

# Lexical Bundles in English Language Textbooks: A Systematic Review

Mingxuan Zhang<sup>1</sup>, Jayakaran Mukundan<sup>1</sup>, Laleh Khojasteh<sup>2</sup>, Jasmine Anak Jain<sup>1</sup>

<sup>1</sup> School of Education, Faculty of Social Sciences and Leisure Management, Taylor's University, Selangor, Malaysia

<sup>2</sup> Department of English Language, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

Correspondence: Mingxuan Zhang, School of Education, Faculty of Social Sciences and Leisure Management, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia. E-mail: 0365972@sd.taylors.edu.my

Received: December 28, 2024

Accepted: March 26, 2025

Online Published: April 28, 2025

doi:10.5430/wjel.v15n6p100

URL: <https://doi.org/10.5430/wjel.v15n6p100>

## Abstract

Lexical bundles (LBs) are recurrent continuous sequences of two or more words that frequently appear in discourse. Acquiring these bundles helps compensate for short-term memory limitations, supports efficient language processing, and promotes both comprehension and production in English. Lexical bundles can be learned through English language (EL) textbook input. Numerous researchers advocate for the inclusion and emphasis of native-like lexical bundles in EL textbooks; however, relatively few studies have evaluated lexical bundles in these textbooks. This study seeks to review the literature on lexical bundles in EL textbooks over the past two decades (2004-2024) to identify research trends and gaps. The review relied on three databases: Web of Science Core Collection, Scopus, and ERIC. After a rigorous screening process, 18 relevant studies were identified and coded based on study identification, language context, research aims, textbook types, methodologies, target lexical bundles, and main findings (cf. Appendix A). The findings recommend future research in this domain continue to employ the corpus-assisted method to investigate lexical bundles in EL textbooks and conduct a comprehensive evaluation using multiple analytical dimensions to explore three- and four-word bundles in foundational-level English for General Purposes (EGP) textbooks and English for Specific Purposes (ESP) textbooks.

**Keywords:** lexical bundles, English language textbooks, systematic review

## 1. Introduction

English language (EL) textbooks occupy a central and multifaceted role in the ecology of language education. They are not merely instructional tools but act as the visible heart of English language instruction for both teachers and learners (Sheldon, 1988). They represent a blend of disciplinary and instructional discourse (Wood & Appel, 2014) and constitute the primary source of English language input for learners (Tomlinson, 2011). This is particularly salient in contexts such as China, where opportunities for naturalistic language exposure are limited, rendering textbooks a critical medium for language acquisition (Xu et al., 2022). In such contexts, textbooks not only scaffold classroom language practice (Nguyen, 2011) but also function as the most direct and substantial channel for formal language development (Yang & Coxhead, 2022). The selection of EL textbooks significantly impacts the quality of English teaching and learning (Mukundan & Kalajahi, 2013).

Nevertheless, the pedagogical efficacy of EL textbooks has been increasingly scrutinized. Concerns have been raised about the tendency of textbook authors to reproduce previously successful materials or to rely on subjective intuition, rather than adhering to systematic pedagogical frameworks grounded in explicit objectives and evidence-based methodologies (Tomlinson, 2012b; Prowse, 2011). These practices contribute to recurring issues, including factual inaccuracies (Litz, 2005), constrained opportunities for authentic language development (Tomlinson, 2008), and oversimplified representations of linguistic knowledge (Allan, 2017), thereby limiting their responsiveness to the diverse and evolving needs of learners (Tomlinson, 2012a).

Lexical bundles (abbreviated as LBs) are phraseological units within the broader category of formulaic sequences. They refer to recurrent continuous sequences of two or more words found in discourse, such as *"as well as"*, *"and this is a"*, *"do you want me to"* etc. Previous studies have investigated lexical bundles by comparing their usage across individuals with different English proficiency levels (e.g., Nguyen, 2023), examined their role in language processing (e.g., Appel et al., 2024), compiled several lexical bundle lists to support English instruction and learning (e.g., Alasmarty, 2024), as well as explored the features and variations of these bundles across different genres (e.g., Gholaminejad, 2021) and disciplines (e.g., Esfandiari & Barbary, 2023). Moreover, previous studies have also summarized their structural and functional classifications across various texts. For instance, Biber et al. (1999) categorized the structures of lexical bundles in academic prose and conversation. Based on Biber et al.'s (1999) work, Hyland (2008) refined structural classifications into verb phrase-related bundles, clause-related bundles, and noun/preposition-related bundles, and functional classifications were categorized as research-oriented, text-oriented, and participant-oriented.

As pre-fabricated elements (Bolinger, 1976), lexical bundles constitute approximately 70% of language expression in adult native speakers (Altenberg, 1990). They are essential for achieving fluent and idiomatic proficiency in English (Hyland & Jiang, 2018), as

mastery of lexical bundles helps compensate for memory limitations (Wray & Perkins, 2000). Moreover, it enables learners to process language more rapidly (Northbrook & Conklin, 2019), facilitating fluent and accurate comprehension and production in English (Boers & Lindstromberg, 2012). For English learners who primarily rely on textbooks as their primary input source, the quality of textbooks directly influences the effectiveness of English language acquisition (Northbrook & Conklin, 2018). Only when input materials are real and natural can they foster more authentic English acquisition and output (Ma & Meng, 2021). Previous research has also highlighted issues with lexical bundles in EL textbooks, such as a lack of representativeness of native English usage (McAleese, 2013). A primary reason for this discrepancy is that textbook developers often rely on intuition, and experience, leading to potentially arbitrary selections of lexical bundles in EL textbooks (Prowse, 2011; Reppen, 2010). This mismatch of lexical bundles in textbooks and native English hinders learners from developing proficiency at a native-like level (Northbrook & Conklin, 2019). Given these pedagogical and linguistic concerns, the systematic analysis and evaluation of lexical bundles in EL textbooks is not merely desirable but essential. Such analysis can inform curriculum development, improve textbook design, and support more effective instructional practices. However, despite the proliferation of individual studies, there is a conspicuous lack of a comprehensive synthesis of this body of work. Key questions remain unresolved: What analytical frameworks and methodological approaches have researchers employed? Which learner populations and educational stages have been prioritized? What types of EL textbooks have been subject to scrutiny, and along what evaluative dimensions? To date, no systematic review has addressed these issues.

To fill this critical gap, the present study aims to conduct a comprehensive systematic review of research on lexical bundles in EL textbooks. By synthesizing existing findings and identifying areas of convergence and divergence, this review seeks to provide a robust foundation for future research and practice in EL textbook development and language pedagogy.

A systematic literature review begins with clearly predefined questions, involves the identification of all relevant previous studies, and conducts an explicit, systematic, and critical appraisal of these studies, followed by a comprehensive synthesis of the findings (Gough et al., 2017; Khan et al., 2003). This study adheres to the PRISMA 2020 statement (Page et al., 2021a), the latest framework applicable to various types of systematic reviews, including those on social and educational interventions. Since explicit and precise research questions are fundamental to the review process (Khan et al., 2003), this study is guided by the following questions:

- (1) What methodologies have been employed in studies on lexical bundles in English language textbooks?
- (2) What contextual features (English language contexts & educational levels) have been examined in these studies?
- (3) What types of textbook samples have been analyzed in these studies?
- (4) What are the length and analytical dimensions of the target lexical bundles examined in these studies?
- (5) What research trends and gaps have been identified in these studies?

These research questions address the gap in understanding how lexical bundles are represented and evaluated in EL textbooks. Given that many EL learners rely on textbooks as their primary source of language input, examining these questions is crucial for English learners to foster native-like fluency and enhance language acquisition, particularly in non-native English contexts.

**2. Method**

*2.1 Identification of Databases and Search Strings*

The literature search for this review was conducted using academic databases, including the Web of Science Core Collection, Scopus, and the Educational Resources Information Centre (ERIC). Both Web of Science and Scopus are comprehensive citation databases that cover a wide range of academic disciplines, facilitating a better summary of research trends. ERIC is a specialized citation database focused on the field of education, which further ensures the specificity of this review. The key search terms used were lexical bundles and English textbooks.

Due to the extensive terminology associated with phraseological units in phraseology, which encompasses more than fifty terms (Wood, 2015), this study draws on the method of Pearson and Abdollahzadeh (2023) to establish a comprehensive set of search terms. The initial search terms “lexical bundles” and “English textbooks” were applied to the article title, abstract, and keywords in the Scopus database. Subsequently, the author-provided keywords in these articles were analyzed, refined, and compiled into the final search strings. These search strings, in combination with Boolean operators, were then applied for a preliminary literature search across three databases. The detailed search strings and Boolean operators are provided in Table 1.

Table 1. Search strings and Boolean operators used in preliminary literature retrieval

Database	Search strings
Web of Science Core Collection	TITLE("formulaic language" OR "formulaic sequences" OR "multi-word expressions" OR
Scopus	"multi-word formulaic sequences" OR "clusters" OR "n-grams" OR "lexical chunks" OR
ERIC	"lexical bundles")
	AND
	TITLE("English textbooks" OR "textbooks" OR "coursebook" OR "school textbook" OR
	"language textbook" OR "textbook evaluation" OR "textbook analysis" OR "reference books")

*2.2 Study Retrieval, Eligibility Assessment and Selection*

The eligibility criteria for the retrieval need to be specified (Page et al., 2021b). First, the publication year is limited to the past 20 years,

with a cut-off date from 2004 to 2024 (the search process was conducted in October 2024). Second, the language of the study must be in English. Additionally, the study types are restricted to published empirical research found in academic journal articles, book chapters, and conference proceedings, all of which are considered primary repositories for empirical studies (Pearson & Abdollahzadeh, 2023). Furthermore, the research must fall within the fields of education, linguistics, or language learning. Finally, the studies must focus on the specific types of lexical bundles defined in this study and their occurrence in English textbooks. The aforementioned criteria are detailed in Table 2.

Table 2. Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Publication year	2004 to 2024	2004 and earlier
Language	English	Non-English
Research types	Published empirical academic journal articles, book chapters & conference proceedings	Unpublished non-empirical studies
Research domains	Education, Linguistics & Language learning	Non-Education, Linguistics & Language learning
Research subject	Lexical bundles in English textbooks	Other formulaic languages in non-English textbooks

Based on the aforementioned eligibility criteria and using the search strings across three databases, a total of 182 studies were yielded (86 from Web of Science, 18 from Scopus, and 78 from ERIC).

According to the PRISMA 2020 guidelines (Page et al., 2021a), the entire study selection process is illustrated in the PRISMA 2020 flow diagram (cf. Figure 1).

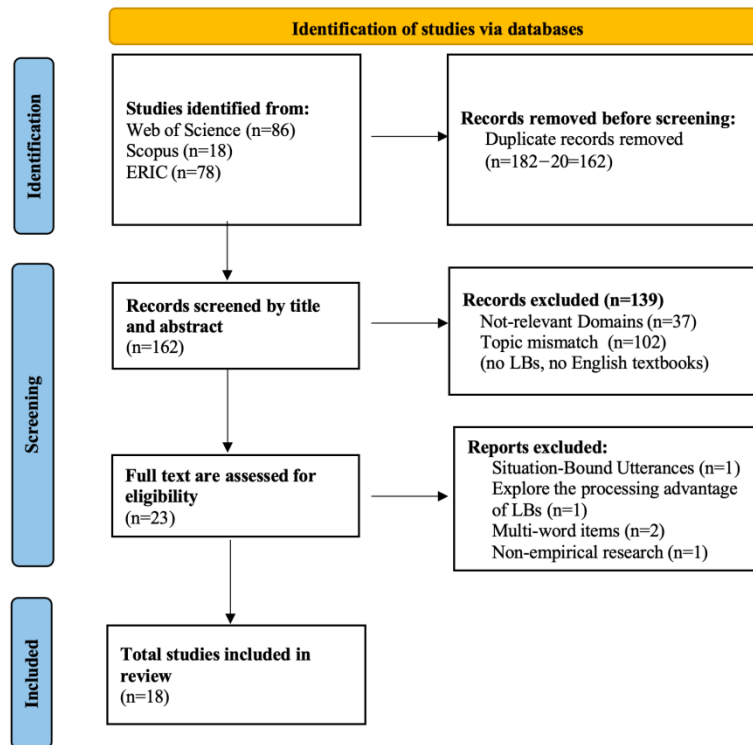


Figure 1. PRISMA diagram of the study selection process

First, all studies from the three databases were compiled, and duplicate records were removed (n=20). Next, the titles and abstracts of the studies were read to exclude those outside the domains of Education, Linguistics and Language Learning domains, as well as those which focused on non-English textbooks (n=139). This process left a total of 23 articles. Finally, the full texts were examined, excluding one study on situation-bound utterances (Ding & Chen, 2019), one on the processing advantage of lexical bundles (Northbrook & Conklin, 2019), two focusing on multi-word items (Hsu, 2006; Koprowski, 2005), and one non-empirical study (Xu, 2016), resulting in 18 articles were selected as the focus of this review.

### 2.3 Data Coding and Analysis

The coding scheme adopted in this study is informed by the framework developed by Plonsky and Gass (2011), as well as relevant previous review studies (e.g., Liu & Brown, 2015). The final coding scheme comprises the following eight categories (cf. Table 3 for

details).

Table 3. Coding scheme

Categories	Variables	Values
Study identification	Author(s)	Open-ended
	Title	Open-ended
	Year of publication	Open-ended
	Source of publication	Journal articles, book chapters & conference proceedings
Context	English language context	Native, ESL, EFL
	Educational level	Primary school, Secondary school (junior & senior), Vocational school, University
Research aims	/	Open-ended
Type of Textbooks	EGP	Yes/No
	EAP	Yes/No
	ESP	Yes/No
Methodologies	Corpus analysis	Yes/No
Target LBs	Length	2-word, 3-word, 4-word etc.
	Analytical dimensions	Open-ended
Main findings	/	Open-ended

Each category in the coding scheme is selected to capture distinct aspects of lexical bundle research in EL textbooks, enabling detailed cross-study comparisons based on various variables. For example, the ‘*English language context*’ category classifies studies based on the English language settings (e.g., ESL, EFL, or native English contexts), which allows for analysis of how lexical bundles are represented across textbooks for learners with different language backgrounds. For instance, the study by Hoang and Crosthwaite (2024) is coded as ‘*EFL*’ due to its focus on Vietnamese EL textbooks, while the study by Wood and Appel (2014) is coded as ‘*native*’ as it examines textbooks within a native English context in Canada. The ‘*methodologies*’ category is used to identify whether corpus analysis serves as the primary research method, given its prominence in this field. Studies employing tools such as AntConc (Anthony, 2018) or SketchEngine (Kilgarriff et al., 2014) for bundle extraction are coded under ‘*corpus analysis*’, indicating the reliance on empirical, data-driven approaches. For instance, Priyatno et al. (2023) utilized AntConc to examine lexical bundle frequency and structures, making corpus analysis central to their methodology. This systematic coding facilitates a comprehensive analysis of research trends and key focus areas within the field.

The coding process was conducted by the lead author, who ensured proper attribution and performed iterative checks on all reviewed articles. To mitigate the risks of unreliability and misinterpretation, all variables underwent iterative verification by the lead author following a two-week interval.

### 3. Results

#### 3.1 Publication Year, Source and Methodology

A review of research data from the past twenty years reveals a fluctuating trend in this field (cf. Table 4 for details). In terms of publication years, 2024 saw two studies (Hoang & Crosthwaite, 2024; Wang & Chan, 2024), while 2023 and 2022 each had one (Alasmay, 2022; Priyatno et al., 2023). In 2021 and 2020, there were three (Gholaminejad, 2021; Hussain et al., 2021; Shirazizadeh & Amirfazlian, 2021) and two (Lee, 2020; Siricharoen & Wijitsopon, 2020) studies published, respectively. The years 2018, 2015, and 2004 each saw one study (Biber et al., 2004; Northbrook & Conklin, 2018; Peng, 2016), while 2010 had two (Chen, 2010; Wood, 2010). Notably, 2014 marked a peak in research activity, with four studies published (Hsu, 2014; Miao, 2014; Ribeck & Borin, 2014; Wood & Appel, 2014), indicating a particularly active year for the field. In comparison, there were seven studies (38.88%) published between 2004 and 2014, and eleven studies (61.11%) between 2015 and 2024.

In terms of publication sources, journal articles dominate the field, accounting for thirteen studies (72.22%). Book chapters make up three studies (16.66%), while conference proceedings represent two studies (11.11%). This distribution indicates that journal articles are the primary medium for studies in this field, reflecting the emphasis on high-quality, peer-reviewed output. At the same time, the presence of book chapters and conference papers highlights the efforts of researchers to share their work across various platforms.

In terms of research methodology, all selected studies (100%) have employed corpus analysis. The application of this method typically involves various corpus concordance software, such as AntConc (Anthony, 2018), SketchEngine (Kilgarriff et al., 2014), and WordSmith Tools (Scott, 2016).

Table 4. Publication year, source and methodologies of reviewed studies

Variables	Values	Number of Studies	Percentage
Decade of publication	2015-2024	11	61.11%
	2004-2014	7	38.88%
Source of publication	Journal article	13	72.22%
	Book chapter	3	16.66%
	Conference proceedings	2	11.11%
Methodologies	Corpus analysis	18	100%

### 3.2 Contextual Features

#### 3.2.1 English Language Contexts

As shown in Figure 2, the English language contexts in related research over the past two decades can be divided into two main categories: native and non-native. For the native context, there are a total of eight studies (Biber et al., 2004; Chen, 2010; Gholaminejad, 2021; Hsu, 2014; Lee, 2020; Shirazizadeh & Amirfazlian, 2021; Wood, 2010; Wood & Appel, 2014), published across the years 2021, 2020, 2014, 2010, and 2004. Notably, five of these studies (62.5%) were published in 2014 or earlier. In the non-native context, there are ten studies with publication years ranging from 2014 to 2024. However, the majority of these studies are concentrated in the recent decade (2015-2024), with eight studies (80%) falling within this timeframe.

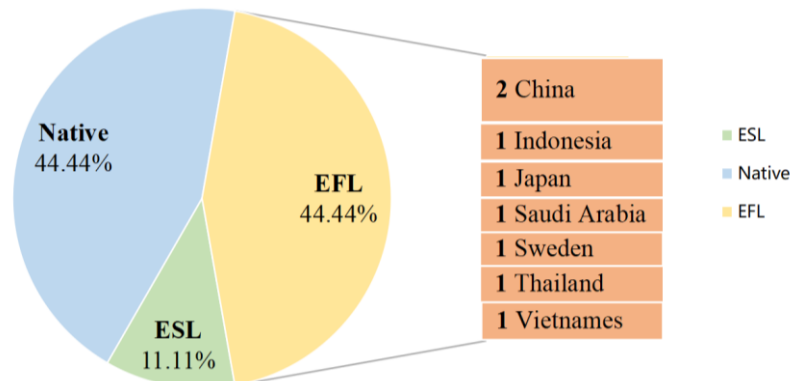


Figure 2. Proportional distribution of different English language contexts

The non-native context can be further categorized into English as a Second Language (ESL) and English as a Foreign Language (EFL). Among them, the number of EFL studies (eight) (Alasmary, 2022; Hoang & Crosthwaite, 2024; Miao, 2014; Northbrook & Conklin, 2018; Peng, 2016; Priyatno et al., 2023; Ribeck & Borin, 2014; Siricharoen & Wijitsopon, 2020) is significantly higher than that of ESL studies (two) (Hussain et al., 2021; Wang & Chan, 2024).

#### 3.2.2 Educational levels

Over the past two decades, the textbooks evaluated in relevant studies have predominantly targeted two educational levels: university (including graduate level) and secondary school (encompass junior and senior secondary school). As shown in Figure 3, the majority of research has focused on the university level (n=14), with only four studies examining textbooks used in secondary school (Hussain et al., 2021; Northbrook & Conklin, 2018; Priyatno et al., 2023; Ribeck & Borin, 2014).

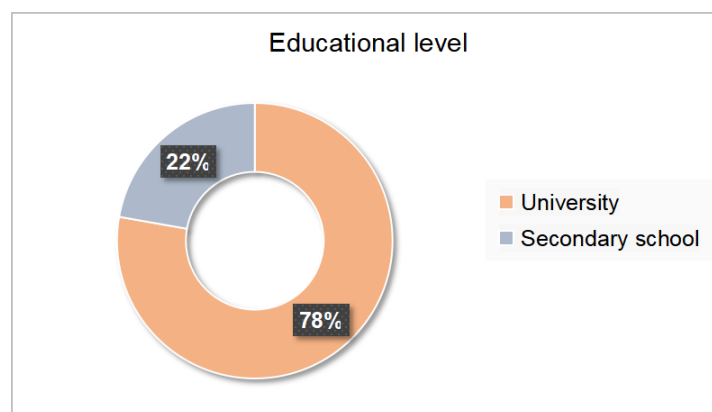


Figure 3. Proportional distribution of different educational levels

### 3.3 Types of textbook samples

The textbooks analyzed fall into three categories: English for General Purposes (EGP), English for Academic Purposes (EAP), and English for Specific Purposes (ESP). There are six studies focused on EGP, including two at the university level (Hoang & Crosthwaite, 2024; Miao, 2014), with the remaining four at the secondary level (Hussain et al., 2021; Northbrook & Conklin, 2018; Priyatno et al., 2023; Ribeck & Borin, 2014).

Eight studies analyze EAP textbooks across various disciplines, including English as a Lingua Franca (Wang & Chan, 2024), mathematics

(Alasmary, 2022), applied linguistics (Gholaminejad, 2021; Lee, 2020; Shirazizadeh & Amirfazlian, 2021), and the integration of multi-disciplines (Biber et al., 2004; Hsu, 2014; Wood, 2010).

Specialized fields in ESP textbooks include business English (Siricharoen & Wijitsopon, 2020), medical English (Peng, 2016), business and engineering (Wood & Appel, 2014), and electrical engineering (Chen, 2010). Details are provided in Table 5.

Table 5. Types of textbooks in reviewed studies

Types	Number of Studies	Percentage
EGP	6	33.3%
EAP	8	44.44%
ESP	4	22.22%

### 3.4 Length and Analytical Dimensions of Target Lexical Bundles

Lexical bundles are the combination of two or more words. The length of target lexical bundles varied in previous studies. Through summary (cf. Figure 4), the most frequently examined length is four words (n=8), while the remaining studies focus on three to six words, with the majority also addressing four-word bundles.

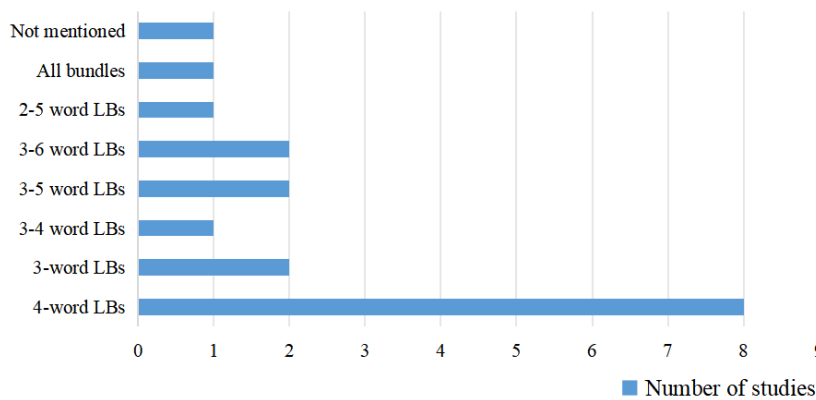


Figure 4. Length of target lexical bundles in reviewed studies

Except for Hsu (2014), which established and tested a pedagogically useful list of lexical bundles, the remaining seventeen studies conducted analyses of lexical bundles across various dimensions. Although these dimensions differ, most (16) of the remaining studies (17) all tend to involve the structure and function of lexical bundles, examining either function (n=9), structure (n=2), or focusing on both (n=5).

The functional classifications utilized in these studies include those proposed by Biber et al. (2004) and Hyland (2008) and the structural classifications refer to Biber et al. (1999) and Hyland (2008).

In addition to the aforementioned dimensions, previous studies have also considered other aspects, such as the frequency of lexical bundles. A detailed overview is provided in Table 6.

Table 6. Analytical dimensions of lexical bundles in reviewed studies

Analytical Dimensions	Number	Studies
Functions	9	Biber et al. (2004), Chen (2010), Gholaminejad (2021), Hoang and Crosthwaite (2024), Hussain et al. (2021), Miao (2014), Wang and Chan (2024), Wood (2010), Wood and Appel (2014)
Frequency/ Occurrence	7	Biber et al. (2004), Hoang and Crosthwaite (2024), Miao (2014), Peng (2016), Shirazizadeh and Amirfazlian (2021), Wood (2010), Wood and Appel (2014)
Functions & structures	5	Alasmary (2022), Lee (2020), Peng (2016), Shirazizadeh and Amirfazlian (2021), Siricharoen and Wijitsopon (2020)
Structures	2	Northbrook and Conklin (2018), Priyatno et al. (2023)
Distribution of Bundle Lengths	2	Northbrook and Conklin (2018), Priyatno et al. (2023)
Bundle density	1	Ribeck and Borin (2014)

## 4. Discussion

### 4.1 Research Trends in Previous Studies on Lexical Bundles in English Language Textbooks

#### 4.1.1 General Research Trends

Over the past two decades (2004-2024), a total of 18 relevant studies have been filtered. The increasing number of studies in this field underscores the growing interest in formulaic language and its significance in English language acquisition. The dominance of corpus

analysis reflects a methodological preference for empirical, data-driven approaches, enabling researchers to objectively extract and quantitatively analyze lexical bundles. Additionally, the high concentration of studies in EFL contexts—particularly in countries with large numbers of EFL learners such as China—demonstrates the challenge these learners face in achieving native-like fluency in English. The following sections provide a more in-depth analysis of these research trends.

#### 4.1.2 Methodological Trend

All selected studies (100%) have employed corpus analysis as the research methodology. This trend is largely attributed to the emergence of corpus linguistics, facilitated by advancements in computer technology from the late 1990s to the 2000s. Corpus linguistics involves the empirical analysis of language using large, electronically stored text collections known as corpora (McEnery & Brookes, 2024). It enables researchers to uncover language patterns from a fresh and insightful perspective (Hunston, 2022).

During the 1980s COBUILD dictionary project, Sinclair pioneered a corpus-based method to identify phraseological units such as lexical bundles. This method, which employs quantitative analysis to replace traditional subjective recognition and extraction techniques, is referred to as the “*frequency-driven approach*” (Biber et al., 2004). It analyzes lexico-grammatical patterns emerging from corpus data by using computational software. By revealing the collocational nature of vocabulary and grammar, it shifts the focus of phraseological studies away from fixed subcategories and incorporates a broader range of lexical combinations. To date, this corpus-based approach has become one of the core methods for identifying lexical bundles in the field. The extraction criteria include bundle length, cut-off frequency, dispersion and other metrics.

#### 4.1.3 Trend in English Language Contexts

On the one hand, whether from the perspectives of publication year or the number of publications (see Section 3.2.1), there is a clear trend towards studies on lexical bundles in non-native English contexts, particularly in EFL context (e.g., Hoang & Crosthwaite, 2024). This trend stems from the distinct learning environments of EFL and ESL. Drawing on his teaching experiences in both contexts, Krieger (2012) noted that ESL learners are multilingual and immersed in the culture of the target language, while EFL learners are typically monolingual and live in a non-target language environment, making it challenging for them to experience continuous target language exposure outside the classroom. Consequently, a specifically tailored teaching approach is necessary for these two groups of learners. This distinction also highlights the growing attention on EFL instruction, such as EFL textbooks that serve as primary sources of English language input (Yang & Coxhead 2022). Secondly, this difference is largely attributable to the varying populations of English learners. Globally, EFL learners are primarily concentrated in densely populated countries across Asia and South America, such as China, Russia, and Brazil. In contrast, ESL learners are more prevalent in countries like India, South Africa, and the Philippines, leading to a significantly higher number of EFL learners compared to ESL learners. Furthermore, given that proficiency in lexical bundles is a key indicator of English language competency among non-native English learners (Conklin & Schmitt 2012; Omidian et al., 2018; Priyatno et al., 2023), this trend underscores the greater challenges EFL learners face in English acquisition and emphasizes the necessity of lexical bundle instruction to support them in achieving native-like fluency.

On the other hand, among all EFL countries, China accounts for two studies (Miao, 2014; Peng, 2016), while the others have only one each. Considering the development of lexical bundle research in the Chinese academic field, studies in this area remained in their infancy before 2009. However, after 2009, empirical research saw a sharp increase, accompanied by a trend toward interdisciplinary integration, as studies increasingly intersected with fields such as translation, culture, and lexicography (Zhu & Pan, 2019). In a recent review of English lexical bundle research in China from 2013 to 2023, Wang and Du (2024) observed a notable rise in both the volume and diversity of related studies since 2013. It is precisely the increasing attention to this area in recent years that can be attributed to the growing interest among Chinese researchers in the topic of lexical bundles in English textbooks. Moreover, given that China has the largest number of EFL learners worldwide, the evaluation of lexical bundles in Chinese EFL textbooks is particularly significant, offering valuable insights into non-native English contexts that should not be overlooked.

#### 4.1.4 Length of Target Lexical Bundles

The length of target lexical bundles ranges from two words to longer multi-word sequences, with the majority (83.33%) focusing on four-word bundles. This suggests that four-word bundles are the most commonly analyzed length in previous research. This preference can be attributed to the high frequency of two-word lexical bundles, which often appear embedded within longer sequences of three or more words (Simpson-Vlach & Ellis 2010) with functions that are challenging to identify (Allan, 2017; Biber et al., 2004). Similarly, three-word lexical bundles are also prevalent (Biber et al., 2004), making the exclusion of both two- and three-word bundles a practical choice for maintaining data manageability. In contrast, five- and six-word bundles are relatively rare and typically consist of extensions of four-word bundles (Biber et al., 1999; Biber et al., 2004). As a result, researchers have suggested that four-word bundles represent the optimal length for lexical bundle analysis (Biber et al., 1999), as they occur more frequently than longer bundles and more clearly capture the structural and functional characteristics of lexical bundles compared to two- or three-word sequences (Cortes, 2004; Hyland, 2008).

However, some researchers have emphasized the significant pedagogical value of three-word lexical bundles. For instance, Simpson-Vlach and Ellis (2010), in their development of the Academic Formulas List (AFL), observed that the 50 most frequent lexical bundles consisted of three words. Similarly, Allan (2017) noted that three-word lexical bundles are particularly well-suited for analysis in English self-study materials, where limited text length constrains the occurrence of four-word bundles.

Therefore, to ensure a comprehensive analysis of lexical bundles in English textbooks, it is recommended to examine both three- and four-word bundles (i.e.,  $3 \leq N \leq 4$ ), with flexibility in adjustment based on sample size.

#### 4.1.5 Evaluative Dimensions

A review of lexical bundle research reveals that (cf. Table 6), after extracting lexical bundles, conducting an in-depth examination of their structure and function has become a well-established tradition in the field (e.g., Biber et al., 1999; Biber et al., 2004; Biber & Barbieri, 2007 etc.). In studies analyzing lexical bundles in English language textbooks, the majority of researchers (16/18) have followed this tradition, reflecting a trend toward integrating form and function in bundle analysis. Additionally, 38.88% (7/18) of studies emphasize bundle frequency or occurrence as an evaluative dimension. This is because frequency is a prerequisite for identifying lexical bundles in corpus linguistics research (Hejazi, 2022), as their recognition hinges on the frequent co-occurrence of lexical units within a specific corpus (Biber et al., 1999).

Thus, it is indicated that researchers should base their analysis on bundle frequency, in line with the established tradition of examining lexical bundles in terms of both structure and function. At the same time, researchers should flexibly select multiple dimensions for comprehensive analysis, according to their research objectives and the characteristics of the target corpus. This multidimensional approach (e.g., Shirazizadeh & Amirfazlian, 2021) enhances the accuracy, comprehensiveness, and practical value of the findings, providing deeper insights into the usage patterns and functional roles of lexical bundles.

#### 4.2 Research Gap in Previous Studies on Lexical Bundles in English Language Textbooks

Research on lexical bundles in EL textbooks has predominantly concentrated on tertiary-level (78%), particularly those designed for English for Academic Purposes (EAP) (e.g., Alasmary, 2022). In contrast, studies on tertiary-level English for General Purposes (EGP) textbooks (e.g., Hoang & Crosthwaite, 2024) and English for Specific Purposes (ESP) textbooks (e.g., Siricharoen & Wijitsopon, 2020) remain limited. Furthermore, research on foundational-level English for General Purposes (EGP) textbooks is scarce (only 22%), with limited studies, such as Priyatno et al. (2023), focusing on the senior secondary school level. This imbalance underscores a significant research gap.

Several structural and contextual factors may account for this imbalance. Firstly, the demand for lexical bundles is arguably more pronounced in academic contexts at the tertiary level, where mastery of specialized academic English is essential for disciplinary literacy (Mok & Jain 2023). Lexical bundles serve as key linguistic features that distinguish specific registers (Hyland & Jiang, 2018). In academic English discourse, they frequently occur and reflect the linguistic norms shared by members of the academic community (Conrad & Biber 2005; Hyland, 2008). As such, they serve as essential scaffolds for EAP learners navigating complex academic texts (Hussain et al., 2021) and facilitates their integration into a specific academic discourse community (Wray, 2013). In contrast, secondary-level education primarily emphasizes foundational English language skills, leading to a relatively lower academic demand for lexical bundles. Secondly, research resources are concentrated at the tertiary level. Most academic researchers are affiliated with universities, where funding and publication opportunities for studies on tertiary-level textbooks are more readily available (Hui, 2020). Meanwhile, secondary-level teachers face heavy workloads, limiting their time for conducting research.

Studies indicate that both native and non-native English learners are sensitive to frequency effects (Siyanova et al., 2011). Frequency also significantly influences the acquisition of lexical knowledge, including both individual words and phraseological units such as lexical bundles (Durrant & Schmitt, 2010). Given their frequent occurrence, lexical bundles are particularly beneficial for learners at beginner to intermediate proficiency levels. Northbrook and Conklin (2019) found that beginner-level secondary school EFL learners process lexical bundles in their textbooks with greater speed and accuracy. At the secondary and lower educational levels, foundational language instruction prepares learners for advanced proficiency and future academic studies. Therefore, greater emphasis should be placed on the selection and presentation of lexical bundles in textbooks at these early educational stages.

In addition, ESP courses cater to the specific language demands of various fields and industries. To effectively support students' acquisition of specialized English, more attention should also be given to the use of lexical bundles in ESP textbooks, such as in the field of Traditional Chinese Medicine.

### 5. Conclusion

This systematic literature review examines research on lexical bundles in English language textbooks from 2004 to 2024 across three databases: Web of Science Core Collection, Scopus, and ERIC. Studies are analyzed based on methodology, context, textbook type, and the characteristics of target lexical bundles, including their length and analytical dimensions. The detailed cross-study comparisons based on various variables identify current trends and research gaps, providing insights and directions for future studies.

#### 5.1 Recommendations

Based on this review, we propose the following recommendations for future studies on lexical bundles in EL textbooks:

Over the past two decades (2001-2020), corpus-assisted research on lexical bundles has shown sustained popularity within corpus linguistics (Crosthwaite et al., 2023). As corpus analysis remains the most widely utilized and effective methodology for investigating lexical bundles (Peng & Wang, 2021), we recommend its continued application for exploring lexical bundles in EL textbooks. Researchers should focus on both three- and four-word bundles, with flexibility based on sample size. We also encourage the combined



use of multiple corpus concordance software to enhance the reliability of lexical bundle extraction and data analysis.

The lack of consensus on extraction criteria—such as bundle length, cut-off frequency, dispersion, and other statistical metrics—has led to inconsistencies and reduced methodological rigour in lexical bundle research (Ädel & Erman, 2012; Yan, 2021). Addressing these discrepancies presents a crucial area for future exploration.

We see significant potential in placing greater emphasis on the evaluation of lexical bundles, particularly in non-native English language textbooks, such as EFL textbooks. Given the greater challenges EFL learners face in English acquisition compared to ESL learners, this focus is especially crucial. Additionally, foundational-level English for General Purposes (EGP) textbooks (e.g., secondary school EGP textbooks) and English for Specific Purposes (ESP) textbooks have received insufficient attention and warrant further exploration.

Future studies should adopt multiple analytical dimensions to comprehensively evaluate lexical bundles in EL textbooks, conducting an in-depth examination of their structures and functions, as proposed by earlier pioneers (e.g., Biber et al., 1999).

### 5.2 Limitations

This review prioritizes rigorous research and high-quality publications, relying on studies from only three databases. As a result, it may not fully capture all relevant literature. Future studies could broaden the range of databases to achieve a more comprehensive and in-depth understanding of lexical bundle research in EL textbooks.

### Acknowledgments

We would like to express our appreciation for the valuable contributions of all our team members in completing this study.

### Authors' contributions

Mingxuan Zhang was responsible for data collection, interpretation, and the initial drafting of the manuscript. Prof. Jayakaran Mukundan, Dr. Laleh Khojasteh, and Dr. Jasmine Anak Jain were responsible for reviewing and supervising the manuscript. All authors read and approved the final version.

### Funding

Not applicable.

### Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Informed consent

Obtained.

### Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

### Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

### Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### Data sharing statement

No additional data are available.

### Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

### References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Alasmary, A. (2022). Academic lexical bundles in graduate-level math texts: A corpus-based expert-approved list. *Language Teaching Research*, 26(1), 99-123. <https://doi.org/10.1177/1362168819877306>
- Alasmary, A. (2024). Sustaining vocabulary knowledge growth through corpus-generated lists of lexical bundles and keywords in the law

- of contracts. *Heliyon*, 10(9). <https://doi.org/10.1016/j.heliyon.2024.e29944>
- Allan, R. (2017). From do you know to I don't know: An analysis of the frequency and usefulness of lexical bundles in five English language self-study books. *Corpus Pragmatics*, 1(4), 351-372. <https://doi.org/10.1007/s41701-017-0016-9>
- Altenberg, B. (1990). Speech as linear composition. In *Proceedings from the fourth Nordic conference for English studies* (Vol. 1, pp. 133-143).
- Anthony, L. (2018). *AntConc* (version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Appel, R., Geluso, J., & Feng, H. H. (2024). An examination of phrase-frames in L2 english academic writing: Exploring relationships with writing quality. *System*, 123, 103349. <https://doi.org/10.1016/j.system.2024.103349>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83-110. <https://doi.org/10.1017/S0267190512000050>
- Bolinger, D. (1979). Meaning and memory. In G. G. Haydu (Ed.), *Experience forms: Their cultural and individual place and function* (pp. 95-111). Berlin, Germany: Walter de Gruyter. <https://doi.org/10.1515/9783110815733>
- Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In A. Wray (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 107-125).
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61. <https://doi.org/10.1017/S0267190512000074>
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423. <https://doi.org/10.1016/j.esp.2003.12.001>
- Crosthwaite, P., Ningrum, S., & Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research in arts and humanities. *International Journal of Corpus Linguistics*, 28(3), 344-377. <https://doi.org/10.1075/ijcl.21072.cro>
- Ding, S. Y., & Chen, J. (2019). Formulaic language in second language teaching—A case study of situation-bound utterances in China's EFL textbooks. In *Proceedings of the 4th International Conference on Education Science and Development (ICESD 2019)*. <https://doi.org/10.12783/dtssehs/icesd2019/28082>
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188. <https://doi.org/10.1177/0267658309349431>
- Esfandiari, R., & Barbary, F. (2023). A corpus-driven, diachronic analysis of recurrent word combinations across academic disciplines. *Corpora*, 18(3), 263-295. <https://doi.org/10.3366/cor.2023.0288>
- Gholaminejad, R. (2021). A comparison of two genres: Lexical bundles in the discourse of applied linguistics. *Atlantis: Journal of the Spanish Association for Anglo-American Studies*, 43(2), 90-109. <https://doi.org/10.28914/Atlantis-2021-43.2.05>
- Gough, D., Thomas, J., & Oliver, S. (2017). *An introduction to systematic reviews* (2nd ed.). SAGE Publications Ltd. <https://doi.org/10.53841/bpsptr.2017.23.2.95>
- Hejazi, H. H. (2022). *A corpus-based investigation of lexical bundles and keyness in B1, B2 and C1 ESL learners' academic writing* (Doctoral dissertation). University of Liverpool.
- Hoang, H., & Crosthwaite, P. (2024). A comparative analysis of multiword units in the reading and listening input of English textbooks. *System*, 121, 103224. <https://doi.org/10.1016/j.system.2024.103224>
- Hsu, J. Y. (2006). An Analysis of the Multiword Lexical Units in Contemporary ELT Textbooks. *Online Submission*. Retrieved from <https://eric.ed.gov/?id=ED497440>
- Hsu, W. (2014). The most frequent opaque formulaic sequences in English-medium college textbooks. *System*, 47, 146-161. <https://doi.org/10.1016/j.system.2014.10.001>
- Hui, Y. (2020). A twenty-year study of English textbooks at home and abroad. *Teaching Research*, 43(4), 38-49.
- Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781108616218>
- Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse Functions of Lexical Bundles in Pakistani Chemistry and Physics Textbooks.

- GEMA Online Journal of Language Studies*, 21(1). <https://doi.org/10.17576/gema-2021-2101-13>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Jiang, F. (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>
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), 118-121. <https://doi.org/10.1177/014107680309600304>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59(4), 322-332. <https://doi.org/10.1093/elt/cci061>
- Krieger, D. (2012). Teaching ESL versus EFL: Principles and practices. *English Teaching Forum*, 43(2), 9. Retrieved from <https://silviaspence.tripod.com/sitebuildercontent/sitebuilderfiles/teflarticle.htm>
- Lee, H. K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research*, 37(1). <https://doi.org/10.17250/khisli.37.1.202003.005>
- Litz, D. R. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal*, 48(1), 1-53. Retrieved from [https://www.researchgate.net/publication/228383638\\_Textbook\\_evaluation\\_and\\_ELT\\_management\\_a\\_South\\_Korean\\_case\\_study](https://www.researchgate.net/publication/228383638_Textbook_evaluation_and_ELT_management_a_South_Korean_case_study)
- Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30, 66-81. <https://doi.org/10.1016/j.jslw.2015.08.011>
- Ma, X. M., & Meng, H. R. (2021). A study on lexical chunks in dialogue sections of primary school English textbooks. *English Learning*, 4, 48-53.
- McAleese, P. (2013). Investigating multi-word items in a contemporary ELT course book. In *JALT2012 Conference Proceedings*, Tokyo, Japan.
- McEnery, T., & Brookes, G. (2024). Corpus linguistics and the social sciences. *Corpus Linguistics and Linguistic Theory*. Advance online publication. <https://doi.org/10.1515/cllt-2024-0036>
- Miao, H. (2014). An Investigation of Formulaic Sequences in Multi-modal Chinese College English Textbooks. *Journal of Language Teaching & Research*, 5(6). <https://doi.org/10.4304/jltr.5.6.1308-1314>
- Mok, S. J., & Jain, J. (2023). Malaysian In-Service Science Teachers' Conception of Nature of Science (NOS). In M. Koc, O. T. Ozturk & M. L. Ciddi (Eds.), *Proceedings of ICRES 2023-- International Conference on Research in Education and Science (pp. 1049-1062)*, Cappadocia, Turkiye. ISTES Organization.
- Mukundan, J., & Kalajahi, S. A. R. (2013). Evaluation of Malaysian English language teaching textbooks. *International Journal of Education and Literacy Studies*, 1(1), 38-46. <https://doi.org/10.7575/aiac.ijels.v.1n.1p.38>
- Nguyen, M. T. T. (2011). Learning to communicate in a globalized world: To what extent do school textbooks facilitate the development of intercultural pragmatic competence? *RELC Journal*, 42, 17-30. <https://doi.org/10.1177/0033688210390265>
- Nguyen, T. M. H., & Coxhead, A. (2023). Evaluating multiword unit word lists for academic purposes. *ITL-International Journal of Applied Linguistics*, 174(1), 83-111. <https://doi.org/10.1075/itl.21041.ngu>
- 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(3), 311-334. <https://doi.org/10.1075/ijcl.16024.nor>
- Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out?—Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, 40(5), 816-833. <https://doi.org/10.1093/applin/amy027>
- Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for Academic Purposes*, 36, 1-14. <https://doi.org/10.1016/j.jeap.2018.08.002>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021a). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <http://dx.doi.org/10.1136/bmj.n160>
- Pearson, W. S., & Abdollahzadeh, E. (2023). Metadiscourse in academic writing: A systematic review. *Lingua*, 293, 103561. <https://doi.org/10.1016/j.lingua.2023.103561>
- Peng, Y. M. (2016). A comparative study on lexical chunks in medical English textbooks and general college English textbooks.

- Educational Modernization*, (33), 235-237. <https://doi.org/10.16541/j.cnki.2095-8420.2016.33.084>
- Peng, Y. M., & Wang, Y. (2021). A corpus-based review of studies on English lexical chunks. *English Square: Academic Research*, (19), 47-49. <https://doi.org/10.16723/j.cnki.yygc.2021.19.014>.
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61(2), 325-366. <https://doi.org/10.1111/j.1467-9922.2011.00640.x>
- Priyatno, A., Dinda, O. Y., Nugraheni, W., & Utami, W. (2023). Lexical bundles in Indonesian EFL textbooks: A corpus analysis. *Journal of Language and Education*, 9(2 (34)), 25-39. <https://doi.org/10.17323/jle.2023.16305>
- Prowse, P. (2011). How writers write: Testimony from authors. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 151-173). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781139042789.009>
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781139042789.003>
- Ribeck, J., & Borin, L. (2014). Lexical bundles in Swedish secondary school textbooks. In *Human Language Technology Challenges for Computer Science and Linguistics: 5th Language and Technology Conference, LTC 2011, Poznań, Poland, November 25-27, 2011, revised selected papers 5* (pp. 238-249). Springer International Publishing. [https://doi.org/10.1007/978-3-319-08958-4\\_20](https://doi.org/10.1007/978-3-319-08958-4_20)
- Scott, M. (2016). *WordSmith Tools* (version 7.0). [Computer software]. Liverpool, UK: Lexical Analysis Software. Retrieved from <http://www.lexically.net>
- Sheldon, L. E. (1988). Evaluating ELT Textbooks and Materials. *ELT Journal*, 42(4), 237-246. <https://doi.org/10.1093/elt/42.4.237>
- Shirazizadeh, M., & Amirfazlian, R. (2021). Lexical bundles in theses, articles and textbooks of applied linguistics: Investigating intradisciplinary uniformity and variation. *Journal of English for Academic Purposes*, 49, 100946. <https://doi.org/10.1016/j.jeap.2020.100946>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>
- Siricharoen, A., & Wijitsopon, R. (2020). A Corpus-Based Comparative Study of Lexical Bundles in Authentic and Textbook English Business Emails. *LEARN Journal: Language Education and Acquisition Research Network*, 13(2), 41-63. Retrieved from <https://eric.ed.gov/?id=EJ1258791>
- Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776. <https://doi.org/10.1037/a0022531>
- Tomlinson, B. (2012a). Materials development. In J. C. Richards & A. Burns (Eds.), *The Cambridge guide to pedagogy and practice in second language teaching* (pp. 269-278). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781009024778.034>
- Tomlinson, B. (2012b). Materials development for language learning and teaching. *Language Teaching*, 45(2), 143-179. <https://doi.org/10.1017/S0261444811000528>
- Tomlinson, B. (Ed.) (2008). *English Language Learning Materials: A Critical Review*. London, England: Continuum.
- Tomlinson, B. (Ed.). (2011). *Materials development in language teaching* (2nd ed.). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781139042789>
- Wang, Q., & Du, H. (2024). A review of lexical bundle research in China over the past decade. *Modern English*, 10(5), 64-66.
- Wang, Y., & Chan, N. C. L. (2024). *Formulaic language in university seminars: A comparison of EAP textbook coverage and authentic language use in ELF settings*. Bloomsbury Collections. <https://doi.org/10.5040/9781350300330.0011>
- Wood, D. (2010). Lexical clusters in an EAP textbook corpus. *Perspectives on formulaic language: Acquisition and communication*, 88, 88-106.
- Wood, D. (2015). *Fundamentals of formulaic language*. Bloomsbury Academic, an imprint of Bloomsbury Publishing Plc.
- Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, 15, 1-13. <https://doi.org/10.1016/j.jeap.2014.03.002>
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46(3), 316-334. <https://doi.org/10.1017/S0261444813000013>
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28. [https://doi.org/10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4)
- Xu, Q., Dong, X. Q., & Yuan, Y. (2022). The impact of task motivation on Chinese English learners' audiovisual retelling output. *Modern Foreign Languages*, 5, 645-658.
- Xu, Y. (2016). Contrastive analysis of lexical chunks in computer English textbooks and College English textbooks. In *2016 5th International Conference on Social Science, Education and Humanities Research (SSEHR 2016)* (pp. 593-597). Paris, France:

Atlantis Press. <https://doi.org/10.2991/ssehr-16.2016.127>

Yan, Y. (2021). The impact of extraction criteria and statistical methods on second language lexical bundle research. *Shandong Foreign Language Teaching*, 42(2), 55-66.

Yang, L., & Coxhead, A. (2022). A corpus-based study of vocabulary in the new concept English textbook series. *RELC Journal*, 53(3), 597-611. <https://doi.org/10.1177/0033688220964162>

Zhu, L., & Pan, L. (2019). A decade of research on lexical bundles (2009-2018): *Review and prospect. Education Modernization*, 4(104), 265-268.

## Appendix A

Author(s)	Title	Year	Source	Context	Research aims	Textbooks Corpus	Method	Length of LBs	Analytical dimensions of LBs	Main findings
Wang, Y.	Formulaic Language in University Seminars: A Comparison of EAP Textbook Coverage and Authentic Language Use in ELF Settings	2024	Book chapter	ESL	Evaluate formulaic sequences (FSs) in EAP textbooks against authentic language use (university seminars) and consider potential implications for the teaching of EAP	Seminar discussion sections in ten EAP textbooks	Corpus Analysis	3-word clusters	Compare the speech functions in the textbooks and seminar	EAP textbooks present a limited range of formulaic sequences (FSs)
Hoang, H.	A comparative analysis of multiword units in the reading and listening input of English textbooks	2024	Journal article	EFL	Compare the multiword units (MWUs) between reading and listening input of English textbooks	EFL textbooks are commonly used by English majors at a Vietnamese university	Corpus Analysis	3-5 word MWUs	Occurrence & functional classifications	MWUs are more frequent in oral input, while reading input has more MWU types and a greater variety but with lower recurrence
Priyatno, A.	Lexical Bundles in Indonesian EFL Textbooks: A Corpus Analysis	2023	Journal article	EFL	Examines the patterns and structural classifications of LBs in Indonesian senior high school EFL textbooks	Five Indonesian senior secondary school English language teaching textbooks	Corpus Analysis	3-6 word LBs	LB length and structures	Three-word LBs are and prepositional phrase bundles are the most frequent
Alasmarty, A.	Academic Lexical Bundles in Graduate-Level Math Texts: A Corpus-Based Expert-Approved List	2022	Journal article	EFL	Synthesize and analyse the most frequently occurring, widely dispersed and pedagogically useful lexical bundles in mathematical texts	Thirty-six advanced graduate-level mathematics textbooks & Seminar discussion sections in ten EAP textbooks	Corpus Analysis	4-word LBs	functional and structural analyses	The dominant structures of these LBs are clausal, and their primary function is research-oriented
Shirazizadeh, M.	Lexical bundles in theses, articles and textbooks of applied linguistics:	2021	Journal article	Native	Investigated the intra-disciplinary variations of LBs across three genres within applied linguistics	Thirty applied linguistics English textbooks.	Corpus Analysis	4-word LBs	Overall frequency information and variation in structures and functions	LBs vary across genres within the same discipline but still share significant commonalities

	Investigating intradisciplinary uniformity and variation										
Hussain, G.	Discourse functions of lexical bundles in Pakistani chemistry and physics textbooks	2021	Journal article	ESL	Identify four-word common core lexical bundles occurring in textbooks and classify the discourse functions	Four Pakistani higher secondary school certificates (HSSC) at intermediate-level textbooks of chemistry and physics	Corpus Analysis	4-word LBs	Functional taxonomies.	The selected textbooks contain 102 LBs, with discourse organizers and referential expressions being the most dominant functional categories	
Gholamin ejad, R.	A Comparison of Two Genres: Lexical Bundles in the Discourse of Applied Linguistics	2021	Journal article	Native	Identify the lexical bundles used in the two genres (textbooks and research articles) in applied linguistics and compare their functions	one hundred and thirty-nine textbooks in applied linguistics	Corpus analysis	4-word LBs	Functions	The occurrence of LBs in textbooks is lower than in research articles, and the discourse functions in both genres exhibit distinct characteristics	
Lee, H.	Lexical bundles in linguistics textbooks	2020	Journal article	Native	To develop a lexical bundle lists in linguistics and assist linguistics students enhancing their reading proficiency	Five linguistics textbooks	Corpus Analysis	3-4 word LBs	Structures and functions	NP-based and PP-based bundles and referential bundles are dominate	
Siricharoen, A.	A Corpus-Based Comparative Study of Lexical Bundles in Authentic and Textbook English Business Emails	2020	Journal article	EFL	Identify and compare the types of lexical bundles found in authentic English business emails and sample emails from business English textbooks	Business email samples taken from seventy-seven business English textbooks	Corpus Analysis	4-word LBs	Structural and functional categories.	Business emails in textbooks are more formal than those in authentic emails and bundle functions differ significantly	
Northbrook, J.	“What are you talking about?”: An analysis of lexical bundles in Japanese junior high school textbooks	2018	Journal article	EFL	Whether the LBs in communicative Japanese junior high school textbooks are representative of conversational English	Eighteen Japanese junior high school English textbooks	Corpus Analysis	3-6 words LBs	Distribution of LBs by length, overlap with the same-length LBs in the reference corpus and lexico-grammatical constructions	Textbooks fail to reflect authentic language use and lack genuine samples	
Peng, Y.	A Comparative Study on Lexical Chunks in Textbooks of Medical English and College	2016	Conference proceedings	EFL	Compares the LBs in textbooks of medical English and college English	Medical English and college English textbooks from different publishers	Corpus Analysis	Not mentioned	Number, structures and functions	Suggestions are provided for improving bundle usage in medical English textbooks	

Hsu, W.	English The most frequent opaque formulaic sequences in English-medium college textbooks	2014	Journal article	Native	Establish a pedagogically useful list of the most frequent semantically non-transparent formulaic sequences for non-English majors in an EFL context	Two hundred college textbooks across forty subject areas	Corpus Analysis	2-5 words LBs	Without the dimension of bundle analysis (presents the process of constructing a bundle list)	A total of 475 opaque bundles are compiled into a bundle list
Miao, H.	An investigation of formulaic sequences in multi-modal Chinese college English textbooks	2014	Journal article	EFL	Explore formulaic sequences (FSs) in EGP textbooks to see whether there are register differences in different modal Chinese EFL textbooks	Integrated textbooks & listening-speaking textbooks	Corpus Analysis	3-word LBs	Frequency, overlap structures and functions	Prepositional and noun-based FSs are more common in integrated textbooks, while verb-based formulaic sequences appear more frequently in the listening-speaking textbooks
Ribeck, J.	Lexical Bundles in Swedish Secondary School Textbooks	2014	Conference proceedings	EFL	Proposes a novel approach to identifying LBs and describes the characteristics of these bundles	Swedish corpus of secondary textbooks (two history books and two physics books)	Corpus Analysis	all bundles	Bundle density, variation, distributions of structures and functions.	History uses more NP/PP-based and less dependent-clause-based bundles than physics
Wood, D.	Multiword constructions in first-year business and engineering university textbooks and EAP textbooks	2014	Journal article	Native	To what extent the extracted LBs in ESP textbooks are present in their EAP counterparts	Ten first-year business and engineering textbooks and five EAP books for intermediate and advanced learners	Corpus Analysis	3-5 word LBs	Occurrence rate & functions	LBs identified in university textbook books are under-represented in intermediate/advanced EAP materials
Wood, D.	Lexical Clusters in an EAP Textbook Corpus	2010	Book chapter	Native	To assess the exposure EAP learners receive to LBs in academic prose from typical EAP textbooks in North America, and to determine whether these materials promote the noticing, processing, and use of LBs relevant to academic reading and writing	Six intermediate-advanced level textbooks (four EAP textbooks and two reading and writing skills-focused textbooks)	Corpus Analysis	4-word LBs	Frequency and functions	The highest frequency LBs are found in instructional materials, not the texts and Textbooks lack exercises, reference materials, or activities focused on LBs, either explicitly or implicitly
Chen, L.	An Investigation of Lexical Bundles in ESP	2010	Book chapter	Native	How lexical bundles with consistent pragmatic functions	Two electrical engineering preparatory textbooks and two English for	Corpus Analysis	4-word LBs	Comparison of functions	LBs in ESP textbooks are inadequate in both occurrence and function, covering

---

	Textbooks and Electrical Engineering Introductory Textbooks				construct EE texts and whether these bundles and their functions are present in ESP textbooks	specific purposes textbooks				only about one-third of the LBs identified in EE preparatory textbooks
Biber, D.	If you look at ...: Lexical bundles in university teaching and textbooks	2004	Journal article	Native	Investigate the use of LBs in two university se registers: classroom teaching and textbooks	Classroom teaching and textbook texts from four universities in the USA across six disciplines and three undergraduate levels (lower, middle, and graduate)	Corpus Analysis	4-word LBs	Number, frequency, distribution of structural patterns, and functions across different registers	Present a functional taxonomy to classify bundle discourse functions. LBs in classroom teaching are around four times as many as textbooks

---