Article

https://doi.org/10.52696/YWRO3156
Reprints and permission:
The Malaysian English Language Teaching Association
✉ Mingxuan Zhang 0365972@sd.taylors.edu.my

Examining EFL Textbooks from a Phraseological Perspective: A Corpus Analysis of Lexical Bundle Features

Mingxuan Zhang
School of Education,
Faculty of Social Sciences and Leisure Management,
Taylor's University,
Selangor, Malaysia

Jayakaran Mukundan
UNITAR International University,
Selangor, Malaysia

Chong Liu
School of Arts,
Weihai Vocational College,
Weihai, China

Laleh Khojasteh
Department of English Language,
School of Paramedical Sciences,
Shiraz University of Medical Sciences,
Shiraz, Iran

Jasmine Anak Jain
School of Education,
Faculty of Social Sciences and Leisure Management,
Taylor's University,
Selangor, Malaysia

ABSTRACT

English language (EL) textbooks are the main or sole source of formal English input in non-native contexts. Their quality significantly influences teaching and learning. As recurrent continuous sequences, lexical bundles (LBs) are important phraseological units in the English language. They support language fluency and accuracy by reducing memory load and enhancing processing, particularly for non-native learners. Despite growing interest in LBs within English language education, research on their use in foundational-level English for General Purposes

(EGP) textbooks has remained insufficient. Given the importance of EL textbooks in non-native contexts and China's large EFL learner population, investigating LBs in Chinese EFL textbooks holds significant national and global value. Accordingly, this study examines the most frequent three- and four-word LBs in widely used Chinese secondary school EFL textbooks. Guided by corpus linguistics, two textbook corpora were compiled. Bundle data were extracted using WordSmith Tools and Collocate, and analyzed quantitatively and qualitatively for frequency, structure, and function. Findings show that three-word bundles are more frequent, with stable distribution across grade levels, while bundle diversity increases in advanced-level textbooks. Structurally, verb phrase fragments dominate, reflecting a phrasal style that lags behind professional native academic texts. Functionally, LBs with a limited functional range and pedagogical instruction bundles are most frequent, highlighting the teacher-centered nature of Chinese education. This study enriches the understanding of bundle features in Chinese EFL textbooks, offering insights for enhancing bundle instruction and optimizing textbook bundle presentation as well as contributing to EL textbook bundle research in non-native contexts.

KEYWORDS: EFL Textbooks, Phraseological Perspective, Corpus Analysis, Lexical Bundle Features

INTRODUCTION

Language learning and teaching materials encompass anything that supports language acquisition. Among these, textbooks, as printed resources, play a vital role in the language learning process (Tomlinson, 2012), particularly in non-native language contexts, where they act as a central link in managing the learning process and serve as effective tools for optimizing the teaching of non-native languages in school settings (Shamov & Shimichev, 2024).

In English language acquisition, English language (EL) textbooks serve as a fundamental component within the English language ecology (Zhang et al., 2025b). Their significance is also particularly pronounced in non-native English contexts, where EL textbooks often provide the most direct and sometimes the only source of formal English language input (Yang & Coxhead, 2022). Therefore, the quality of non-native EL textbooks is critically linked to the overall effectiveness of English language teaching and learning (Zhang et al., 2025a).

Among various phraseological units, lexical bundles (abbreviated as LBs) refer to recurrent continuous sequences of two or more words found in discourse, such as *"as well as"* and *"and this is a"*. In the English language, phrases, rather than individual words, primarily convey meaning. As a phraseological unit, LBs hold considerable importance in English language acquisition and instruction and should be given priority in English language teaching and learning (Zhang et al., 2025b).

Mastering these bundles provides learners with a distinct processing advantage, as they function similarly to a *"cache"* akin to how a computer stores frequently used data to enhance efficiency. LBs leave lasting impressions on memory, reduce cognitive load, and speed up recognition (Siyanova et al., 2011). Moreover, they boost overall English proficiency (Liu et al., 2022). Therefore, numerous scholars have concentrated on LBs in language teaching and learning. In recent years (2020-2024), researchers have extensively explored LBs using English learner corpora (e.g., Joharry, 2021), examined their processing contributions in language acquisition and instruction (e.g., Appel et al., 2024), and compiled bundle lists to supplement traditional word lists (e.g., Alasmary, 2022).

To effectively emphasize LBs in English language teaching and learning, EL textbooks should serve as the starting point. The use of EL textbooks that overlook or inadequately address LBs may hinder learners from understanding their proper usage across different registers and from recognizing their functional roles (Hoang & Crosthwaite, 2024). Therefore, researchers should place greater emphasis on LBs in EL textbooks.

However, a systematic review of LBs in EL textbooks over the past two decades (2004-2024) reveals insufficient attention in this realm, especially in foundational-level English for General Purposes (EGP) textbooks (Zhang et al., 2025b). This finding highlights a significant research gap, emphasizing the need for further exploration. Specifically, within English language contexts, there is a growing trend toward examining non-native English settings, with a particular focus on EFL textbooks (Zhang et al., 2025b). Considering that China has the largest number of EFL learners worldwide, evaluating LBs in Chinese EFL textbooks holds substantial global value in non-native English contexts. Furthermore, a review of the literature in China reveals a gap in the comprehensive analysis of bundle features in Chinese secondary school EFL textbooks, prompting the present study to select these textbooks as research samples.

The corpus approach is currently one of the most widely utilized and effective methodologies for investigating LBs (Peng & Wang, 2021). Frequency is the prerequisite for identifying lexical bundles within a corpus (Hejazi, 2022). After the frequency-based bundle extraction, it has become a well-established tradition to conduct a detailed analysis of their structure and function (e.g., Biber et al., 1999). Consequently, research dimensions typically involve bundle co-occurrence/frequency, bundle structure, and bundle function (Zhang et al., 2025b).

In light of this, the present study selects Chinese secondary school EFL textbooks as textbook samples, aiming to extract LBs from these textbooks and analyse their structural and functional characteristics using corpus concordance software. Accordingly, the study seeks to address the following three research questions:

(1) What lexical bundles are identified in Chinese secondary school EFL textbooks in terms of frequency and length?

(2) What structural patterns characterize these lexical bundles?

(3) What functional classifications are assigned to these lexical bundles?

LITERATURE REVIEW

Theoretical and Methodological Review from a Phraseological Perspective

Theoretical developments in phraseology have evolved from Structuralism through Universal Grammar to Construction Grammar. Alongside the emergence of these theories, phraseology has followed a trajectory from its origin to marginalization and subsequently to a renewed central position within linguistic research.

With the rise of Structuralism, European linguists emphasized the significance of phraseological units in language, viewing them as structural patterns or lexical combinations. In the 1940s, a comprehensive framework for phraseology emerged in the Soviet Union, where Vinogradov (1947) argued that phrases convey specific meanings and functions.

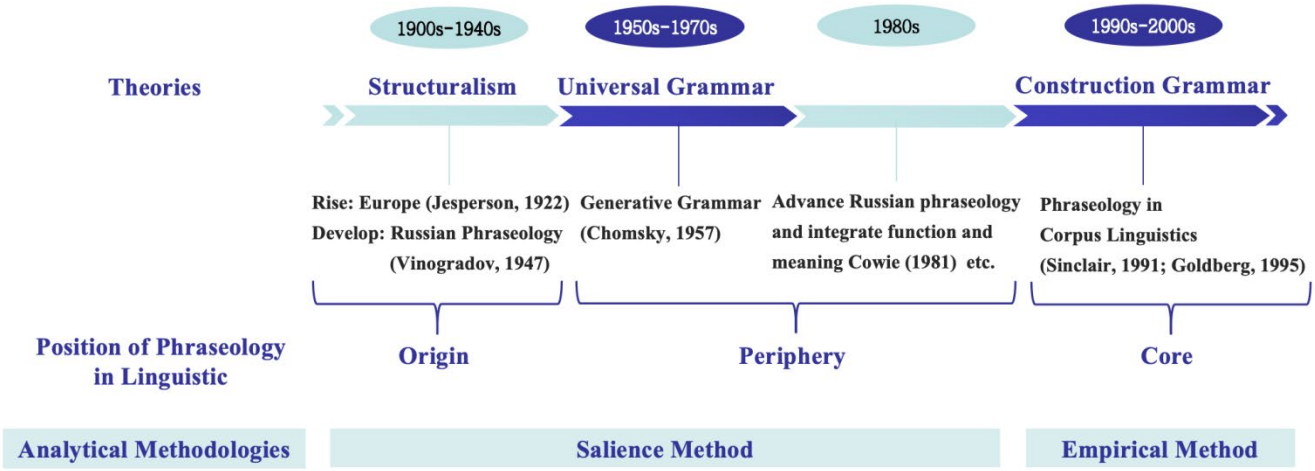
From the 1950s to the 1970s, Chomsky (1957) introduced Generative Grammar, highlighting that language can generate infinite sentences through a limited set of grammatical rules, and proposed Universal Grammar. During this period, phraseology was categorized as peripheral grammar, leading to its marginalization in linguistic studies.

In the 1980s, European scholars incorporated function and meaning into phraseological studies. During this period, phraseological units were identified primarily based on syntactic patterns, employing a top-down approach grounded in linguists' observations, judgments, and inductive summarization of texts. This approach, commonly referred to as the *"phraseological method"* (Granger & Paquot, 2008), rendered phraseological studies predominantly qualitative and subjective. As a result, phraseology remained marginal within linguistic research.

From the late 1990s to the 2000s, advancements in computer technology facilitated the rise of corpus linguistics. Digitized corpora enabled phraseological studies to transition from manual induction and text analysis to a new phase characterized by data-driven quantitative analysis. Notably, Sinclair's *"frequency-driven approach"* (Biber et al., 2004), also referred to as the *"frequency-based approach"* (Granger & Paquot, 2008) provided researchers with a novel perspective to analyze phraseological units without relying on predefined grammatical structures. Concurrently, Construction Grammar (Goldberg, 1995) reconceptualized phraseological units as constructions that inherently integrate form and meaning. As a result, phraseology has gradually shifted from the periphery to the core of linguistic research. Figure 1. below illustrates the entire developmental path.

Figure 1

The Developmental Path of Phraseology in Linguistics



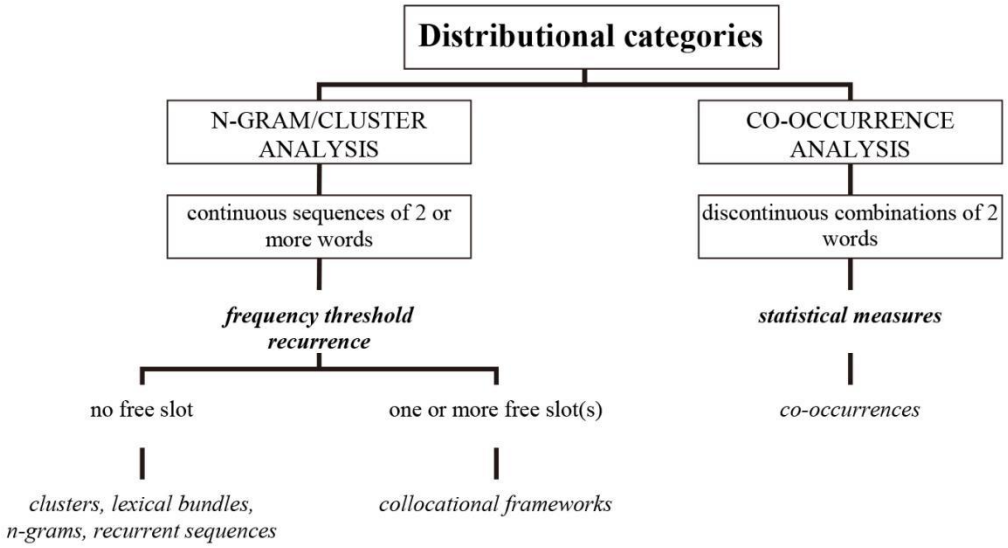
Therefore, as a phraseological unit, the object of this study, LBs, represents a combination of form and meaning. This study will employ a corpus linguistics approach, adopting Sinclair's *"frequency-driven approach"* to extract and analyze lexical bundles.

Definition of Lexical Bundles in Corpus Linguistics Research

As a type of phraseological unit, researchers have yet to reach a consensus on the definition of lexical

bundles, increasing the fuzziness of this field. To help clarify the scope and research focus of phraseological units, Granger and Paquot (2008, p. 39) propose a categorization framework. This classification is illustrated in Figure 2.

Figure 2
Categorization in Phraseological Units (Granger & Paquot, 2008)



This categorization is distribution-based, aligning with the *"frequency-driven approach"* employed for bundle extraction and analysis in the present study. Accordingly, drawing on previous corpus studies that focus on the same phraseological units as the research object, this study adopts the term *"lexical bundles"* (Biber et al., 1999, p. 990). Specifically, the object of this study refers to continuous sequences of two or more words without free slots under N-gram analysis in EFL textbooks. The operational definition is as follows:

LBs are recurrent, continuous sequences of two or more words found in discourse. They exhibit fixed structures, convey relatively complete meanings, and fulfil specific discourse functions.

Global Research on Lexical Bundles in English Language Textbooks

Globally, previous corpus linguistics studies have examined lexical bundles across various contexts (e.g., Chen & Baker, 2010, on academic writing). However, relatively few studies have focused specifically on LBs in EL textbooks (Hoang & Crosthwaite, 2024).

Nevertheless, a systematic review of the past two decades (2004–2024) indicates an upward trend in research on LBs in EL textbooks (Zhang et al., 2025b). On the one hand, this trend reflects a growing scholarly interest in LBs and their role in English language acquisition. On the other hand, it demonstrates a clear methodological preference for empirical, data-driven corpus analysis, which allows researchers to objectively extract and quantitatively analyze lexical bundles, thereby providing a robust methodological foundation in this realm.

Specifically, from the perspective of English textbooks, a higher proportion are set in a non-native English language context, with EFL contexts being the most prevalent. In terms of educational levels, most studies use tertiary-level EL textbooks as research samples (78%). Regarding the types of sample textbooks, English for Academic Purposes (EAP) textbooks constitute the highest proportion. In contrast, studies focusing on English for General Purposes (EGP) and English for Specific Purposes (ESP) textbooks at the tertiary level are relatively limited, with foundational-level EGP textbooks being particularly scarce. Therefore, based on existing research on LBs in EL textbooks worldwide, selecting foundational-level EFL textbooks, particularly EGP textbooks, as research samples to investigate lexical bundles within them represents a field of significant value that has long been neglected. This study will align with the research trend focusing on EFL contexts by selecting foundational-level EGP textbooks within this context as research samples.

From the perspective of target bundle selection, prior studies have examined bundles of varying lengths, from two-word to longer multi-word sequences, with most (83.33%) concentrating on four-word bundles. On the one hand, longer LBs tend to have lower occurrence (Biber et al., 1999), making the selection of four-word bundles conducive to maintaining data manageability. On the other hand, four-word bundles more clearly reflect the structural and functional characteristics of lexical bundles. They are widely considered by researchers to represent the optimal bundle length for lexical bundle analysis. Moreover, researchers have highlighted the significant pedagogical value of three-word lexical bundles (Simpson-Vlach & Ellis, 2010). Therefore, it is recommended that researchers select both three- and four-word lexical bundles as target bundles (Zhang et al., 2025b). This rationale underlies the selection of three-word and four-word lexical bundles as the target bundles in the present study.

From the perspective of analytical dimensions, previous studies usually focus on in-depth analysis of the structure and function of LBs after bundle extraction (e.g., Biber et al., 1999). Meanwhile, various scholars have proposed different classification frameworks for structural and functional analysis (e.g., Biber et al., 1999; Biber et al., 2004). Since frequency is the primary prerequisite for bundle extraction (Hejazi, 2022), this study prioritizes frequency as the first analytical dimension when exploring the characteristics of LBs in textbooks. Following this, in line with the existing research traditions (Zhang et al., 2025b), a detailed analysis of the structure and function of LBs will be conducted. In summary, the analysis of bundle features in this study will be carried out from three dimensions: bundle frequency, structure, and function.

Previous Research on Lexical Bundles in Chinese EFL Textbooks

In the EFL context, China has the largest number of EFL learners worldwide. Therefore, focusing on LBs in Chinese EFL textbooks has significant global value in non-native English settings (Zhang et al., 2025b). Based on this, Chinese EFL textbooks were selected as the research objects of the present study.

The review of the relevant literature from the past two decades (2004–2024) in the core database of China National Knowledge Infrastructure (CNKI) reveals that research on LBs in Chinese textbooks is relatively limited.

Specifically, regarding the educational levels of the target textbooks, most studies (75%) focus on the tertiary educational stage (vocational colleges and universities) (Peng, 2016; Peng & Wang, 2020; Ren, 2014). In contrast, studies on the elementary educational stage (primary and secondary schools) are

relatively scarce (25%). This phenomenon aligns with the global research trend, which similarly lacks sufficient attention to LBs in foundational-level textbooks. From the perspective of the types of textbooks, existing research focuses on ESP textbooks (62.5%) and EGP textbooks (37.5%), with no studies addressing EAP textbooks. The main reason for this phenomenon lies in the long-standing delay in EAP textbook research and development in China, resulting in a limited selection of available EAP textbooks (Wang, 2020). Research on EAP textbooks, including their evaluation and teaching practices, has gradually emerged after 2010, and the number of studies remains limited (Zuo & Ning, 2019). Moreover, in studies involving the corpus analysis approach, the analytical dimensions similarly involve examining the structural and functional features of lexical bundles after their extraction (Ren, 2014; Peng, 2016).

To sum up, the existing relevant Chinese studies on LBs exhibit a lack of data-driven corpus analysis approaches. The target textbooks of the existing corpus research involve university EGP textbooks, junior secondary school EGP textbooks dialogue section, and ESP textbooks (Peng, 2016 etc.). However, there is a notable absence of comprehensive and detailed corpus analyses of lexical bundle features across all sections of foundational-level EGP textbooks published by the People's Education Press (PEP) in mainland China (2019) as a research sample. Given the widespread use of this textbook series nationwide, this study aims to fill the gap by selecting it as the research sample.

Structural and Functional Analysis Framework for Lexical Bundles

Previous studies have adopted various classifications, with the foundational work beginning with Biber et al. (1999). The present study adopts the structural and functional classification framework proposed by Biber et al. (2004). This framework aligns closely with the focus of this study, as it includes the analysis of LBs in university-level English textbooks, making it highly relevant to the analysis of textbook bundles. This framework classifies the structure and function of LBs into three groups each, as detailed in Table 1 and Table 2, respectively.

Table 1
Structural Classification of Lexical bundles (Biber et al., 2004)

Broad Classifications	**Sub-classifications**	**Examples**
LBs that incorporate verb phrase (VP) Fragments (VP fragments)	(connector +)1st/2nd person pronoun + VP fragment	*you don't have to, well I don't know*
	(connector +) 3rd person pronoun + VP fragment	*it's going to be, that's one of the*
	Discourse marker + VP fragment	*I mean you know, you know it was*
	Verb phrase (with non-passive verb)	*have a lot of, is going to be*
	Verb phrase with passive verb	*can be used to, is based on the*
	Yes-no question fragments	*are you going to, do you want to*
	WH-question fragments	*what do you think, how many of you, what does that mean*

LBs that incorporate dependent clause fragments (Dependent clause fragments)	1st/2nd person pronoun + dependent clause fragment	*I want you to, I don't know why, you might want to*
	WH-clause fragments	*when we get to, what I want to*
	If-clause fragments	*if you want to, if you have a*
	(verb/adjective +) to-clause fragment	*to be able to, to come up with*
	That-clause fragments	*that there is a, that I want to*
LBs that incorporate noun phrase & prepositional phrase fragments (NP & PP fragments)	(connector +) Noun phrase with of-phrase fragment	*one of the things, the end of the, a little bit of*
	Noun phrase with other post-modifier fragment	*a person who, a little bit about, those of you who*
	Other noun phrase expressions	*the main point, a little bit more*
	Prepositional phrase expressions	*in the future, of the things that*
	Comparative expressions	*as well as the, as far as the*

Table 2

Functional Classification of Lexical bundles (Biber et al., 2004)

Broad Classifications	**Sub-classifications**		**Examples**
Stance Expressions	Epistemic Stance		*Personal: I don't know if, I don't think so*
			Impersonal: are more likely to, in the fact that the
	Attitude/ Modality Stance	Desire	*Personal: I want you to, do you want to*
		Obligation/ Directive	*Personal: I want you to, you have to be*
			Impersonal: it is possible to, it is necessary to
		Intention/ Prediction	*Personal: I'm not going to, are you going to*
			Impersonal: it's going to be, going to have a
		Ability	*Personal: to be able to, to come up with*
			Impersonal: can be used to, it is possible to
Discourse Organizers	Topic introduction/focus		*what do you think, if you look at*
	Topic elaboration/clarification		*has to do with, I mean you know*
Referential Expressions	Identification/Focus bundles		*that's one of the, and one of the*

	Imprecision		*or something like that, and things like that*
	Specification of attributes	Quantity specification	*have a lot of, how many of you*
		Tangible framing attributes	*the size of the, in the form of*
		Intangible framing attributes	*the nature of the, in the case of, on the basis of*
	Time/place/text reference	Place reference	*in the United States, of the United States*
		Time reference	*at the same time, at the time of*
		Text deixis	*show in figure N, as show in figure*
		Multi-functional reference	*the end of the, at the end of*
Special Conversational Functions	Politeness		*thank you very much*
	Simple inquiry		*what are you doing*
	Reporting		*I said to him/her*

METHODOLOGY

Corpus Linguistic

This study adopts corpus linguistics as its guiding methodology. Corpus linguistics involves the empirical analysis of large, electronically stored, and representative text collections, known as corpora, as the basis for analysis. As reviewed in the literature, digitized corpora offer a robust *"empirical method"* for examining phraseological units such as LBs .

Specifically, this study is a corpus-based study under Sinclair's *"frequency-driven approach"* (Biber et al., 2004). In corpus-based research, a potentially significant linguistic phenomenon is pre-selected as the research object. By examining authentic corpus data, this approach aims to reveal the usage patterns of the selected phenomenon (Biber, 2009). The data are analysed both quantitatively and qualitatively, with corpus concordance software serving as the analytical tools. Details regarding the selected textbooks and software are provided below.

Target Textbooks and Self-Constructed Textbook Corpora

In line with the literature review and prevailing research trends focusing on non-native EFL contexts, and in response to addressing the gap in studies on foundational-level EGP textbooks, this study selects the widely used secondary school EFL textbooks published by the People's Education Press (PEP) in mainland China since 2019 as research samples: five from junior secondary school and three from senior secondary school (Compulsory Volumes One to Three).

The full content of each textbook was included in the corpora, except for word lists and

gramm*ar/pronunciation explanations in the appendices (tapescripts were kept). The target textbooks corpora consist of two components: Chinese Junior Secondary School EFL Textbooks Corpus (CJSTC) and Chinese Seni*or Secondary School EFL Textbooks Corpus (CSSTC). Detailed information is summarized in Table 3.

Table 3
Composition of Target Corpora

Corpora	Size in tokens (running words)	Size in types (distinct words)
Chinese Junior Secondary School EFL Textbooks Corpus (CJSTC)	107,760	4,221
Chinese Senior Secondary School EFL Textbooks Corpus (CSSTC)	84,166	6,415
Total	**191,926**	**7,768**

Selection of Corpus Concordance Software

Various statistical software tools facilitate the fully automated extraction of LBs, as demonstrated in previous studies, including WordSmith Tools (Scott, 2016). In the present study, the extraction process begins with WordSmith Tools (Scott, 2016) and is subsequently processed in Collocate (Barlow, 2004) using its *'batch collocation'* function to analyze the Mutual Information Score (MI scores).

Developed by Mike Scott, WordSmith Tools (available at http://www.lexically.net) is a widely used and reputable concordance software in corpus linguistics research (e.g., Alasmary, 2022.). It provides three analysis functions: *Concord/KWIC (Key Word in Context), Keywords, and Wordlist.* For this study, the *Wordlist* function will generate both a wordlist and an index wordlist from the plain text. The clustering settings within the *Wordlist* function will help identify LBs by adjusting parameters such as bundle length, cut-off frequency and dispersion. Furthermore, the *Concord* function will display the immediate context of each bundle, aiding in the manual classification of their structure and function within the sample texts.

Developed by Michael Barlow, the software Collocate (available at https://www.athel.com) allows for the extraction of bundles by setting a cut-off frequency, while also providing three statistical measures: MI scores, Log Likelihood Values, and the T-Score. However, it does not consider the frequency and dispersion of bundles, making it a complementary software to WordSmith Tools.

Target Lexical Bundles and Bundle Extraction Criteria

Drawing on the literature review, this study targets 3- and 4-word LBs. Due to the large number of LBs and the precedent set by previous studies focusing on high-frequency bundles, this study limits its analysis to the top 50 LBs. In other words, the 50 most frequent three- and four-word bundles are selected as the target bundles for detailed analysis.

The criteria for bundle extraction involve cut-off frequency, dispersion, and Mutual Information (MI) scores. Specifically, regarding cut-off frequency, previous studies have not established a consistent

threshold due to variations in research objectives and corpus size, with cut-offs ranging from 2 to 40 occurrences per million words (pmw). The cut-off threshold should be adjusted according to the specific characteristics and requirements of the target corpus. In this study, a conservative threshold of 40 pmw is adopted to ensure the extraction of a sufficient number of 3- and 4-word LBs.

Concerning dispersion, bundles that appear frequently within a specific discourse often reflect discourse-specific concerns rather than general LBs (Biber et al., 1999). Therefore, it is necessary to set a restriction. Dispersion can be quantified by the number or proportion of texts in which a bundle appear. Given that relatively short texts are in the target textbooks, this study follows Hoang & Crosthwaite (2024) by setting a minimum dispersion requirement of appearing in at least two textbooks.

MI scores are used as a supplementary statistical measure during the bundle extraction process, assessing the strength of association between words to determine whether their co-occurrence is statistically significant rather than coincidental (Manning & Schuetze, 1999). Simpson-Vlach and Ellis (2010) demonstrated a strong correlation between MI scores and the pedagogical value of LBs, based on expert rankings. It is widely accepted that an MI score of 3 or higher indicates a significant association between words (Hsu, 2014). Therefore, this study selects bundles with MI scores ≥ 3 as the threshold.

After the software extraction, all bundles undergo manual screening. Following the guidelines provided by Chen and Baker (2010) and the seven guiding questions outlined by Hsu (2014), LBs are manually filtered. The manual filtering process is conducted by two peers. To ensure the accuracy of lexical bundle data collection, peer consensus is achieved.

FINDINGS

Descriptive Statistics of Extracted Lexical Bundles

An initial extraction in two textbook corpora yielded 606 three-word types and 183 four-word lexical bundle types using the WordSmith Tool. These bundles were then entered into the software Collocate to obtain their MI scores. A total of 246 lexical bundle types were missing in Collocate (171 three-word types and 75 four-word types), likely due to differences in recognition and extraction models between the tools. For example, *"attention to the"* is recognized in the WordSmith tool but does not meet the collocation criteria of the software Collocate, which emphasizes statistically significant collocational strength. Consequently, these missing bundles (246 types in total) were excluded from further analysis, along with two bundles that did not meet the MI threshold (*"it to the"*, MI = 2.762377 and *"this is the"*, MI = 2.650701).

A subsequent manual screening identified 97 overlapping bundles (e.g., *"again and answer"*, *"can you think"*) and 30 bundles are semantically incomplete bundles (e.g., *"about what you", "is there a"*), and 2 bundles that crossed sentence boundaries (*"long have you"*, *"this unit you will"*). Following these exclusions, the final dataset included 305 three-word types and 107 four-word types. Table 4 compares lexical bundle statistics before and after refinement through manual screening.

Table 4
Overall Statistics of Target Bundles in Two Textbook Corpora

Corpora	LBs	Initial Extraction		Post-Refinement	
		Types	Tokens	Types	Tokens
Two Textbook Corpora	3-word LBs	606	10,849	305	4,611
	4-word LBs	183	3,078	107	2,032
	Total	789	13,927	412	6,643

As shown in Table 4, three-word bundles (4,611 bundle tokens) are significantly more frequent than four-word bundles (2,032 bundle tokens) in the Chinese secondary school EFL textbooks.

To enable comparative analysis between the two textbook corpora with different corpus sizes, the types and tokens of each bundle length have been normalized to occurrences per million words (pmw). Additionally, the normalized lexical bundle type-token ratio (TTR) has been calculated to measure bundle diversity (Hoang & Crosthwaite, 2024). A higher TTR indicates greater bundle diversity and lower repetitiveness. The normalized statistical results are displayed in Table 5.

Table 5
Normalized Statistics of Top 50 Target Bundles in Two Textbook Corpora

Corpora	LBs	Normalized Bundle Types (pmw)	Normalized Bundle Tokens (pmw)	Normalized TTR
CJSTC	3-word LBs	260.52	6,434.77	0.040
	4-word LBs	255.31	5,257.23	0.049
	Total	260.52	11,692.01	0.022
CSSTC	3-word LBs	234.47	2,751.06	0.085
	4-word LBs	197.99	2,490.54	0.079
	Total	260.52	5,241.60	0.050

According to Table 5, the top 50 most frequent three-word and four-word LBs appear significantly more frequently in junior secondary school EFL textbooks (11,692.01 normalized bundle tokens) compared to their senior-level counterparts (5,241.60 normalized bundle tokens). This difference is particularly pronounced in three-word LBs (6,434.77 > 2,751.06).

In terms of bundle TTR, CSSTC exhibits a higher overall TTR (0.050) compared to CSSTC (0.022), indicating that the top 50 most frequent LBs in senior textbooks are less repetitive and demonstrate greater diversity than those in junior textbooks.

Structural Patterns of Lexical Bundles in Textbook Corpora

Based on the classification framework, this study categorizes the target LBs into three broad categories: Verb Phrase Fragments (VP fragments), Dependent Clause Fragments, and Noun Phrase and Prepositional Phrase Fragments (NP & PP fragments). Their distribution is visualized in Figures 3 & 4.

Figure 3
Broad Structural Distribution of Three-word Lexical Bundles in Textbooks

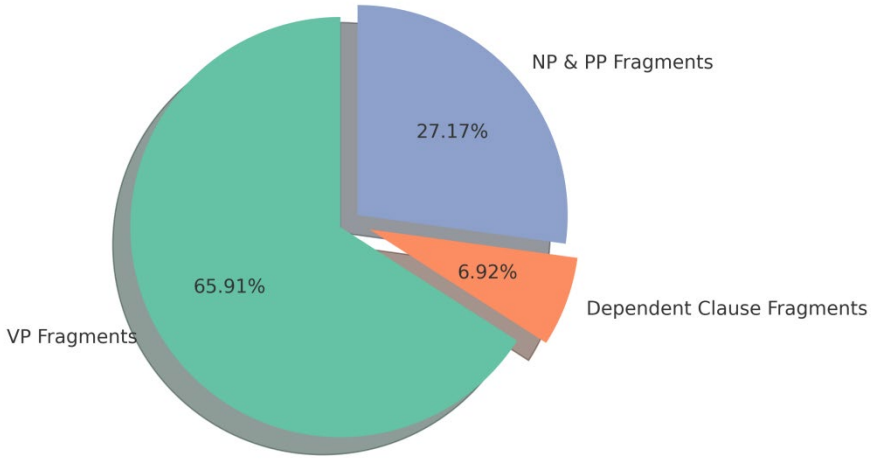
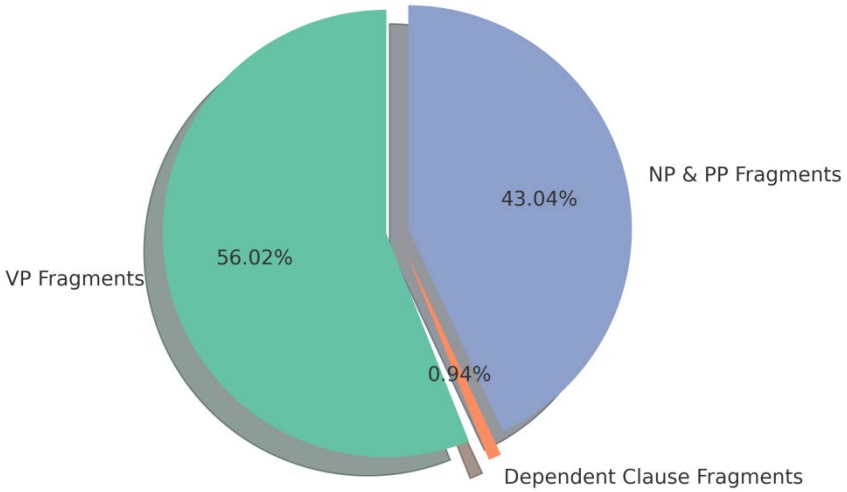


Figure 4
Broad Structural Distribution of Four-word Lexical Bundles in Textbooks



Statistical data show that both the most frequent three-word and four-word LBs cover all three broad structural categories, exhibiting a relatively consistent distribution between the two bundle lengths. The leading structure is VP fragments (65.91% in 3-word bundles, 56.02% in 4-word bundles). NP & PP fragments occupy the second position (27.17% in 3-word bundles, 43.04% in 4-word bundles). Dependent clause fragments represent the smallest proportion (6.92% in 3-word bundles, 0.94% in 4-word bundles).

Further analysis of the sub-categories reveals distributional differences between three-word and four-word LBs. The results of this detailed examination are summarized in Tables 6 and 7.

Table 6
Sub-structural Distribution of Three-word Lexical Bundles by Token in Textbooks

Classifications	Raw Frequency	Percentage
VP fragments	**1162**	**65.91%**
(connector +)1st/2nd person pronoun + VP fragment (connector +) 3rd person pronoun + VP fragment Discourse marker + VP fragment	44	2.50%
Verb phrase (with non-passive verb)	724	41.07%
Verb phrase with passive verb Yes-no question fragments	247	14.01%
WH-question fragments	147	8.34%
Dependent clause fragments	**122**	**6.92%**
1st/2nd person pronoun + dependent clause fragment WH-clause fragments	23	1.30%
If-clause fragments (verb/adjective +) to-clause fragment That-clause fragments	99	5.62%
NP & PP fragments	**479**	**27.17%**
(connector +)Noun phrase with of-phrase fragment	67	3.80%
Noun phrase with other post-modifier fragment	70	3.97%
Other noun phrase expressions	53	3.01%
Prepositional phrase expressions	268	15.20%
Comparative expressions	21	1.19%

Table 7
Sub-structural Distribution of Four-word Lexical Bundles by Token in Textbooks

Sub-classifications	Raw Frequency	Percentage
VP fragments	**833**	**56.02%**
(connector +)1st/2nd person pronoun + VP fragment (connector +) 3rd person pronoun + VP fragment Discourse marker + VP fragment		
Verb phrase (with non-passive verb)	374	25.15%
Verb phrase with passive verb Yes-no question fragments	306	20.58%
WH-question fragments	153	10.29%
Dependent clause fragments	**14**	**0.94%**
1st/2nd person pronoun + dependent clause fragment WH-clause fragments	14	0.94%
If-clause fragments (verb/adjective +) to-clause fragment That-clause fragments		
NP & PP fragments	**640**	**43.04%**
(connector +)Noun phrase with of-phrase fragment	135	9.08%

Noun phrase with other post-modifier fragment	258	17.35%
Other noun phrase expressions		
Prepositional phrase expressions	247	16.61%
Comparative expressions		

First, within the VP fragments, the Verb phrase (with non-passive verb) accounts for the largest proportion in both bundle lengths (41.07% in three-word bundles and 25.15% in four-word bundles). Examples include *"get to school", "talk about the", "is one of the", and "look at the picture"*. Additionally, Yes-no question fragments also constitute a significant proportion (14.01% in three-word bundles and 20.58% in four-word bundles), featuring highly interactive expressions such as *"have you ever", "are you doing", "do we need", "can you come"*. Second, while the Dependent clause fragments represent the smallest proportion in both bundle lengths, it is notably higher in three-word bundles than in four-word bundles (6.92% > 0.94%). Of particular interest is the (verb/adjective +) to-clause fragment, which accounts for 5.62% of three-word bundles but does not appear in four-word bundles. Common examples include "*used to be", "going to do"*, and "*be able to"* Lastly, within the NP & PP fragments, three-word bundles cover all subcategories, with Prepositional phrase expressions comprising the highest proportion (15.20%). In contrast, while four-word bundles also encompass multiple subcategories, they lack Other noun phrase expressions and Comparative expressions.

Functional Classification of Lexical Bundles in the Textbook Corpora

Based on the functional classification proposed by Biber et al. (2004), this study categorizes target LBs into four broad classifications: Stance Expressions, Discourse Organizers, Referential Expressions and Special Conversational Functions. Tables 8 and 9 present the distribution of these categories by bundle type respectively.

Table 8
Functional Distribution of Three-Word Lexical Bundles by Type in Textbooks

Bundle Functions		**Bundle Types Count**
Stance Expressions		**24**
Epistemic Stance		2
Attitude/	Desire	
Modality Stance	Obligation/Directive	17
	Intention/Prediction	3
	Ability	2
Discourse Organizers		**12**
Topic introduction/focus		2
Topic elaboration/clarification		10
Referential Expressions		**19**
Identification/Focus bundles		7
Imprecision		1
Specification of attributes	Quantity specification	2
	Tangible framing attributes	4
	Intangible framing attributes	4

Time/place/ text reference	Place reference Time reference Text deixis Multi-functional reference	6 6
Special Conversational Functions		**11**
Politeness Simple inquiry Reporting		11

Table 9 Functional Distribution of Three-Word Lexical Bundles by Type in Textbooks

Bundle Functions		**Bundle Types Count**
Stance Expressions		**33**
Epistemic Stance		2
Attitude/ Modality Stance	Desire	3
	Obligation/Directive	27
	Intention/Prediction	1
	Ability	
Discourse Organizers		**27**
Topic introduction/focus		3
Topic elaboration/clarification		24
Referential Expressions		**10**
Identification/Focus bundles		3
Imprecision		1
Specification of attributes	Quantity specification	
	Tangible framing attributes	
	Intangible framing attributes	
Time/place/ text reference	Place reference	2
	Time reference	4
	Text deixis	
	Multi-functional reference	
Special Conversational Functions		**7**
Politeness		1
Simple inquiry		7
Reporting		

Overall, both lengths of LBs cover four functional categories. Among them, Stance Expressions are the most dominant, with 24 types in three-word bundles and 33 types in four-word bundles. Notably, the sub-functional category Obligation/Directive is the most prevalent, accounting for 75% (18/24) of the three-word bundles and 81.8% (27/33) of the four-word bundles. Next, Discourse Organizers occupy a significant position in both bundle lengths, with the sub-category Topic Elaboration/Clarification being the most prominent. Furthermore, Referential Expressions are particularly prominent in three-word bundles, mainly concentrated in Identification/Focus bundles (7 types) and Time/place/text reference (12 types). Finally, Special Conversational Functions represent the least frequent functional category across both bundle lengths.

DISCUSSION

This study aims to extract LBs from Chinese secondary school EFL textbooks and analyse bundle structural and functional characteristics. The findings will be discussed in alignment with the research questions.

RQ1: What lexical bundles are identified in Chinese secondary school EFL textbooks in terms of frequency and length?

In terms of bundle frequency (i.e., bundle occurrence or bundle tokens), three-word LBs are more prevalent than four-word bundles in textbooks. On the one hand, this result verifies the occurrence patterns of LBs of varying lengths, indicating that shorter bundles tend to occur more frequently in discourses (Biber et al., 1999.). This is because three-word LBs are *"extended collocational association"*, while longer bundles are *"more phrasal in nature and correspondingly less common"* (Conrad et al., 2002). On the other hand, this finding aligns with the results found in EFL textbooks from other countries. For instance, Priyatno et al. (2023) reported that three-word bundles were the most common among 3- to 6-word bundles in five Indonesian senior high school EFL textbooks.

Although the high token count of the most frequent LBs in junior-level textbooks aligns with the principle that LBs should be repeatedly presented in foundational-level EL textbooks to facilitate bundle acquisition among low-level or beginner English learners (Siyanova et al., 2011; Northbrook et al., 2022), the decrease in high-frequency bundle tokens in senior-level textbooks may hinder learners' mastery of these expressions.

López-Jiménez (2013) examined LBs in English and Spanish as a second language (L2) textbooks and found that the reduced presence of LBs in advanced-level textbooks stems from the belief that higher-proficiency language learners are better equipped with these units, leading textbook developers to omit them and do not need further reinforcement. However, relevant studies have suggested that the frequency of vocabulary (i.e., tokens) in EL textbooks should increase progressively with grade levels, while remaining lower at the initial educational stage to reduce learners' cognitive burden. This approach ensures a balanced distribution of vocabulary across different grade-level textbooks, thereby facilitating more effective vocabulary acquisition (Biseko, 2025; Tang & Liang, 2021). Therefore, the frequency of LBs should follow a similar incremental pattern, enabling students to gradually develop proficiency in lexical bundle usage through continuous exposure and practice as they progress to higher grade levels, ultimately promoting bundle acquisition.

Regarding lexical bundle diversity (i.e., bundle TTR), the target LBs in senior secondary school EFL textbooks exhibit a lower repetition rate, indicating greater bundle diversity. This arrangement meets the criteria for an ideal textbook. Previous studies have pointed out that the richness of vocabulary and phraseological units (e.g., LBs) in EL textbooks should increase with the advancement of educational stages, as this contributes to the improvement of learners' language proficiency (Tang & Liang, 2021).

RQ2: What structural patterns characterize these lexical bundles?

Aligned with the structural rankings of LBs in Chinese university-level EFL textbooks reported by Ren (2014), these findings indicate that VP fragments and NP & PP fragments are the most prominent structural categories in EFL textbooks at both the secondary and tertiary education levels. This suggests that, compared to clause fragments, Chinese EFL textbook compilers show a preference for

phrase fragments, reflecting a phrasal style. In terms of the distribution of sub-structural categories, four-word LBs, compared to three-word bundles, lack two sub-structures. This absence suggests that three-word bundles exhibit greater flexibility and diversity in expressing noun phrases and their modifiers within textbooks.

Previous studies have shown that the structural characteristics of LBs vary across different registers. Spoken discourse (e.g., conversation) tends to use more clausal bundles (Biber et al., 2004), while written discourse, particularly academic writing, relies more on phrase fragments (Chen & Baker, 2016). Therefore, the findings of bundle structural distribution in this study indicate that Chinese secondary school EFL textbooks exhibits a tendency toward academic writing, which is consistent with the register characteristics of Chinese university-level EFL textbooks (Peng, 2016).

However, studies have shown that native academic writing or professional academic writing tends to use NP and PP fragments more intensively (Biber et al., 1999; Biber et al., 2004; Lee, 2020; Chen & Baker, 2016; Shirazizadeh & Amirfazlian, 2021). In contrast, non-native speakers (ESL and EFL) are more likely to use verb-based bundles (Lu & Deng, 2019). For example, Lee (2020) examined the structures of three- and four-word LBs in native university-level EAP textbooks and found that NP and PP fragments predominated (80%), in contrast to the VP fragments that dominate Chinese secondary school EFL textbooks. Similarly, Biber et al. (1999) noted that noun phrase with of-phrase fragments constitute the largest proportion of academic prose. In the target textbooks analyzed in this study, the proportion of VP fragments within phrase fragments is higher than that of NP and PP fragments. Specifically, the *"Noun phrase with of-phrase fragment"* accounts for only 3.80% among three-word bundles and 9.08% among four-word bundles. This indicates that although the linguistic features of Chinese secondary school English textbooks show a tendency toward written academic style, as non-native authored texts, their writing style is immature, highlighting a gap compared to professional academic writing by native speakers.

The structural proportional distribution of lexical bundle in the target Chinese textbooks also differs from those found in textbooks from similar EFL contexts. In their analysis of Indonesian senior secondary school EFL textbooks, Priyatno et al. (2023) found that NP & PP fragments were the most common structures, followed by dependent clause fragments, with VP fragments being the least frequent. In contrast, VP fragments appear most frequently in Chinese secondary school textbooks. One reason for this is that textbooks need to provide instructional language for students (e.g., *"answer the questions"* and *"read the passage"*). Additionally, VP fragments such as *"get to school"*, *"go to the movies"*, and *"go to the party"* are situational expressions that are closely related to students' daily experiences. Other VP fragments, such as *"is one of the"*, are often used to categorize specific items. These fragments are simple in structure, highly practical, and easy to memorize, making them foundational structures that frequently appear in Chinese EFL textbooks at this early stage of English language learning in Chinese secondary school.

RQ3: What functional classifications are assigned to these lexical bundles?

Based on the functional distribution of target LBs by type, there are differences in the bundles' functional tendencies across bundle lengths. These differences are mainly reflected in two categories: Discourse Organizers and Referential Expressions. Three-word bundles show a stronger tendency toward Referential Expressions, particularly in Time/Place/Text Reference and Identification/Focus bundles. In contrast, four-word LBs contain a greater variety of bundle types on Discourse Organizers.

Discourse Organizers play a crucial role in clarifying discourse relationships (Hoang & Crosthwaite, 2024; Biber et al., 2004), enhancing the logicality and coherence of textbook content (Peng, 2016), as well as improving reader-friendliness (Hussain et al., 2021). Examples of such bundles in the textbooks include *"correct forms of the"*, *"with the correct forms"*, and *"forms of the words"*. These bundles are primarily used to further clarify instructional language, thereby significantly contributing to the enhancement of textual coherence and logicality. This finding reflects the specific functional orientation of different bundle lengths within textbook language.

Regardless of whether they are three-word or four-word LBs, the most frequent functional category is Stance Expressions, particularly the Obligation/Directive sub-category. Representative bundles include *"fill in the blanks"*, *"correct forms of the"*, and *"with the correct forms"*. These bundles are instructional language, serving important functions in organizing learning tasks, guiding classroom activities, and facilitating interactions between teachers and students. This reflects the strong emphasis on instructional language in the design of secondary school English textbooks.

This phenomenon can be attributed to the long-standing emphasis on teacher-centered pedagogical models in traditional Chinese education, which highlights the authority of teachers. Therefore, this tendency is reflected in the design of LBs in textbooks to further reinforce the dominant role of teachers in the classroom.

In contrast, Special Conversational Functions exhibit the fewest bundle types. Examples include *"what kind of"*, *"have you ever"* and *"why do you like"*. These bundles predominantly appear in the Listening, Speaking, and Exercises sections of the target textbooks, thereby enhancing the communicative function of the textbooks. They familiarize students with common question-and-answer patterns, thus improving their language communication skills. However, the relatively limited types of such bundles may constrain students' ability to express themselves effectively in authentic communicative contexts.

CONCLUSION

This study explored EFL textbooks from a phraseological perspective, focusing on LBs. Corpus concordance software, including WordSmith Tools and Collocate, was employed to extract and analyse lexical bundle features in the eight most widely used Chinese secondary school EFL textbooks. The analysis was conducted along three dimensions: bundle frequency, structural and functional classifications.

The findings conclude that, overall, three-word bundles occur significantly more frequently than four-word bundles in these EFL textbooks, a trend consistent with the occurrence patterns of LBs and findings from previous research in other EFL contexts. Further analysis reveals that as the grade level of the textbooks advances, the diversity of LBs increases (i.e., repetition rates decrease), which is beneficial for enhancing students' language proficiency. However, the observed decrease in high-frequency bundle tokens in senior-level textbooks may hinder learners' mastery of these bundle expressions.

In terms of structural patterns, the leading structure is verb phrase fragments, followed by noun phrase and prepositional phrase fragments, with dependent clause fragments being the least frequent. This ordering aligns with the findings on bundle structural features in Chinese university-level EFL textbooks. Notably, it suggests that Chinese EFL textbook compilers tend to favor phrase fragments, reflecting a phrasal style and indicating that Chinese secondary school EFL textbooks exhibit a tendency toward academic writing in terms of register. However, as non-native authored texts, these

textbooks still exhibit a relatively immature writing style compared to professional academic writing by native speakers, with an excessive emphasis on verb phrase fragments driven by their pedagogical orientation. Additionally, compared to three-word bundles, four-word bundles show greater flexibility and diversity in expressing noun phrases and their modifiers within the textbooks.

Regarding bundle functions, LBs of different lengths exhibit distinct functional tendencies in the textbook language. Three-word LBs predominantly feature Referential Expressions, while four-word LBs contain a greater variety of bundle types on Discourse Organizers. However, across both lengths, the Obligation/Directive sub-category holds the greatest bundle types, primarily used for issuing pedagogical instructions, reflecting the teacher-centered pedagogical models characteristic of Chinese education. Additionally, the target textbooks demonstrate a deficiency in facilitating students' acquisition of bundle types related to communicative expression.

Situated within the field of phraseology, this study explores the features of bundle usage in the textbook genre. By examining Chinese secondary school EFL textbooks as samples, this study offers insights into bundle usage in non-native English context textbooks. This exploration not only enhances instructors' awareness of the textbooks they use, facilitating more targeted pedagogical approaches to LBs but also provides textbook developers insights into the bundle usage in these textbooks, guiding future revisions and improvements.

Based on the findings drawn from the bundle features in this study, future research could further examine LBs in textbooks. Potential directions include assessing the appropriateness of bundle selection in textbooks, comparing the differences in bundle usage across various textbook editions, and evaluating students' acquisition of the bundles presented in the textbooks.

REFERENCES

Alasmary, A. A. (2022). Comparing lexical bundles across three advanced mathematical text types: a corpus-based genre-focused investigation. *SAGE Open, 12*(3). https://doi.org/10.1177/21582440221113824

Barlow, M. (2004). *Collocate (Version 1.0)* [Computer software]. Athelstan. https://www.athel.com

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

Biseko, J. M. (2025). Vocabulary learning in EFL context: Do primary school English subject textbooks provide structured support? *Cogent Education, 12*(1), 2455047. https://doi.org/10.1080/2331186X.2025.2455047

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 2010, 14*(2), 30–49.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Conrad, S., Biber, D., & Leech, G. N. (2002). *Longman student grammar of spoken and written English workbook*. Harlow: Longman.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27–49). Amsterdam, The Netherlands: John Benjamins Publishing Company.

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]. 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, 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
Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse functions of lexical bundles in Pakistani chemistry and physics textbooks. *Journal of Language Studies, 21*(1), 191–209. https://doi.org/10.17576/gema-2021-2101-13
Lee, H. K. (2020). Lexical bundles in linguistics textbooks. *Linguistic Research, 37*(1), 121–145. https://doi.org/10.17250/khisli.37.1.202003.005
Liu, Y., Zhang, L. J., & Yang, L. (2022). A corpus linguistics approach to the representation of western religious beliefs in ten series of Chinese university English language teaching textbooks. *Frontiers in Psychology, 12*, 789660. https://doi.org/10.3389/fpsyg.2021.789660
Lu, X., & Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes, 39*, 21–36. https://doi.org/10.1016/j.jeap.2019.03.008
López-Jiménez, M. D. (2013). Multi-word lexical units in L2 textbooks. *Revista española de lingüística aplicada,* (26), 333-348.
Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
Northbrook, J., Allen, D., & Conklin, K. (2022). 'Did you see that?'—The role of repetition and enhancement on lexical bundle processing in English learning materials. *Applied Linguistics, 43*(3), 453–472. https://doi.org/10.1093/applin/amab063
Peng, Y. M. (2016). A comparative study on lexical chunks in medical English textbooks and general college English textbooks. *Educational Modernization (Electronic Version),* (33), 235-237. https://doi.org/10.16541/j.cnki.2095-8420.2016.33.084.
Peng, Y. M., & Wang, Y. (2020). Construction and application of a high-frequency medical English terminology lexical chunk list. *University: Research and Management,* (46), 48-50.
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), 25–39. https://doi.org/10.17323/jle.2023.13972
Ren, X. H. (2014). A study on lexical chunks in college English textbooks. *Language Education,* (3), 41-46.
Scott, M. (2016). *WordSmith Tools (Version 7.0)* [Computer software]. Lexical Analysis Software. http://www.lexically.net
Shamov, A. N., & Shimichev, A. S. (2024). Foreign language textbook at school: Conditions for its qualitative creation. *Tambov University Review. Series: Humanities, 29*(6), 1608–1621. https://doi.org/10.20310/1810-0201-2024-29-6-1608-1621
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
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), 776–786.
Tang, M. H., & Liang, M. C. (2021). Analysis of the level difference in lexical complexity in college

English textbooks. *Frontiers in Foreign Language Education Research, 4*(1), 61-68.

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language teaching, 45*(2), 143-179. https://doi.org/10.1017/S0261444811000528

Vinogradov, V. V. (1947). *Russkij jazyk: Grammaticheskoye uchenie o slove [The Russian language: Grammatical teaching of the word]*. Moscow: Uchpedgiz.

Wang, J. (2020). Research on EAP teaching for undergraduate and graduate students in ordinary universities in China. *Journal of Chongqing University of Science and Technology (Social Sciences Edition)*, (5), 101-105. https://doi.org/10.19406/j.cnki.cqkjxyxbskb.2020.05.023

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

Zhang, M., Liu, C., & Mukundan, J. (2025a). Using corpus concordance software for low-effort multidimensional analysis of linguistic style in Chinese secondary EFL textbooks. *International Journal of Learning and Teaching, 11*(5), 279–283. https://doi.org/10.18178/ijlt.11.5.279-283

Zhang, M., Mukundan, J., Khojasteh, L., & Jain, J. A. (2025b). Lexical bundles in English language textbooks: A systematic review. *World Journal of English Language, 15*(6), 100-115. https://doi.org/10.5430/wjel.v15n6p100

Zuo, X., & Ning, Q. (2019). Research on academic English teaching in China since the 21st century: Review and prospects. F*oreign Language World,* (2), 45-52.