

## The Structure and Function of Lexical Bundles in Communicative Saudi High School EFL Textbooks

Alaa Alzahrani\*

Department of English Language and Literature, College of Arts, King Saud University, Riyadh 12441, Saudi Arabia

Corresponding Author: Alaa Alzahrani, E-mail: alaa-zahrani@hotmail.com

### ARTICLE INFO

#### Article history

Received: May 14, 2020

Accepted: July 21, 2020

Published: September 30, 2020

Volume: 9 Issue: 5

Advance access: September 2020

Conflicts of interest: None

Funding: None

### ABSTRACT

Current English Language Teaching (ELT) textbooks have largely adopted the communicative approach by using authentic materials to foster EFL students' communicative competence. However, the communicative status of Saudi high school English textbooks has been underexplored. One way to assess the authenticity of Saudi EFL textbooks is by considering their use of a frequent linguistic item known as lexical bundles. Thus, the present study investigated whether the lexical bundles in communicative Saudi high school textbooks are representative of conversational English. This comparative corpus study used a lexical bundle approach to compare the ten most frequent lexical bundles in the textbooks to those in an English reference corpus. Results show that three and four-word lexical bundles are less frequent in the textbooks compared to the reference corpus and that there is considerable variation in the structural and functional patterns of the bundles in the two corpora. Pedagogical implications are discussed in light of the findings.

**Key words:** Lexical Bundles, High School EFL Textbooks, Authentic EFL Textbooks

### INTRODUCTION

In an EFL context, textbooks are one of the primary sources of language input. Realizing this, current English Language Teaching (ELT) textbooks have adopted mainly the communicative approach by using authentic materials to foster EFL students' communicative competence (e.g., Mitchell & Malkogianni, 2019; Soars & Soars, 2009; Spencer, 2016). In this language learning approach, exposure to real-life language use is expected to help EFL learners in meeting their future language needs. However, most studies on ELT textbooks have shown that the language presented in them does not reflect real language use, failing to equip students with the necessary communicative skills to engage in real-life tasks (Alquraishi, 2014; L. Chen, 2010; Coxhead, Yen Dang, & Mukai, 2017; Gouverneur, 2008; Wood, 2010; Wood & Appel, 2014; Yoo, 2013).

Due to the principal role of textbooks in the EFL context, it is important that ELT textbooks present to EFL learners communicative language that show how language is used in everyday life. Exposure to authentic materials has been reported to facilitate the acquisition and use of language forms that are communicatively effective (e.g., Gilmore, 2011; Widodo, 2012). Although there are pedagogical advantages to introducing authentic materials, some EFL textbooks still fail to represent real language use (e.g., Northbrook & Conklin, 2018b). One way to assess the authenticity of textbooks is by considering their use of a frequent linguistic item known as lexical bundles (e.g., *I would like to*) which cover

around 30% of spoken conversation and 20% of academic writing (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Lexical bundles can, therefore, provide a suitable ground for testing the degree to which a text is communicative due to their frequency and utility in spoken and written English.

Lexical bundles are a type of formulaic language and defined as the most frequently occurring sequences of three or more words (Biber, 2006a). There are several characteristics that distinguish lexical bundles (e.g., *I'm gonna*) from other forms of formulaic language such as phrasal verbs (e.g., *get up*), collocations (e.g., *adopt policies*) and idioms (e.g., *kick the bucket*). First, the meaning of lexical bundles is transparent as it can be readily predicted from each word constituting the bundle. Another defining feature of lexical bundles is that they usually do not form complete structural and semantic units. They come in different grammatical shapes, including clauses (e.g., *I don't want to*) or phrases (e.g., *in the case of*). Although lexical bundles do not have a fixed grammatical structure, they are strongly associated with specific structures (Biber, Conrad, & Cortes, 2004). Having noticed that lexical bundles tend to occur in specific structures, Biber et al. (1999) proposed a structural classification for grouping lexical bundles into the structural categories they mostly tend to occur in. The present study used this classification to detect further patterns in the data.

A further feature of lexical bundles is that they serve three primary functions for the construction of discourse. Follow-

ing an inductive approach, Biber et al. (2004) identified three functions for lexical bundles, and they are: (a) stance expressions, (b) discourse organizers, and (c) referential expressions. While stance bundles express one's point of view and attitude, organization bundles connect two pieces of discourse. Referential bundles identify a physical or abstract object or parts of it to highlight the importance of it. More explanation for these categories is provided in the methodology section. This functional classification was implemented in the analysis of bundles to account for the discourse factors motivating their use.

### Research Problem and Study Aim

The present study focuses on analyzing lexical bundles in Saudi EFL textbooks as their communicative nature has rarely been examined. Little investigation has been carried to understand whether EFL high school students are exposed to authentic content in their classrooms, allowing them to be more proficient speakers in real-life tasks (Gilmore, 2011). As the textbook is the primary source of learning in the Saudi context, examining the EFL textbooks taught in Saudi public high schools is important for enhancing the process of language learning. Therefore, this study aims to investigate whether communicative Saudi high school English textbooks represent authentic lexical bundles used frequently by native speakers.

## STUDIES EXAMINING LEXICAL BUNDLES IN ELT TEXTBOOKS

### University-level Textbooks

Several studies have evaluated ELT university textbooks' use of lexical bundles. Most of them have found a mismatch between the lexical bundles presented in ELT materials and that of a reference corpus (Allan, 2017; Alquraishi, 2014; L. Chen, 2010; Coxhead et al., 2017; Wood, 2010; Wood & Appel, 2014), with only one study reporting a match in their data (Nekrasova-Beker & Becker, 2019). For example, Allan (2017) analyzed lexical bundles in five different self-study books for English language learners to examine their frequency, structural and functional patterns by comparing them to those used in a spoken corpus of conversational English as a Lingua Franca (ELF). Results show that some of the pragmatic functions of lexical bundles in the books analyzed were misrepresented, including hedges (e.g., *I don't know*) and vague language (e.g., *a little bit*) compared to the ELF corpus. While the examined ELF bundles were characteristic of interactive and conversational language (e.g., *do you have, you want to*), those presented in the self-study books had an instructional focus (e.g., *check your answers*).

Similar results were reported by studies examining lexical bundle use in English for Academic Purposes (EAP) textbooks. For instance, Coxhead, Yen Dang, & Mukai (2017) compared the actual use of three and four-word lexical bundles in spoken university tutorials and laboratories and that presented in EAP speaking and listening course books. To do so, the study collected three corpora. The lab corpus

comprises 137,399 words and is compiled from three online corpora of academic spoken English (Michigan, the Limerick-Belfast, and Newcastle Corpus of Academic Spoken English) while the tutorials corpus comprises 380,078 words and is based on two corpora of academic spoken English (the Limerick-Belfast, and Hong Kong Corpus of Spoken English). The third compiled corpus constitutes 15 series of EAP speaking and listening coursebooks published by different companies and one set of English for specific purposes (ESP) textbooks. Results reveal that most of the EAP/ESP textbooks gave little space for recommending lexical bundles suitable for tutorial and laboratory settings. It was found that only three of the examined textbooks suggested to the learner the use of 176 useful lexical bundles for speaking in tutorials, such as ways to keep the discussion on topics (e.g., *let's get back to, we are getting a little off track*), and introduce a new topic (e.g., *let's start with, we need to discuss*). However, there was little correspondence between many of the textbook-based 176 lexical bundles and those occurring in university tutorial and laboratory talk.

Likewise, Wood & Appel (2014) compared the use of three and four-word lexical bundles in five EAP textbooks to those found in ten first-year university business and engineering course books. This was done to examine whether EAP textbooks contain useful language that can support business and engineering students in their later studies. It was reported that the bulk of the lexical bundles used in the first-year business and engineering university books were absent from the EAP textbook readings, revealing that the EAP textbooks gave little or no pedagogical treatment of lexical bundles by not including activities which focus on bundles and their functions. This corroborates the findings of a previous study. In an earlier study, Wood (2010) investigated lexical bundles in a 539,210-word corpus of EAP textbooks and found that the majority of the bundles had referential functions, dealing with location and tangible framing. This indicated that lexical bundles in EAP textbooks are limited in function and tend to focus more on classroom instruction.

Similarly, Alquraishi (2014) compared the functions of lexical bundles in English as a Second language (ESL) textbooks to those in engineering academic texts. In order to do this, the study created 65,000 -word corpus of ESL textbooks and 1.26 million-word corpus of engineering textbooks. Using the functional categorization proposed in Biber et al. (2004), it was shown that there is minimal overlap of lexical bundles in terms of function. Although bundles serving a referential function (e.g., *in the form of, as shown in fig, at the end of*) were common in both corpora, amounting to 48% in the engineering texts and 30% in the ESL course books, the other functions were not distributed evenly across the two corpora. This comparison revealed a gap between the formulaic language that university students encounter in an ESL coursebook and what they will encounter in their engineering textbooks.

### School-level Textbooks

While most of the works analyzing lexical bundles presented in ELT textbooks focused on university-level materials,

little attention has been paid to school ELT textbooks. Only one study looked at lexical bundles in an EFL school textbook. Northbrook & Conklin (2018) examined three-, four-, five- and six-word lexical bundles in Japanese middle school communicative ELT textbooks and compared them with a spoken American English corpus the SUBTLEXus. Findings show that lexical bundles are more frequent in the examined textbooks compared with conversational English. However, although the textbooks include 3-word lexical bundles that structurally and functionally match those found in the reference corpus, this is no longer the case when it comes to longer bundles as they mismatch the conversational English corpus. Northbrook & Conklin concluded that the language used in Japanese junior high school English textbooks does not reflect language use outside the classroom.

One main observation of the literature review reveals that most of the earlier works have focused on lexical bundles used in textbooks for advanced language learners and/or in a university setting with only one work (Northbrook & Conklin, 2018a) examining beginning learners of English in a Japanese middle school context. This indicates that the nature of the language of high school and middle school textbooks needs further examination.

## THE PRESENT STUDY

Based on the literature above, the communicative status of Saudi high school English textbooks has been underexplored. In the Saudi context, much emphasis is given to textbooks as they are considered as the main tool in the English language classroom in high schools. For this reason, the English textbooks series used in Saudi public high schools are co-published with companies located in a native-English speaking country, aiming to ensure that the language of the school textbooks is as native-like as possible. This is shown in a statement mentioned in one of the main English textbooks series used in Saudi high schools stating that the series represents “how English is used in real-life situations” (Mitchell & Malkogianni, 2019: 2). This series is called the Traveler series KSA edition, which aims to show the communicative use of British English as used in real-life situations by native speakers to enable Saudi school-aged learners to transfer this authentic use of language to their everyday tasks. However, the claim that the Traveler series KSA edition represents how natives use the language in authentic situations has not yet been investigated. If the language of this series is communicative and authentic, as mentioned by its designers (Mitchell & Malkogianni, 2019), then it is expected that we will find more similarities than differences between the Traveler series KSA edition and a corpus of communicative British English. This comparison would make it possible to establish whether Saudi high school EFL textbooks are as communicative as the language used by English natives.

Thus, a comparative corpus study using a lexical-bundle approach can offer valuable insights into the nature of English materials in Saudi. The present study aims to examine the lexical bundles presented in the Traveler series KSA edition by comparing them with the output of native language users. The language of the Traveler series is primarily British

English with the purpose of demonstrating to high school students how native speakers use the language to “establish relations, exchange information and express ideas, attitudes, and feelings” (Mitchell & Malkogianni, 2019: 2). The language functions presented in Traveler textbooks are more common in the register of conversational English compared to the registers of fiction, news, or academic English (Biber et al., 1999). Therefore, the language of the Traveler series would be compared with a reference corpus representing conversational language as spoken by native speakers of English. This study aims to do so by answering the following two research questions:

1. Are the most frequent three and four-word lexical bundles in the textbooks used as frequently by native English speakers (as found in British National Corpus 2014)?
2. Are the most frequent three and four-word lexical bundles in the textbooks similar in their structure and function to the ones frequently used by native English speakers (as found in British National Corpus 2014)?

## METHODOLOGY

### Corpus Construction

The present study compiled two corpora, as can be shown in Table 1. The first corpus contains the six high school English textbooks along with their listening components from the 2019/2020 Traveler series KSA edition (Mitchell & Malkogianni, 2019). As Table 1 shows, the total number of words in this textbook corpus is 290,053 words. It should be noted that although there are two other English textbooks series (Macmillan and McGraw Hill) used in Saudi public high schools, only the Traveler series were analyzed to limit the scope of the study. The Traveler series is approved by the Saudi Ministry of Education and is jointly produced by a UK-based publication company (MM publications) and a local one (Tatweer Company for Educational Services). Consequently, the language used in the Traveler series KSA edition is principally British English intending to show “how English is used in real-life situations, thus enabling learners to use it in meaningful contexts” (Mitchell & Malkogianni, 2019: 2). Another defining aspect of the Traveller series is that it follows the requirements of the Common European Framework of Reference for Languages (CEFR). Thus, it focuses on presenting communicative language to students and ultimately enhance their socio-cultural understanding of everyday life patterns of their age group. Also, the designers maintain that the written and audio dialogues contained in the series present real spoken English.

The process of constructing and cleaning the textbook corpus was as follows. The textbooks were downloaded from the web as a pdf format and were converted to Word files to allow manual clean-up. The six textbooks were manually cleaned from the data irrelevant for linguistic analysis, including textbook cover page, textbook titles, names of authors, headings, wordlists. The listening sections in each textbook were included in the corpus to ensure the representativeness of the corpus and to truly reflect the number/type

**Table 1.** Composition of the two corpora, showing the number of texts, *source of texts* and words in each corpus.

Corpus	Number of texts	Source of texts	Number of words
Textbook corpus	12	6 student textbooks 6 transcripts of listening files	290,053
The Sampled British National Corpus 2014 spoken (BNC2014)	54	Sampled from the year 2012 - 2016	290,057

of lexical bundles learners encounter in their ELT textbooks. Six listening components were downloaded and transcribed using web-based software and were further manually checked for error.

As it was mentioned above that the examined Saudi high school textbooks mainly use British English; the British National Corpus 2014 spoken (BNC2014) was selected as the baseline for comparison (Love, Dembry, Hardie, Brezina, & McEnery, 2017). When comparing two corpora, some corpus linguists call for using normalized frequencies rather than raw ones (e.g., Gries, 2010; McEnery & Hardie, 2011). However, normalizing word frequencies in the present study may not be the best approach to compare between the two corpora. This is because comparing the number of occurrences of lexical bundles in the 11.5-million-word BNC2014 with the 290,053-word textbook corpus would inevitably inflate the results and would not allow an accurate comparison. A more appropriate approach, in this case, is creating a representative sample of the BNC2014 similar in size to the textbook corpus (e.g., Allan, 2017; Northbrook & Conklin, 2018b). As such, a sample of BNC 2014 was created using a stratified sampling technique (see Table 2), resulting in a smaller sized sample corpus containing 290,057 words. To sum up, using a representative sample of BNC2014 (290,057 words) rather than the full corpus (11.5-million words) was done to ensure a more accurate comparison with the small-sized textbook corpus (290,053 words).

Following the creation of the two corpora, the next step involved uploading the two corpora to a web-based corpus tool called Sketch Engine, which offers an “n-grams search” to identify the lexical bundles in the selected corpora.

### Unit of Analysis

Only three- and four-lexical bundles are considered here. This is because the majority of previous research has investigated four-word sequences (e.g., Biber et al., 2004; Conrad & Biber, 2004; Coxhead et al., 2017; Wood & Appel, 2014) as five-six word lexical bundles mostly incorporate a three-four sequence, which may not add new insights on the corpus analyzed. Therefore, focusing on three- and four-word bundles was expected to capture a more detailed picture of how this type of formulaic language is used in textbooks.

### Identifying Lexical Bundles

The minimal frequency cut-off for identifying lexical bundles seems to be arbitrary and mostly dependent on the aim of each study (Biber, 2006b). Therefore, to account for the small size of the compiled corpus in the present study,

**Table 2.** Sampling the BNC2014 according to the year of recording

Year of recording	Number of texts	Number of words
2012	12	68,207
2013	3*	18,052
2014	13	67,775
2015	13	68,178
2016	13	67,845
Total	54	290,057

Note 1.

a three-four-word sequence would be qualified as a lexical bundle if it occurred at least 40 times in the corpus. Also, following Northbrook & Conklin (2018), this study only analyzed lexical bundles that occurred in three out of the six Traveler textbooks. That is, this study identified a sequence as a lexical bundle if it was presented in three Traveler textbooks, at least, guarding against the analysis of idiosyncratic uses of lexical bundles. The criteria for identifying lexical bundles in this study are shown in Table 3.

Table 3 shows that the criteria for identifying both three and four-word lexical bundles are identical. Three and four-word bundles should have a frequency of 40 occurrences or more in the 290,000- word textbook corpus to be identified as one. This was expressed in Table 3 by the criterion “40 per 290,000”. The second criterion in Table 3 is range which specifies the spread/distribution of lexical bundles in the six different textbooks in the Traveler series. In this study, a lexical bundle is identified as one if it occurred with a range of three, i.e., across three textbooks. By setting this range limit, the study can ensure that the identified lexical bundles are “representative of the corpus as a whole, and not confined to only a high number of occurrences in a small amount of text or by an individual writer” (Wood & Appel, 2014: 5).

Furthermore, several guidelines were followed in extracting bundles. One involved the treatment of overlapping lexical bundles. Some three-word lexical bundles were repeated in a four-word cluster. For example, three-word units such as “and answer the” as well as “answer the questions” occurred in a longer bundle “and answer the questions”. To deal with this overlap, this study followed Y. H. Chen & Baker’s (2010) approach by combining overlapping three-word sequences into a 4-word one to minimize the risk of inaccurate results. Another aspect followed in extracting lexical bundles in this study is the treatment of contractions (e.g., *don’t*) as a separate word (e.g., *do not*). Therefore, the bundle “*I don’t*” is counted as three words, and “*I don’t know*” as a four-word bundle.

**Data Analysis**

The top ten three- and four-word lexical bundles were extracted from the two corpora using Sketch Engine. Because the two corpora have a similar number of words, only raw frequencies were reported to enable comparison of the bundles between the textbook and BNC corpora. To analyze similarity, the top ten in each bundle size were compared between the two corpora, both structurally and functionally. As mentioned in the introduction, Biber et al. 's (1999) structural as well as Biber et al. 's (2004) functional classifications of lexical bundles are implemented in the present analysis.

Biber et al. 's (1999) structural classification is presented in Table 4 in which the common structures of lexical bundles are listed. Eleven structural patterns are included in this classification, each specifying the grammatical role of the words constituting a bundle. For instance, the bundle "I don't know what" is broken down in terms of its structural elements and is described as a bundle that starts with a personal pronoun (e.g., I, we) followed by a lexical verb phrase (e.g., don't know, cannot go) and with an optional slot for a complement clause (e.g., what he said, that you played). In Table 4, the bracketed words indicate that the inclusion of the structure is optional [e.g., (complement clause)], while the plus sign means followed by. This classification was used in categorizing the lexical bundles in the present study in terms of structure.

Table 5 shows Biber et al. 's (2004) functional classification of lexical bundles highlighting their four common discourse roles. It was mentioned in the introduction that lexical bundles serve three primary functions for the construction of discourse, including (1) expressing stance, (2) organizing

discourse, and (3) expressing referential functions. Stance bundles express epistemic and attitudinal perspectives (e.g., I don't know, I don't think). Organization bundles connect between two pieces of discourse (e.g., if you look at, go to the). Referential bundles make direct reference to an object or to the text itself to highlight its importance (e.g., is one of the, a lot of). Another identified function of lexical bundles by Biber et al. 's (2004) in Table 5 is (4) special conversational functions. This discourse function covers inquiring about something (e.g., what are you doing), reporting to someone (e.g., I said to him/her) and using polite forms of language to indicate gratefulness (e.g., thank you very much), smoothing by this the flow of the conversation. The conversational function seems less comprehensive than the first three; thus, it is less emphasized in Biber et al. 's (2004). This functional classification was used in the analysis of bundles in the current study to explain the discourse factors motivating their use.

**ANALYSIS**

This study was set out to answer two research questions. The first question seeks to examine whether the most frequent 3- and 4- word lexical bundles in the 2019/2020 Traveler series KSA edition are as frequent in the output of native English users as found in BNC. Thus, an n-gram analysis of the two corpora was conducted, and the resulting raw frequencies were compared using a chi-square test for association. The second question aims to investigate the extent to which the most frequent 3- and 4- word lexical bundles in the Traveler textbooks are structurally and functionally similar to those used in BNC. To answer this question, structural and functional patterns were manually analyzed in the two corpora. It is expected that the two corpora would be similar as the examined textbooks are designed to represent communicative language like that found in BNC. The following two subsections present findings separately for each research question.

**Are the Most Frequent Three and Four-word Lexical Bundles in the Textbooks used as Frequently by Native English Speakers?**

To answer the first question, a quick look at Table 6 shows that the distribution of the most frequent three- and four-word bundles differs in the two corpora. The number of occurrences of three-word lexical bundles in the BNC2014 (N = 3083) is double of that used in the textbook corpus (N = 1556). Similarly, the frequency of four-word lexical bundles in the

**Table 3.** Criteria for identifying lexical bundles

Length of the bundle	Frequency	Range
3-word	40 per 290,000	Min 3
4-word	40 per 290,000	Min 3

**Table 4.** Structural classification of common lexical bundles along with examples (adapted from Biber et al. (1999: 996))

Structural pattern	Example
Personal pronoun + lexical verb phrase (+ complement clause)	<i>I don't know what</i>
pronoun/NP(+auxiliary) + copula be (+ (auxiliary+) active verb (+)	<i>It was in the</i> <i>Have a look at</i>
yes-no and wh-question fragment (verb +) wh-clause fragment	<i>Can I have a</i> <i>Know what I mean</i>
noun phrase with post-modifier fragment	<i>The nature of the</i>
preposition+ noun phrase fragment	<i>As a result of</i>
anticipatory it + VP/adjective (+complement-clause)	<i>It is possible to</i>
passive verb + PP fragment (verb +) that-clause fragment	<i>Is based on the</i> <i>Should be noted that</i>
(verb/adjective +) to-clause fragment	<i>Are likely to be</i>

**Table 5.** Functional classification of common lexical bundles along with examples (adapted from (Biber et al. (2004: 384–388))

Discourse function	Example
Stance expressions	<i>I don't know</i>
Discourse organizers	<i>If you look at</i>
Referential expressions	<i>Is one of the</i>
Special conversational functions	<i>What are you doing, I said to him/her, thank you very much</i>

BNC2014 ( $N = 1084$ ) is twice that of the textbook corpus ( $N = 623$ ). A chi-squared test on the raw frequencies of both bundle sizes shows that there is a frequency difference between the two corpora ( $\chi^2(1, N = 6346) = 4.833, p < .02$ ). There are significantly more lexical bundles in the BNC corpus across the two bundle sizes. This suggests that the language found in BNC2014 is largely more formulaic than that of the language contained in the textbooks.

### Are the Most Frequent Three and Four-word Lexical Bundles in the Textbooks Similar in their Structure and Function to the Ones Frequently used by Native English Speakers?

To examine similarities in both corpora, the top ten in each bundle size were compared between the two corpora in terms of their grammatical structure and communicative function. These sequences were analyzed based on Biber et al. 's (1999) structural and Biber et al. 's (2004) functional classification of the lexical bundles. The following two subsections present findings separately for three- and four-word lexical bundles.

#### Comparing the structure and function of three-word lexical bundles in the two corpora

A general observation of Table 7 shows that the two corpora use mostly different lexical bundles. Only two lexical bundles frequently occurred in both corpora ("*I don't*," "*a lot of*"). This suggests that, except for "*I don't*" and "*a lot of*," the majority of bundles in the Table above are frequent

**Table 6.** Total number of the ten most frequent three- and four-word bundles in each corpus

Lexical bundle type	Textbook corpus	BNC2014 corpus
3-word	1556	3083
4-word	623	1084
Total	2179	4167

**Table 7.** The top ten three-word lexical bundles in the two corpora. Lexical bundles in brackets occur in the top ten of both corpora.

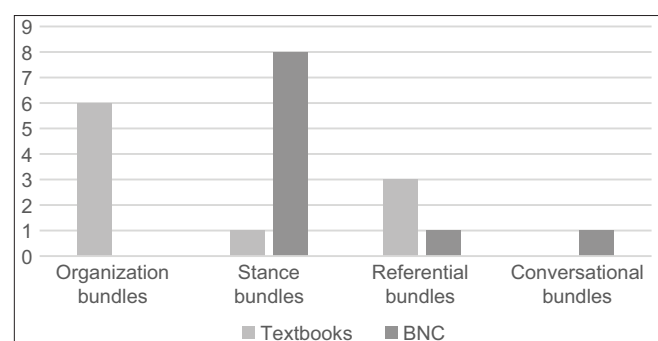
	Textbook corpus	Raw frequency	BNC2014	Raw frequency
1.	Look at the	252	(I don't)	1,036
2.	Do you think	183	Don't know	519
3.	(I don't)	175	Isn't it	248
4.	What do you	153	You don't	205
5.	One of the	148	I didn't	203
6.	(A lot of)	144	I was like	200
7.	In the past	135	I can't	191
8.	You will hear	126	Don't think	167
9.	Go to the	124	(A lot of)	160
10.	Read the text	116	I think it	154
Total		1556		3083

only in one corpus. Not surprisingly, the following analysis reveals that there are structural and functional differences between the top 10 three-word lexical bundles in the two corpora.

The two corpora had different structural patterns for the most common three-word bundles. Structurally, the first personal pronoun + auxiliary (+adverb) pattern dominates the bundles in the BNC2014 column (*I don't*, *you don't*, *I didn't*, *I was like*, *I can't*), while it only occurred once in the textbook corpus (*I don't*). A more preferred structure for the three-word lexical bundles in the textbooks is the lexical verb/verb phrase + (determiner) + noun such as "*look at the*," "*go to the*" and "*read the text*," which is not one of the top bundle structures in BNC2014 data. Another structural difference between the two corpora is observable if we looked at the number of reduced forms in the two columns in Table 7. While the sampled BNC2014 contained eight contracted bundles out of ten, the textbooks had only one.

Due to differences in the structure of the three-word lexical bundles in the two corpora, their functions naturally vary. Figure 1 shows the functional patterns in the two corpora for 3-word bundles. This figure demonstrates that most of the bundles in the BNC corpus (eight out of ten) function to express one's stance and attitude (i.e., *I don't*, *don't know*, *you don't*, *I didn't*, *I was like*, *I can't*, *Don't think*, *I think it*), with only one stance bundle in the textbook corpus (i.e., *I don't*). The conversational nature of the BNC2014 corpus requires the frequent use of bundles that express a speaker's stance towards a proposition and to help the addressee in interpreting that proposition, e.g., "*I think it was really nice just to be at home*" (BNC2014, S2AJ). Unlike the BNC, the textbook corpus frequently presents discourse organization lexical bundles, focusing more on delivering instructions.

Indeed, most of the three-word sequences in the textbook corpus are discourse organization bundles. Specifically, there are six bundles with this function in the textbooks (i.e., *look at the*, *do you think*, *what do you*, *you will hear*, *go to the*, *read the text*). Organization bundles seem to guide students through the material by instructing them to have a look at a specific part in the book (*look at the*), express their thoughts (*do you think*, *what do you*), prepare for the audio to be listened to (*you will hear*), navigate the textbook (*go to the*) and perform an instructional task (*read the text*). According to (Biber et al., 2004), this type of lexical bundles functions to introduce topics and ultimately organize discourse. This



**Figure 1.** Distribution of functions for the three-word bundle in the two corpora

functional analysis of three-word lexical bundles suggests that the Traveller series provide more task instructions to students than demonstrating the use of English in real-life situations, thus probably failing to prepare students to use the language outside the classroom.

Additional insights can be gained when examining the referential bundles used in the two corpora. While the textbook corpus contained three frequent referential bundles, “*a lot of*,” “*one of the*” and “*in the past*,” only one bundle with the same function topped the BNC corpus “*a lot of*.” In other words, two of the referential sequences presented in the textbook corpus seems to be peculiar to it. To understand this distribution, we can investigate the context of the bundle “*in the past*” to know why it is commonly used in the textbooks. Examples (1) and (2) below show two representative contexts of “*in the past*” extracted from the textbook corpus. Looking at these examples, we can hypothesize that the frequent use of this bundle in the textbooks is possibly due to the frequent explanation of the past tense and its markers (Example 1) or discussion of events in the past (Example 2). This suggests that the Traveller series might be more concerned with demonstrating grammar rules than the communicative use of language.

1. We use ‘could’ to express ability *in the past*. (KSA\_TRAVELLER\_1)
2. People travel more now than they did *in the past*. (KSA\_TRAVELLER\_6)

Furthermore, it might be useful as well to consider the referential bundle “*a lot of*” which appeared in both corpora. Although the two corpora contained this bundle with the same function, a closer look suggests that its context of use is different in the textbook and BNC corpora. This difference in use can be noticeable when we examine the top collocate which follows “*a lot of*” in each corpus. The top collocating word for “*a lot of*” in the textbook corpus is “money”, while the word “people” is the most used collocate in the BNC corpus. Two instances of the bundle “*a lot of*” followed by its top collocate in the textbooks (example 3) and the BNC corpus (example 4) are presented below.

3. He came into *a lot of* money when his wealthy uncle died. (KSA\_TRAVELLER\_6)
4. *a lot of* people say I’ve had flu (BNC2014, S3GS)

It should be noted, however, that the word “people” is the fifth most used collocate for this bundle in the textbook corpus. This implies that the textbooks do not completely lack language that is representative of authentic usage as measured by the output of native speakers in BNC2014, but rather that these textbooks should provide more uses of “*a lot of*” that are highly frequent in natural language.

#### **Comparing the structure and function of four-word lexical bundles in the two corpora**

Data on four-word lexical bundles mostly confirm results reported for the three-word bundles. However, unlike the three-word lexical bundles, there are no four-word lexical bundles that appeared in both corpora, as can be seen in Table 8. This suggests that at larger lexical bundle sizes (e.g., four-word), the textbook language gradually becomes less representative of authentic, communicative language.

**Table 8.** The top ten four-word lexical bundles in the two corpora.

No.	Textbook corpus	Raw frequency	BNC2014	Raw frequency
1.	and answer the questions	130	I don’t know	451
2.	What do you think	77	I don’t think	145
3.	The correct form of	61	And I was like	92
4.	The questions that follow	58	don’t know I	65
5.	correct form of the	55	I don’t like	60
6.	answer the questions that	55	Don’t want to	57
7.	Of the words in	51	Don’t know what	56
8.	the text below and	46	But I don’t	56
9.	Form of the words	45	Yeah I don’t	51
10.	Read the text below	45	I can’t remember	51
Total		623		1084

Another observed point in the data is the difference in the type of structure used for four-word lexical bundles in the two corpora. Similar to the three-word bundles’ results, the most dominant structure in the ten most common 4-word bundles in the BNC2014 starts with a first personal pronoun followed by an auxiliary and a lexical verb such as, “*I don’t know*,” “*I don’t think*,” “*I don’t like*,” “*I can’t remember*.” This is a less popular structure in the textbook corpus, appearing only once (i.e., “*you do not need*”). Instead, the textbooks tended to present more bundles that contain a noun phrase with post-modifier fragment including “*(the) correct form of (the)*,” “*the questions that follow*,” “*of the word in*” and “*form of the words*”. Bundles with this structure are more common in academic prose than conversational English (Biber et al., 2004), which suggests that the textbooks represent language typical of academic writing than that of interactive conversations. This seems likely as Table 8 clearly shows that the textbook corpus is much formal with no contracted forms appearing in its column compared to the high frequency of contractions in the BNC2014 corpus.

There is considerable variation as well between the functions of 4-word bundles present in both corpora, replicating the findings for the three-word bundles. All the bundles in the BNC column are stance expressions that imply the speaker’s viewpoint on a given topic, while only one bundle in the textbooks expressed this function, i.e., “*what do you think*”. However, even though both corpora contain stance bundles, their specific context of use is different in the two data sets. Consider the next three examples (5, 6, 7) illustrating the use of the stance bundle presented in the textbooks.

5. *What do you think* the phrase in bold means? (KSA\_TRAVELLER\_6)
6. *What do you think* the verbs in bold mean? (KSA\_TRAVELLER\_4)
7. *What do you think* the dialogue is about? (KSA\_TRAVELLER\_1)

Looking at these examples (5-7), it seems that the main reason for including the stance bundles “*what do you think*” in the textbooks is to engage students with the material rather than to allow them to communicate effectively with people. This finding suggests that even the most commonly used function of lexical bundles in authentic language that of expressing a person’s stance/point of view is distorted in the textbooks. Stance bundles in English textbooks instead became a tool for the authors to deliver instructions.

This, in fact, extends to most of the identified 4-word lexical bundles in the textbook column. All of them seem to be included to fulfill the instructional purposes of a task. The most frequent 4-word bundles learners encounter in their textbooks seem to be limited to the classroom language rather than used in real life. The fact that almost all of the 4-word sequences in the examined textbooks mainly represent the language of teaching makes it difficult for learners to use what is represented in the textbooks beyond the classroom. It is indeed problematic that one of the main functions characterizing authentic bundles that of expressing stance lacks sufficient representation in the textbooks, not allowing learners to grasp some of the communicative features of language through the use of formulaic bundles.

## DISCUSSION

The purpose of comparing lexical bundles in the two corpora is to investigate whether communicative Saudi high school English textbooks represent authentic language use as spoken by native speakers. The examined communicative textbooks are supposed to reflect authentic oral British English language use; thus, the baseline for comparison was the British spoken corpus, the BNC2014. This is important as the communicative status of Saudi high school English textbooks has not been closely examined; hence, a comparative corpus study using a lexical-bundle approach can offer valuable insights on the nature of English materials in Saudi. The comparison of three- and four-word bundles in the two corpora revealed three findings.

### Frequency Differences

First, the analysis of the most frequent 3- and 4- word lexical bundles in the two corpora showed that recurrent word sequences are more frequent in conversational English than in the textbook corpus ( $X^2(1, N = 6346) = 4.833, p < .02$ ). This observation suggests that students encounter in their textbooks less lexical bundles than expected, providing limited opportunities for learning them. The role of repeated exposure in learning lexical bundles has been established, highlighting the fact that the less a student is exposed to a recurrent sequence, the more time it takes her to learn them (Jeong & Jiang, 2019; Northbrook & Conklin, 2018a). The

finding here is contrary to Northbrook & Conklin’s (2018b) results, which reported that Japanese secondary school English textbooks have more bundles than the reference corpus with most of the identified textbooks’ bundles not representing real English use. It is difficult to explain this difference as the present study used the BNC2014 containing real-time spoken conversations while Northbrook & Conklin (2018b) used the SUBTLEX, which is composed of subtitles of American series and movies. The somewhat different spoken registers in the BNC2014 and the SUBTLEX might be one reason for the opposing reports. Another aspect not shared between the two studies is the number of textbooks examined. The present study examined 6 textbooks (290,053 words); meanwhile, Northbrook & Conklin (2018b) analyzed 18 (152,966 words). In other words, the two studies looked at corpora with two different sizes, which makes it difficult to compare their frequency findings.

### Structural and Functional Differences

Another major result is the considerable variation in the structural and functional patterns of 3- and 4-word bundles in the two corpora. Structurally, the first personal pronoun + auxiliary (+lexical verb) featured in most of the 3- and 4- word bundles in the sampled BNC2014, while it was misrepresented in the textbook corpus. In fact, the textbooks preferred (a) the lexical verb/verb phrase + determiner for 3-word bundles and (b) a noun phrase + a post-modifier fragment for 4-word ones. Another observable structural difference is the high frequency of contracted bundles in the BNC2014 and the low number of contractions in the textbook corpus. The analysis of bundles’ structures in the two corpora suggests that the textbooks are mainly full of grammatical patterns associated with more formal and academic language rather than that of conversational English.

Likewise, functionally, stance bundles dominated the BNC2014 sample across both bundle sizes, whereas the textbooks preferred discourse organization and instructional oriented bundles. Variation in the structure and functions of bundles across the two corpora confirms earlier findings. A difference in lexico-grammatical structures is reported in Northbrook and Conklin (2018b), which observed that the textbooks’ bundles are presented in a limited set of structures. Also, Allan’s (2017) study found that stance bundles described as “informal bundles, and largely related to the management of conversation and extended turns” (370) were missing in the analyzed English self-study textbooks.

### Longer Bundles Diverge More

A final observation is that it appears that the longer the bundle, the more divergent from English patterns it gets. The present study noted that while the shorter 3-word lexical bundles in the textbooks shared some similarities with those present in the BNC2014, the longer 4-word ones do not. This supports Northbrook & Conklin’s (2018b) results, finding that 3-word bundles in Japanese English textbooks conform to those in the reference corpus but deviate at longer lengths including 4-, 5- and 6-word bundles.



## PEDAGOGICAL IMPLICATIONS

The above discussion shows the relatively low frequency as well as the limited structure and functions of lexical bundles in a main English textbooks series used in public Saudi high schools. Material designers should take note of this issue and include more bundles that are representative in both structure and function of the bundles used by native speakers. The process of doing so does not require a complete replacement of all the bundles present in the textbooks. Instead, a more practical approach is possible. Focusing on restructuring the existing bundles can be a good starting point for representing more authentic language as students are more familiar with them. To illustrate, let us consider one way of refining an existing bundle “what do you,” which is one of the most frequent 3-word bundles in the textbook. We could place this bundle in a more meaningful context that is more relevant to students’ linguistic needs if we tweaked with its collocates. A quick collocation analysis through Sketch Engine reveals that the top five verbs following the lexical bundle “what do you” are think, know, mean, notice and say. Two examples below illustrate their use.

8. What do you notice about the underlined words? (KSA\_TRAVELLER\_2)
9. What do you mean he disappeared into thin air? (KSA\_TRAVELLER\_4)

Both (8) and (9) illustrate that the sequence “what do you” offers language learners a good starting point to ask about what has someone observed or meant. Nevertheless, more relevant verbs that are related to students’ everyday activities can be incorporated in “what do you”. For instance, textbook designers may present “what do you” with highly frequent verbs that are relevant to activities usually done by students such as playing videogames or watching movies. That is, we can present in the textbooks more relevant bundles such as “what do you play,” “what do you watch,” and so on, enabling students to use the language in meaningful, everyday contexts.

## CONCLUSION

Little attention has been paid to investigate the communicative status of Saudi high school English textbooks. For this reason, the current study aimed to examine the authenticity of lexical bundles in communicative Saudi high school English textbooks that are promoted as representative of native language use. This examination revealed certain features of the lexical bundles present in the examined textbooks. Specifically, the present study compared the structure and function of three- and four-word lexical bundles found in the Traveler series, a set of six books taught across the three years of high school education in Saudi Arabia, to those used in the British spoken corpus, the BNC2014. This comparison showed that three and four-word lexical bundles are less frequent in the textbooks compared to the reference corpus (the BNC2014) and that there is great variation in the structural and functional patterns of the bundles in the two corpora.

Nevertheless, the present study has several limitations. As was mentioned in the methodology, the present study only

focused on one Saudi high school English textbook series out of three to provide a narrow analysis. Future research should examine lexical bundles in the remaining two series used in public Saudi high schools to confirm the present findings. A further limitation is that this study only analyzed the primary functions of lexical bundles, giving a general idea about its context of use. A pragmatic approach to the study of lexical bundles (e.g., Allan, 2017) in secondary school textbooks involving the analysis of pragmatic functions could add insights and complement our knowledge about the nature of lexical bundles in such textbooks. Another possible area of research is the examination of lexical bundles in English textbooks of secondary/middle schools used in other EFL countries.

## END NOTE

1. The BNC contained only three texts recorded in 2013, hence the small number of texts in this category.

## REFERENCES

- 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.
- Alquraishi, M. A. (2014). *Lexical Bundles in an Advanced INTO CSU Writing Class and Engineering Texts: A Functional Analysis* (Unpublished MA thesis). Colorado State University.
- Biber, D. (2006a). *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Biber, D. (2006b). *University Language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing Company.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Chen, L. (2010). An Investigation of Lexical Bundles in ESP Textbooks and Electrical Engineering Introductory Textbooks. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 107–125). New York, NY: Continuum.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56–71.
- Coxhead, A., Yen Dang, T. N., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30, 66–78.
- Gilmore, A. (2011). “I prefer not text”: Developing Japanese learners’ communicative competence with authentic materials. *Language Learning*, 61(3), 786–819.

- Gouverneur, C. (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: a corpus-driven approach to EFL textbook analysis. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 223–243).
- Gries, S. T. (2010). Useful statistics for corpus linguistics. In A. Sanchez & M. Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269–291). Frankfurt: Peter Lang.
- Jeong, H., & Jiang, N. (2019). Representation and processing of lexical bundles: Evidence from word monitoring. *System*, 80, 188–198.
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344.
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. New York, NY: Cambridge University Press.
- Mitchell, H. Q., & Malkogianni, M. (2019). *Traveller 1 KSA Edition Teacher's Manual*. MM publications.
- Nekrasova-Beker, T., & Becker, A. (2019). Lexical bundles in university course materials: From academic English to pathway to mainstream engineering. *International Journal of Corpus Linguistics*, 24(2), 143–168.
- Northbrook, J., & Conklin, K. (2018a). Is what you put in what you get out?—Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, 40(5), 816–833.
- Northbrook, J., & Conklin, K. (2018b). “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.
- Soars, L., & Soars, J. (2009). *New Headway Plus: Upper-intermediate. Student's Book*. Oxford, UK: Oxford University Press.
- Spencer, D. (2016). *Gateway. Student's Book*. (2nd, Ed.). Macmillan Education.
- Widodo, H. P. (2012). The Use of Complaint Letters as an Authentic Source of Input for an Interactive Task in Second Language Learning. *Electronic Journal of Foreign Language Teaching*, 9(2), 245–258.
- Wood, D. (2010). Lexical Clusters in an EAP Textbook Corpus. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 88–106). New York, NY: Continuum.
- Wood, D., & 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.
- Yoo, S. (2013). *Hypothetical would-clauses in Korean EFL textbooks: An analysis based on a corpus study and focus on form approach (Unpublished master's thesis)*. Portland State University.