construction of text" and "the following questions" commonly utilized. Such lexical bundles are crucial for students to cultivate a coherent and organized methodology in academic writing. The prevalence of NP-based bundles in grade 11 raises concerns about the diversity of sentence structures. Students who frequently depend on NP-based bundles may produce sentences that are overly simplistic or exhibit insufficient syntactic diversity. This may restrict their ability to compose more intricate texts.

The study's findings indicate that the grade 10 EFL textbook has a wider variety of bundles, suitable for younger students who are still enhancing their cognitive and linguistic abilities. Textbooks should integrate a broader array of structural kinds, including VP-based and PP-based bundles, to foster students' adaptability in both academic and quotidian communication. The lexical bundles in English textbooks for grades 10 and 11 demonstrate variations in frequency, structure, and function that correspond to the educational objectives at each level. In grade 10, lexical bundles such as "do you think" and "you need to" are more prevalent, emphasizing direct engagement and fostering critical thinking abilities through inquiries and directives. The utilization of noun phrase fragments, such as the subsequent inquiries, enhances the understanding of particular items. In grade 11, complexity and diversity increase, as seen by terms like "construction of text" and "joint construction of," which highlight analysis and collaboration in text comprehension. The prevalence of prepositional phrase fragments, such as "in groups of," signifies an emphasis on collaborative efforts. These contrasts illustrate the evolving cognitive and educational requirements of pupils, with materials designed to progressively develop more sophisticated and comprehensive skills.

Interestingly, the number of lexical bundles identified in the Grade 12 textbook was lower compared to Grades 10 and 11. This could be attributed to several possible explanations. First, Grade 12 materials tend to focus on more advanced and extended writing tasks, where ideas are expressed using more complex sentences or individualized wording rather than highly repetitive formulaic sequences. As students' progress to the final year of secondary school, textbooks may intentionally reduce dependence on prefabricated expressions to encourage greater syntactic flexibility, critical thinking, and independent language production. Nevertheless, this assumption should be validated by comparing the data with relevant reference corpora to determine whether the decreased frequency represents a broader pattern or is unique to this particular textbook.

In general, this study has some implications in the field of language education, particularly in the analysis of lexical bundles. The result of the study can contribute to the theoretical understanding of lexical bundles in educational contexts, particularly in Indonesian secondary school textbooks. For curriculum designers and educators, the findings of this study can provide valuable insights into how lexical bundles are used across different grade levels in Indonesian secondary school textbooks. Curriculum

designers can use this information to create materials that better reflect natural language usage, aiding in developing language learning strategies that emphasize formulaic language. The teachers can benefit from this research by incorporating lexical bundles into their teaching to help improve students' comprehension and production of language. For language researchers, the study offers theoretical contributions to understanding lexical bundles in educational contexts, providing researchers with empirical data that can support further exploration of language acquisition and usage-based linguistics theories.

## 6. Conclusion

This study identified 53 distinct three-word lexical bundles within Indonesian senior high school EFL textbooks aligned with the *Kurikulum Merdeka*. The analysis revealed a clear predominance of noun phrase (NP)-based bundles, a structural pattern consistent with a pedagogical emphasis on analytical and explanatory language. Variations across grade levels further suggest a deliberate pedagogical progression: Grade 10 materials feature a higher proportion of discourse-organizing and interactive bundles to foster engagement, while Grade 11 exhibits a marked increase in text-oriented, academic constructions that scaffold analytical writing. The comparatively lower recurrence of core bundles in Grade 12 may signify a curricular shift towards promoting more varied and independent language use.

The findings of this study are subject to certain limitations. The exclusive focus on three-word bundles precludes insights into longer, potentially more sophisticated multi-word sequences characteristic of advanced academic prose. Furthermore, as the analysis is confined to a single national curriculum, the generalizability of the results to other Indonesian EFL contexts is constrained. Future research would benefit from incorporating four-word bundles, conducting comparative analyses across different curricula or educational regions, and utilizing reference corpora to determine the extent to which these observed patterns reflect broader linguistic or pedagogical trends. Finally, classroom-based observational or interventional studies are recommended to elucidate the processes of bundle acquisition and application by learners, thereby providing actionable insights for instructional practice and curriculum design.

## Acknowledgments

This work was funded by Directorate of Research and Community Service (*Direktorat Penelitian dan Pengabdian kepada Masyarakat*), Ministry of Higher Education, Science, and Technology, Indonesia, under the Grant Number 126/C3/DT.05.00/PL/2025, 0498.12/LL5-INT/AL.04/2025, 060/PTM/LPPM.UAD/VI/2025. The authors express sincere gratitude to Universitas Ahmad Dahlan for the support.

## References

- Acorinti, W. H., & Bocorny, A. E. P. (2024). This paper is organized as follows: Lexical bundles in computer science academic texts produced by novice and expert writers. *Journal of the Undergraduate Linguistics Association of Britain*, 2(2), 195-216. <https://doi.org/10.5281/zenodo.12738429>
- Adawiyah, S. Z., & Oktavianti, I. N. (2023). Gender representation in Merdeka curriculum ELT textbooks: A corpus-assisted study. *Indonesian Journal of EFL and Linguistics*, 8(2), 313-326. <https://doi.org/10.21462/ijefl.v8i2.668>
- Aini, Q. & Adiyono. (2023). Implementation of an independent curriculum in supporting students' freedom to create and learn. *Journal of Scientific Research, Education, and Technology (JSRET)*, 2(3), 999-1008. <https://doi.org/10.58526/jsret.v2i3.187>
- Akhofullah, A., & Oktavianti, I. N. (2023). The frequency distribution of central modal verbs in Kurikulum Merdeka's textbooks in Indonesia: A corpus-based analysis. *PROJECT (Professional Journal of English Education)*, 6(6), 1243-1249.
- Alasmary, A. (2025). Lexical bundles in psychology lectures and textbooks: A contrastive corpus-based study with implications for academic writing. *Frontiers in Psychology*, 16: 1-17. <https://doi.org/10.3389/fpsyg.2025.1545355>
- 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), 1-10. <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), 25-39. <https://doi.org/10.17323/jle.2023.16305>
- Bao, K., & Liu, M. (2024). A contrastive study of lexical bundles expressing gratitude in dissertation acknowledgments produced by Chinese and American PhD students of linguistics. *Sage Open*, 14(1). <https://doi.org/10.1177/21582440241239164>
- 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), 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of Spoken and Written English*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.232>
- Binti Mohamed Ismail, N. F., Binti Kamarulzaman, A. S., Binti Zainol, N. Z., Binti Nordin, N. H., & Binti Mohamad Jamil, K. A. (2024). Lexical bundle patterns in Malaysian educational texts: A corpus-based analysis of English language textbooks. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 9(9), 1-12. <https://doi.org/10.47405/mjssh.v9i9.2986>

- Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, 10(2), 292-303. <https://doi.org/10.17509/ijal.v10i2.28592>
- Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 20, 38–52. <https://doi.org/10.1016/j.jeap.2017.10.008>
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49. <https://doi.org/10.64152/10125/44213>
- 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>
- Csomay, E., & Crawford, W. J. (2024). *Doing Corpus Linguistics* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003363309>
- 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.
- Hassanzadeh, M., & Tamleh, H. (2023). The use of lexical bundles by native English authors in applied linguistics: A corpus-driven study. *LLR 2023*, 13(6), 539-564. <https://doi.org/10.29252/LRR.13.6.17>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1). <https://doi.org/10.1016/j.esp.2007.06.001>
- 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/eej.v11i1.43411>
- Intamanjana, P. D., & Sada, C. (2021). A corpus-based analysis of lexical collocation in conversation texts of English textbooks. *Journal of English as a Foreign Language Education*, 2(2), 68-75. <https://doi.org/10.26418/jefle.v2i2.51634>
- Jakupčević, E., & Čavar Portolan, M. (2024). An analysis of pragmatic content in EFL textbooks for young learners in Croatia. *Language Teaching Research*, 28(1), Article 1. <https://doi.org/10.1177/1362168820986936>
- Lasmita, R., Harahap, A., & Arsyad, S. (2023). Lexical bundles in reading passages of English textbook for senior high school: A comparative study between three textbooks of different grades. *Edu-Ling: Journal of English Education and Linguistics*, 6(2), 157-164. <https://doi.org/10.32663/edu-ling.v6i2.3495>
- Lee, H.-K. (2020). Lexical bundles in linguistics textbooks. *언어연구*, 37(1), 121-145. <https://doi.org/10.17250/KHISLI.37.1.202003.005>
- Lestari, E. S., Oktavianti, I. N., & Aziz, R. A. (2025). Functional categories of lexical bundles in Indonesian EFL Textbooks: A corpus-based study. *Indonesian Journal of EFL and Linguistics*, 10(1), 195–211. <https://dx.doi.org/10.21462/ijefl.v10i1.907>
- Li, L. X. (2022). Improving modal verb treatment in English textbooks used by Chinese learners: A corpus-based approach. *SAGE Open*, 12(1). <https://doi.org/10.1177/21582440221079918>

- Li, Y., & Lei, H. (2025). Lexical bundles in L1 and L2 English academic writing: Convergent and divergent usage. *SAGE Open*, 15(2). <https://doi.org/10.1177/21582440251333850>
- McEnery, T., & Wilson, A. (2022). *Corpus linguistics*. Edinburgh University Press. <https://doi.org/10.1515/9781474470865>
- Meng, F., Zheng, Y., Bao, S., Wang, J., & Yang, S. (2022). Formulaic language identification model based on GCN fusing associated information. *PeerJ Computer Science*, 8, e984. <https://doi.org/10.7717/peerj-cs.984>
- Navarro Gil, N., & Martínez Caro, E. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language & Literature*, 12(1), 65-90. <https://doi.org/10.5565/rev/jtl3.794>
- Nisa, R. (2024). Implementation of an independent learning curriculum in enhancing student creativity in primary schools. *Indonesian Journal of Education (INJOE)*, 4(2), 555–567.
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19. <https://doi.org/10.1177/1609406919899220>
- Oktavianti, I. N., & Fajria, A. (2021). Modal verbs in a curriculum-based EFL textbook of senior high school in Indonesia: A corpus-based study. *TESOL International Journal*, 16(5).
- Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education*, 9(2), 761-783. <https://doi.org/10.24815/siele.v9i2.23995>
- Oktavianti, I. N., Radzuan, N. R. M., Pratolo, B. W., Surono, S., Budiwati, T. R., & Aisyah, S. (2025). Modal verbs in Indonesian and Malaysian English textbooks for secondary schools: A corpus-driven study. *Studies in English Language and Education*, 12(1), 20–33. <https://doi.org/10.24815/siele.v12i1.35258>
- 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), 509-534. <https://doi.org/10.21462/ijefl.v6i2.421>
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60–71. <https://doi.org/10.1016/j.jeap.2015.11.003>
- Puimège, E. (2024). Learning L2 formulaic sequences from meaning-focused activities. *ITL - International Journal of Applied Linguistics*, 175(2), 163–186. <https://doi.org/10.1075/itl.23014.pui>
- 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>
- Wahyudin, D. (2024). *Kajian akademik kurikulum merdeka*. Pusat kurikulum dan pembelajaran, badan standar, kurikulum, dan asesmen pendidikan Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. Bloomsbury Academic.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wulandari, Y., S, R., & Ilham, D. (2024). Unleashing student creativity: A dynamic look at merdeka belajar curriculum's impact. *International Journal of Asian Education*, 5(1), 21–33. <https://doi.org/10.46966/ijae.v5i1.371>