



Retaliated With Tariffs on: A Corpus Analysis of Lexical Bundles in TED Talks and BBC News on Global Business Issues*

Dohwan Shin**

Hallym University

Yu Kyoung Shin

Hallym University

ARTICLE INFO

Received: 4 March 2020

Revised: 21 April 2020

Accepted: 11 May 2020

Examples in: English

Applicable Languages: English

Applicable Levels: Tertiary

KEYWORDS

*lexical bundles/BBC news/
TED talks/academic English/
global business/
말뭉치/BBC 뉴스/
TED 강연/학술 영어/
글로벌 비즈니스*

ABSTRACT

Shin, Dohwan, & Shin, Yu Kyoung. (2020). *Retaliated with tariffs on: A corpus analysis of lexical bundles in TED talks and BBC news on global business issues. Modern English Education, 21(2), 71-84.*

While several studies have compared lexical bundle (LB) usage across academic written texts, few have controlled for semantic context, which inevitably affects the usage of bundles. This study examines LBs in the texts of BBC news articles and TED talks in one semantic domain (i.e., global business). It first compares the use of LBs identified in BBC written news articles and TED talks, and then further explores how the LBs are embedded in context. The results show significantly different types of discourse functions and structures to be characteristic of each corpus. The BBC news corpus shows written features including referential expression bundles and phrasal bundles embedded in multiple PPs, while the TED talks rely heavily on features of spoken genres such as stance expression bundles and clausal bundles. Yet both corpora share several features typical of academic written genres. In addition, the findings show that the structures in which the LBs are embedded serve different syntactic roles in each corpus. The finding of differences in spoken and written discourse conventions in the shared semantic context of global business has pedagogical applications for business-related courses at the university level as well as English for Academic Purposes courses.

I. INTRODUCTION

Formulaic language – recurrent word sequences – comprise a large portion of natural language, both spoken and written (e.g., Hyland, 2012; Sinclair, 1991; Wray, 2002). An increasing number of studies on formulaic language

are taking data-driven and frequency-based approaches to identifying a special type of formulaic sequence, *lexical bundles* (LBs). LBs are groups of three or more words that frequently recur in a register (Biber, Johansson, Leech, Conrad, & Finegan, 1999). For instance, specific LBs are characteristic of academic prose, and the appropriate use

* This research was supported by Hallym University Research Fund (HRF-201912-011).

** First author: Dohwan Shin, Corresponding author: Yu Kyoung Shin

Dohwan Shin (Undergraduate Student)
School of Global Studies, Hallym University, 1 Hallymdaehak-gil, Chuncheon, 24252, Korea
Tel: (033) 248-1880 / Email: jamesthebest@naver.com

Yu Kyoung Shin (Professor)
School of Global Studies, Hallym University, 1 Hallymdaehak-gil, Chuncheon, 24252, Korea
Tel: (033) 248-2486 / Email: yshin@hallym.ac.kr

of LBs in this register, as in others, is essential to academic fluency (Chen & Baker, 2016; Hyland, 2012; Pérez-Llantada, 2014; Simpson-Vlach & Ellis, 2010). Being able to employ bundles competently in context helps writers to produce coherent texts by meeting genre expectations (e.g., Hyland, 2012; E. Park, 2019; Salazar, 2014; Y. K. Shin & Y. Kim, 2017).

Many studies have employed lexical bundles as a comparative tool for examining the academic writing of students of different language groups and different English proficiency levels (e.g., Ädel & Erman, 2012; Hyland, 2008; Salazar, 2014). While this literature offers some useful insights, many questions remain. For one, the findings reported in the literature centre on the use of lexical bundles across academic written genres, with the vast majority of studies focusing on published research articles (e.g., Hyland, 2008; E-S. Kim, 2019; Pan, Reppen, & Biber, 2016; Pérez-Llantada, 2014) and/or undergraduate writing (e.g., Ädel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2016; J. Kim, 2018; Y. K. Shin, 2019; C. Yoon & J.-M. Choi, 2015). There exist relatively few studies on spoken texts (Reppen & Chen, 2019; Wang, 2017), and even fewer that compare texts across different genres (e.g., Biber & Barbieri, 2007).

Among the comparative research of LBs across genres, Biber and Barbieri (2007) examined LBs found in a wide variety of spoken and written genres, such as class management talk and written course syllabi, showing marked differences between the two genres. These differences may be blurred, however, due to the confounding effects of semantic domain (e.g., Salazar, 2014; Y. K. Shin, 2019). While it is evident that *semantic domains* – specific types of contexts – influence the choice and usage of LBs (Cortes & Hardy, 2013), it is unusual for comparative studies to attempt to match domains. Hence, previous studies' findings may be related to genre differences, as the authors claim, but they may also be due to semantic domain differences.

In addition, as noted, LBs are determined solely on the basis of frequency, and are thus often incomplete in structure, being embedded in other structures. Nevertheless, previous studies have focused on their internal structure and largely ignored their context. A few recent studies comparing academic writing produced by different language groups have suggested that even the same bundle can play different *syntactic roles* (i.e., relations to larger structures), and have demonstrated syntactically distinct patterns of LB use by language groups (e.g., Y. K. Shin, 2018a, 2018b). The existing studies, however, have focused on their analyses to academic written genres; little research has considered spoken genres. In addition, the prior research has mostly focused on English learners' deviance from typical (i.e., native) patterns for academic written genres. (See "Syntactic Roles of LBs" in the literature review.)

In order to address some of these gaps, the present study compares the use of LBs in different genres in the same semantic domain. To do so, the study uses two corpora: One is a corpus of TED talks, which are spoken pre-

sentations by experts geared toward a general audience, and the other is of BBC news articles. A majority of the prior research on LBs centres on written genres, and is mostly limited to published research articles and/or college student writings (e.g., Ädel & Erman, 2012; Chen & Baker, 2016; Hyland, 2008; Pan et al., 2016; Qin, 2014; Y. K. Shin, 2019; Wei & Lei, 2011). The present study is the first attempt to examine LBs by comparing new sets of academic-register data (i.e., TED talks and BBC news articles). The authors expect the findings to have direct pedagogical implications in regard to the usage of LBs in terms of genre differences. Furthermore, this study targets one specific semantic domain (i.e., global business). When teaching courses related to global business, one of the authors has witnessed students' difficulties with academic English, and noted the lack of authentic teaching materials.

This study asks three questions. The study first identifies all LBs of three or more words in the two corpora, and then presents a traditional functional and structural analysis of these LBs, following the approach of existing research. Taking a further step, it then investigates the syntactic roles of the bundles in each genre.

- 1) What are the most frequent three-or-more-word lexical bundles identified in the TED and the BBC corpora?
- 2) To what extent do the bundles' discourse functions and internal structures differ in the TED and the BBC corpora?
- 3) To what extent do the bundles' syntactic roles differ in the TED and the BBC corpora?

II. LITERATURE REVIEW

1. Functions and Structures of LBs

LBs are simply the most frequent recurrent multiword sequences in a register, "regardless of their idiomaticity, and regardless of their structural status" (Biber et al., 1999, p. 990). Yet bundles are not merely sequences of individual words, as they have pragmatic functions in discourse (e.g., Biber & Barbieri, 2007).

Several researchers (e.g., Chen & Baker, 2016; Qin, 2014; Salazar, 2014) have investigated the discourse functions of bundles, which vary according to genre. Biber, Conrad, and Cortes (2004), for instance, compared the bundles identified in two types of spoken genres (university classroom teaching and face-to-face conversation) and one written genre (textbooks), categorizing them into three main discourse categories: stance expressions (e.g., *I thought it was*), discourse organizers (e.g., *going to talk about*), and referential expressions (e.g., *the rest of the*). They found that the specific lexical bundles deployed differ greatly depending on genre. Academic writing employs more referential bundles; conversation employs more stance expression bundles. Classroom teachers reg-

ularly use all three types of LBs; the study found about twice as many different bundles in classroom teaching talk as in conversation, and about four times as many as in textbooks. According to the study, stance bundles were extremely common in both classroom teaching and face-to-face conversation. Discourse organizing bundles were most common in classroom teaching, and also frequent, if somewhat less common, in conversation. In short, stance and discourse organizing bundles were used more in speaking. On the other hand, referential bundles appeared frequently in the written academic texts. Overall, the spoken genres showed more use of bundles than the written genre.

Prior research has also examined the structure of lexical bundles. Biber et al. (1999), for example, grouped bundles into two structural types, which differ according to genre: phrasal LBs (i.e., NP- and PP-based bundles) and clausal LBs (i.e., VP-based bundles). These studies showed that conversation uses more clausal bundles including verb phrase fragments (e.g., *I mean you know*) and dependent clause fragments (e.g., *that there is a*). Academic written genres, on the other hand, include more phrasal bundles, as in noun phrases (e.g., *the end of the*) or prepositional phrases (e.g., *of the things that*). Similarly, Biber and Gray (2010), albeit not focusing on LBs, suggested that while spoken genres often include *elaborated* structures with subordinating clauses, written genres are more likely to have *compressed* structures that combine postnominal modifiers in multiple nominal phrases.

As noted, many previous studies have examined either very formal academic writing – mostly published research articles – and/or very casual face-to-face conversation. Consequently, their findings indicate rather extreme differences in the structures and functions of LBs. But we know little about the usage of bundles between these two extremes of formality. The present study addresses this gap by selecting data from this middle ground, but, for this reason, the findings may not correspond to what has been previously reported. In particular, we may observe features of LBs specific to the news genre and the spoken presentation genre. Furthermore, the sets of LBs identified in the global business discourse we analyse may well include sequences unique to that specific domain. The findings may thus reveal new insights on English language usage across academic genres.

2. Syntactic Roles of LBs

As noted earlier, many of the previous studies have focused on LBs' internal structures (i.e., VP-, NP-, and PP-based bundles), isolated from their context. Recent studies, however, have pointed out that because lexical bundles are solely based on frequency, and thus incomplete in structure, they generally occur as constituents of larger structures (e.g., Li, Franken, & Wu, 2018; Y. K. Shin, 2018a; Y. K. Shin et al., 2019). These studies have shown that the same bundle can play different syntactic roles according to where it is embedded in a sentence.

One study found divergent patterns in the use of the same sets of LBs between native and nonnative university students (Y. K. Shin, 2018b). For instance, the nonnative student writers often used NP-based bundles in the roles of subject or subject-predicative (about 70% together), while the native student writers used such bundles in these roles much more rarely (about 20%), instead using the same NP-based bundles more often as objects. The study attributed the difference to the learners' overuse of copula *be*-verbs and/or tendency to place formulaic language in the initial position of sentences (Li et al., 2018). The same study also found that the native writers, but not the nonnative writers, frequently used NP-based bundles embedded in *of*-phrases as postmodifiers, which is a typical feature of academic prose. In addition, the study showed that both groups used VP-based bundles mostly as adverbial clauses and complement clauses (about 60%), which is characteristic of interpersonal spoken genres (Biber, Gray, & Poonpon, 2011) and indicative of novice academic writers.

These recent studies have examined the positions of LBs in written but not spoken contexts. Hence, the current study will extend this line of research by comparing TED talks and BBC articles controlled for semantic domain. Its findings should provide a more concrete picture of how LBs are positioned in context in different genres.

III. METHOD

1. Corpora

This study used two corpora on the topic of global issues related to international trade and business. The BBC corpus was collected from the British Broadcasting Corporation (BBC) website (<https://www.bbc.com>). Two example titles of the news are *Japanese car sales sink in South Korea amid trade rift* (Harrison, 2019, October 4) and *BREXIT: What trade deals has the UK done so far?* (Edgington, 2020, March 2). The TED corpus was collected from talks on global issues related to international trade and business on the TED website (<https://www.ted.com>). TED talks, as mentioned, are talks for general audiences by experts on the topics of their expertise; they are freely available online. The TED talks were first chosen to match the specific topics of the texts in the BBC corpus as closely as possible. Example titles of the talks selected for the corpus include *The real harm of the global arms trade* (Nutt, 2015, November) and *Facebook's role in BREXIT and the threat to democracy* (Cadwalladr, 2019, April). The selected talks were then screened to check if their topics actually did correspond to those of the BBC news articles. Those that were found to greatly differ were excluded from the dataset. The texts selected through these processes from both sources were found to be mostly quite recent – from 2014 to 2020. The final texts chosen for each corpus were then saved as text files and loaded to a concordance software, *AntConc* (Anthony, 2019), in which data can be retrieved and automatically sorted.

As Table 1 shows, the BBC corpus includes 79 texts of 774.7 words on average, with a total of 61,200 words. The TED corpus contains 30 talks, amounting to 61,212 words, with an average text length (2040.4 words) and total number of words very similar to that of the BBC corpus. The texts in the BBC corpus are shorter; it contains more texts than the TED corpus. Because the two corpora each contain approximately 61,200 words, raw rather than normalized frequencies were used.

TABLE 1
Description of the Two Corpora

Corpus	Number of texts	Mean length of texts in words	Total corpus size in words
BBC	79	774.7	61,200
TED	30	2040.4	61,212

2. Identification, Functions, Structures, and Syntactic Roles of LBs

LBs are identified by frequency, and therefore, studies employ frequency cut-offs to find the sequences. Cut-offs in representative literature range from 10 to 40 times per million words for four-word bundles (e.g., Biber et al., 2004; Hyland, 2008; Salazar, 2014). Moreover, each LB must appear in a certain number of different texts in a given register, typically five or more; this step avoids text-based idiosyncratic usages (Biber et al., 2004). Following the previous studies, the present study, with its two corpora of approximately 61,000 words each, employed the cut-off frequency of 10 times per million words, with a range threshold of a minimum of three different texts.

Additionally, previous studies focus on four-word bundles because they have found most shorter bundles to be embedded in four-word bundles. In contrast, however, this study found, first, that many of the three-word bundles were not embedded in longer bundles, and second, that the total number of four-word bundles identified was too small for the analysis. Therefore, to address the first research question, recurrent sequences of three or more words were identified in the two corpora, using *AntConc* (Anthony, 2019). The retrieved bundles were then checked for any overlapping bundles derived from the same longer bundles, which were removed.

To answer the second question, the LBs identified in each corpus were categorized in terms of their discourse functions and structures, following the categorization method used by Biber et al. (2004). Their functional taxonomy contains three main types. According to Biber et al. (2004, p. 384), stance expressions (e.g., *you don't have to*) display “attitudes or assessments of certainty that frame some other proposition”; discourse organizers (e.g., *let's have a look*) “reflect relationships between prior and coming discourse”; referential expressions (e.g., *is one of the*) are used to “identify [an] entity or to single out some particular attribute of [an] entity.” The structural taxon-

omy includes two phrasal structure types (NP-based and PP-based bundles) and one clausal type (VP-based bundles). NP-based bundles can be bundles in noun phrases with an *of*-phrase structure (e.g., *the structure of*) and those with nominal post modifiers (e.g., *the fact that*); PP-based bundles begin with a preposition, followed by a nominal phrase fragment (e.g., *for the first time*). The clausal structure type includes a verb component such as *you think about*.

For the third question, the bundles in each corpus were examined in terms of their syntactic roles, employing the approach described by Y. K. Shin (2018b). For instance, the clausal bundles (e.g., *left the EU*) were first structurally categorized according to whether they were embedded in a main clause or a dependent clause starting with a complementizer or subordinator. Those in the latter category were then subcategorized by the syntactic roles played by the clause such as adverbial as in (1) or the complement of a prepositional phrase as in (2).¹

- (1) “We aren’t able to negotiate any new trade agreements until after we’ve **left the EU**.” (BBC)
- (2) On one level, this is simply down to not having **left the EU**. (BBC)

Similarly, phrasal bundles were examined in terms of their syntactic roles in context. The following examples show various roles of the same NP bundle, *the trade war*, in sentences found in the BBC corpus: subject in (3), subject predicative following a copular *be*-verb in (4), object in (5), complement of a prepositional phrase adverbial in (6), and postnominal modifier in (7).

- (3) **The trade war** that has seen Washington impose tariffs on billions of dollars’ worth of Chinese goods is adding more strain. (BBC)
- (4) Part of the reason is China’s slowdown, but the other reason is increasingly **the trade war**. (BBC)
- (5) The IMF cut its growth forecasts for the global economy for this year and the next, citing **the trade war** as one of the key reasons. (BBC)
- (6) In just the last three months, we’ve seen another wave of data that shows fresh cracks in the global economy due to **the trade war**. (BBC)
- (7) Mr Kuijs says the impact of **the trade war** on Hong Kong and Singapore is “larger than in China itself, even though no one is imposing any tariffs on these countries.” (BBC)

3. Log-likelihood Test

The study then compared the syntactic roles of LBs of the same structure type in the TED corpus and the BBC corpus. In this and the following sections, statistical significance was tested using a log-likelihood calculator (<http://>

¹ Throughout the paper, sources are given in parentheses at the end of the examples; lexical bundles are indicated in bold.

ucrel.lancs.ac.uk/llwizard.html). As a statistical measure, this method is common in corpus-based research (e.g., Bychkovska & Lee, 2017; I. W. Yoo & Y. K. Shin, 2020). Because it allows the comparison of frequency differences between corpora, it enables research to compare different populations' language use (Baker, 2010). In this study, we use the log-likelihood scores to compare LB frequencies in the native corpus and the nonnative corpus, which enables us to identify statistically significant differences in the frequency with which the two language groups use specific LBs. With log-likelihood scores, higher values indicate more significant differences (i.e., 3.84–6.62 is significant at $p < .05$; 6.63–10.82 is significant at $p < .01$; 10.83–15.12 is significant at $p < .001$; above 15.13 is significant at $p < .0001$).

IV. RESULTS

1. Identification of LBs in the BBC and the TED Corpora

To address the first research question, the study identified three- and four-word LBs in each of the two corpora (see Appendix). In the BBC corpus, 77 types of LBs were found; in the TED corpus, 165 types were found. Nine were shared, as shown in Table 2, comprising 11.7% of the BBC bundles and 5.4% of the TED bundles. Thus, 68 LBs were unique to the BBC, and 156 to the TED.

TABLE 2
Shared Bundles in the Two Corpora

LBs identified in NC	LBs in BBC	LBs in TED
<i>around the world</i>	20	90
<i>the impact of</i>	20	10
<i>it is not</i>	16	63
<i>of the world</i>	15	52
<i>the European Union</i>	14	26
<i>it is a</i>	12	20
<i>to work with</i>	10	16
<i>we have to</i>	10	30
<i>in the last</i>	10	20
Total	127	327

2. Functions and Structures of LBs in the Two Corpora

For the second research question, this study examines the discourse functions and grammatical structures of the LBs in the BBC corpus and the TED corpus.

Figure 1 shows distinct differences in the two corpora regarding the three main functional categories: stance expressions, discourse organizers, and referential expressions. In the BBC corpus, bundles of the referential expression type are the most frequent, amounting to over 50% of all bundles, followed by discourse organizer bundles (35.1%) and stance expression bundles (11.7%). In the TED corpus, on the other hand, stance expression and

referential bundles together constitute over 80% of all bundles, with a slightly higher percentage of referential types, while discourse organizer bundles comprise the smallest proportion (15.8%).

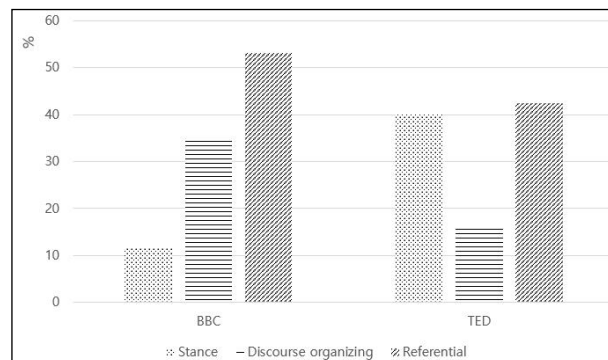


FIGURE 1 Distribution of Main Functional Categories

Table 3 presents the subcategories of the functions of the LBs in each corpus. With respect to the category of stance expressions, the TED corpus has higher portions of both epistemic and attitudinal/modality bundles. As noted earlier, stance expression bundles are considered a feature of spoken registers (e.g., Biber & Barbieri, 2007); this characterization is supported by the high use of stance expression bundles in TED talks compared to BBC news articles. The following examples demonstrate the use of a stance bundle, *we have to*, found in the BBC corpus (8) and in the TED corpus (9). Note that the tokens in this category in the BBC corpus mostly occur in spoken remarks reported in the news articles, as indicated by the use of quote marks in the example.

- (8) “But **we have to** do it with our eyes open and understand the problems as well as the opportunities that we might face.” (BBC)
- (9) Democracy is not guaranteed, and it is not inevitable, and **we have to** fight and **we have to** win and we cannot let these tech companies have this unchecked power. (TED)

Regarding discourse organizers, the BBC corpus includes significantly more topic introductions and elaboration/clarification bundles than the TED corpus. In particular, the written news articles tend to include explicit topic-related LBs, while the spoken talks tend not to. The topic-related bundles in the two corpora display structural differences, as they are usually PP-based in the BBC corpus, and VP-based in the TED corpus. The following examples show topic-introducing LBs in the BBC corpus (10) and the TED corpus (11).

- (10) And in the last decade, the number of inquiries carried out because of fears a child is at risk has increased by 139%, or 120,000. **At the same time**, council budgets have been cut. (BBC)
- (11) **Let’s take a look** at that assumption for just one moment, because you see there has been a boom in

the small-arms trade since the start of the War on Terror. (TED)

TABLE 3
Distribution of Main Functional Subcategories

Categories	Subcategories	Types		Tokens	
		BBC	TED	BBC	TED
Stance expressions	Epistemic**** (e.g., <i>due to the fact</i>)	5.2% (4)	11.5 % (19)	5.4% (71)	10.2% (290)
	Attitudinal/ Modality**** (e.g., <i>the most extraordinary</i>)	6.5% (5)	28.5 % (47)	6.9% (91)	25.6% (724)
Discourse organizers	Topic introduction**** (e.g., <i>when it comes to</i>)	14.3% (11)	10.1% (18)	17.3% (228)	11.2% (318)
	Topic elaboration/ clarification**** (e.g., <i>at the same time</i>)	20.8% (16)	5.4% (9)	16.7% (220)	6.7% (190)
Referential expressions	Identification/ Focus**** (e.g., <i>is one of</i>)	13% (10)	12.7% (21)	15.7% (207)	10.6% (301)
	Framing attributes**** (e.g., <i>in a way that</i>)	4% (3)	0.6% (1)	5.4% (55)	0.5% (13)
	Quantity specification**** (e.g., <i>two largest economies</i>)	18.2% (14)	13.9% (23)	14.3% (188)	13.6% (384)
	Place/Time/ Text-deixis**** (e.g., <i>in South Korea</i>)	18.2% (14)	15.2% (25)	19.6% (258)	20.5% (581)
Conversational function	Politeness**** (<i>thank you very much</i>)	0% (0)	1.2% (2)	0% (0)	1.1% (31)
Total		100% (77)	100% (165)	100% (1318)	100% (2832)

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

The referential expression category is the largest in both corpora (BBC: 53.2%, TED: 42.7%); this finding adds to previous studies' findings that the most frequently occurring bundles in academic written prose tend to be referential expression bundles (e.g., Biber et al., 2004; Chen & Baker, 2016; Y. K. Shin, 2019). Both types of referential expression bundles (i.e., quantifying bundles and place/time/text-deixis bundles) occur often, together comprising over 30% of the LBs in both corpora (BBC: 33.9%, TED: 34.1%).

Place-related bundles also appear at notably high rates in both corpora, which makes sense given the semantic domain (i.e., global issues in business and trade). Examples (12) and (13) demonstrate the use of the same bundle, *in the world*, in the BBC corpus and the TED corpus, respectively.

- (12) India imposes very high tariffs on imports – some of the highest **in the world**, according to US President Trump. (BBC)
- (13) So African agriculture, which is the place of most hunger **in the world**, has actually, fallen precipitously as hunger has risen. (TED)

Figure 2 shows the differences in the use of the three main structural types of LBs in the two corpora: phrasal (i.e., NP- and PP- based bundles) and clausal (i.e., VP-

based bundles). NP- and PP-based bundles together comprise over 70% of the bundles used in the BBC corpus, while VP-based bundles amount to about half of the bundles in the TED corpus.

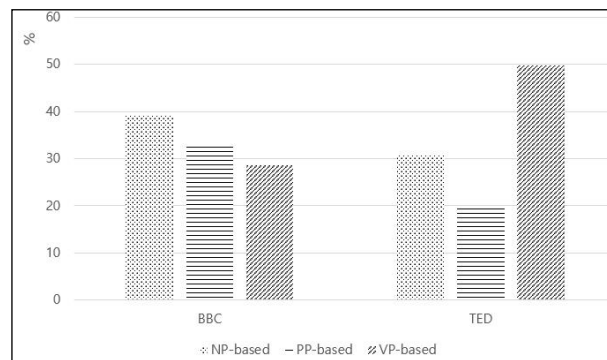


FIGURE 2 Distribution of Main Structural Categories

As already mentioned, phrasal bundles are considered to be typical of the written academic register, as clausal bundles are of the spoken conversational register (e.g., Biber et al., 1999; Chen & Baker, 2010, 2016; Pan et al., 2016; Salazar, 2014; Y. K. Shin, 2019). When Biber and his colleagues (2004) compared academic written prose to face-to-face conversation, they demonstrated that almost 90% of the LBs used in conversation were clausal, while 70% of the LBs in the academic writing were phrasal, whether NP-based or PP-based. In the current study's TED corpus, phrasal bundles, in particular PP-based bundles, appear much less often than they do in the BBC corpus (BBC: 32.5%, TED: 19.7%).

Table 4 shows sub-structural subcategories of the bundles in the two corpora, illustrating specific types of bundles preferred by each group.

Regarding NP-based bundles, the 'other noun phrase' category, which includes nouns with an attributive adjective as a pre-modifier, shows considerable differences, in particular. Many of the bundles in this category are related to the semantic domain used in the study (i.e., global business/trade issues). The following examples show the use of an LB of this type, *the European Union*, from the BBC corpus (14) and from the TED corpus (15).

- (14) **The European Union (EU)** has about 40 free trade deals, covering more than 70 countries.(BBC)
- (15) There's not even any discussions of it joining the European Union. (TED)

As for PP-based bundles, the BBC corpus shows a greater use of those with *of*-phrase fragments (BBC: 10.4%, TED: 4.2%), echoing the literature that reports such bundle usage to be common in academic writing genres (Biber et al., 2004; Biber & Barbieri, 2007; Salazar, 2014). Examples (16) and (17) show the use of *of*-bundles in the BBC corpus and in the TED corpus, respectively. It is noteworthy that in (16), the PP-based bundle, *of the world*, is preceded by multiple PPs (i.e., *with the regular-*

isation/of the UK's membership); compressed structures embedded in multiple PPs functioning as postnominal modifiers are known to be a feature of academic writing genres. In contrast, the PP-based bundle in example (17), from the TED corpus, is preceded by a nominal phrase (i.e., a wonderful example).

- (16) The department did make quick progress with the regularisation of the UK's membership **of the World Trade Organisation**. (BBC)
- (17) A wonderful example **of this is** coming from El Salvador and Los Angeles. (TED)

TABLE 4
Distribution of Main Structural Subcategories

Categories	Subcategories	Types		Tokens	
		BBC	TED	BBC	TED
NP-based	Noun phrase with <i>of</i> -phrase fragment**** (e.g., <i>members of the</i>)	10.4% (8)	14.5% (24)	10.7% (141)	13.6% (386)
	Noun phrase with other postmodifier fragment*** (e.g., <i>trade between the</i>)	9.1% (7)	7.3% (12)	12.1% (159)	5% (141)
	Other noun phrase**** (e.g., <i>the global economy</i>)	19.5% (15)	9.1% (15)	23.7% (312)	8.7% (248)
PP-based	Prepositional phrase with embedded <i>of</i> -phrase**** (e.g., <i>by the end of</i>)	10.4% (8)	4.2% (7)	8.1% (107)	2.8% (80)
	Other prepositional phrase fragment**** (e.g., <i>in the second quarter</i>)	22.1% (17)	15.2% (25)	21% (277)	21.3% (603)
VP-based	Personal pronoun + verb phrase**** (e.g., <i>you will notice that</i>)	0% (0)	10.9% (18)	0% (0)	11.8% (335)
	Existential- <i>there</i> construction**** (e.g., <i>there has been</i>)	3.9% (3)	0% (0)	3.6% (47)	0% (0)
	Anticipatory <i>it</i> + verb phrase/adjective phrase**** (e.g., <i>it is a</i>)	2.6% (2)	0.6% (1)	2.1% (28)	2.2% (63)
	Copula <i>be</i> + noun phrase/adjective phrase**** (e.g., <i>is one of the</i>)	1.3% (1)	17.6% (29)	0.7% (10)	16.9% (479)
	(Verb phrase) + active verb**** (e.g., <i>the US has imposed</i>)	20.8% (16)	20.6% (34)	18% (237)	17.5% (497)
Total	100% (77)	100% (165)	100% (1318)	100% (2832)	

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

Another significant difference between the two corpora is found in the use of VP-based bundles, which comprise the largest category in the TED corpus (49.7%), and the smallest in the BBC corpus (28.6%). In addition, each corpus favours different subcategories of VP-bundles. One notable difference is that VP-based bundles with personal pronouns are frequently used in the TED corpus (10.9%), but never in the BBC corpus. Instances with the personal pronoun in the TED corpus include *you think about*, *I think it's*, *you know what*, *I can tell*, and *you will notice that*, all of

which are indicative of face-to-face conversation, according to the literature. Another difference is that the BBC corpus contains LBs with existential-*there* constructions such as *there has been* and *there will be*, although only a few (3.9%), while the TED corpus includes no LB of this type.

On the other hand, the most frequently occurring category of VP-based bundles is the same in both corpora: active-verb structures (BBC: 20.8%, TED: 20.6%). Examples (18) and (19) show instances of active-verb LBs in the BBC corpus and in the TED corpus respectively.

- (18) Beijing has consistently denied that it engages in unfair trade practices, and has **retaliated with tariffs on** a wide range of US products. (BBC)
- (19) The vast majority, 90 percent, **will be happening in** the South, in cities of the South. (TED)

3. Syntactic Roles of LBs in the BBC and the TED Corpora

This section examines the syntactic roles of the three main types of bundles (i.e., NP-, PP-, and VP-based bundles) in context in the two corpora.

1) NP-based LBs

Table 5 presents the frequency of the several syntactic roles played by NP-based bundles in the two corpora. The order of frequency of the structures in which NP-based LBs are embedded in both the BBC corpus and the TED corpus is the same: subject in active voice is the most frequent (BBC: 48.1%, TED: 29.5%), followed by prepositional phrase functioning as adverbial (BBC: 19.4%, TED: 24.5%) and object (BBC: 13.2%, TED: 17.4%).

TABLE 5
Distribution of Syntactic Roles of NP-based LBs in the BBC and TED Corpora

Syntactic role	BBC	TED
subject in active voice****	248 (48.1%)	199 (29.5%)
subject in passive voice****	4 (0.8%)	11 (1.6%)
subject predicative****	38 (7.4%)	65 (9.6%)
object****	68 (13.2%)	117 (17.4%)
PP as adverbial****	100 (19.4%)	165 (24.5%)
complement of <i>of</i> -phrase****	57 (11.1%)	117 (17.3%)
Total	515 (100%)	674 (100%)
we have to	10	30
in the last	10	20
Total	127	327

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

The results of the log-likelihood calculations, however, display significant differences between the two corpora. First, in the BBC corpus, about half the tokens of NP-based LBs are embedded in subject in active voice structures. The following examples illustrate the syntactic roles of NP-based bundles in the two corpora. The shared bundle, *the impact of*, plays the role of subject in the BBC ex-

amples (subject in active voice in [20], subject in passive voice in [21]), and subject predicative (22) and object (23) in the TED examples.

- (20) Mr Kuijs says **the impact of** the trade war on Hong Kong and Singapore is “larger than in China itself, even though no one is imposing any tariffs on these countries”. (BBC)
- (21) But he was known to be frustrated by **the impact of** no-deal tariffs on the rollover. (BBC)
- (22) The challenge is **the impact of** China’s rise – the discombobulation this will cause the United States and the international order, of which the US has been the principal architect and guardian. (TED)
- (23) Who will feel **the impact of** this rise of China most directly? (TED)

Second, while the role of PP-adverbial ranks second for NP-based LBs in both corpora, the proportion is significantly higher in the TED corpus, at 24.5%, than in the BBC corpus. Examples of the same bundle (i.e., *the impact of*) in PPs functioning as adverbials appear in (24), from the BBC corpus, and in (25), from the TED corpus.

- (24) Companies from Japan to the US were worried about **the impact of** the ban on their supply chains, as they too carried products for Huawei with US parts inside. (BBC)
- (25) I’m going to focus, therefore, on **the impact of** China’s rise, on the US, on the international order and on the order and on the prospects for war and peace. (TED)

In addition, the TED corpus shows a significantly higher use of NP-based bundles embedded in *of*-phrases serving as postnominal modifiers. In examples (26–27), the part underlined is an NP with an *of*-phrase and the LB **in bold** is embedded in the *of*-phrase.

- (26) “The one thing that is affecting business plans is the uncertainty of the US-China trade war, probably more important than the tariffs.” (BBC)
- (27) Events like a nuclear war or a global pandemic that could permanently derail civilization or even lead to the extinction of the human race. (TED)

2) PP-based LBs

Table 6 shows the syntactic roles of PP-based bundles in the two corpora. As can be seen in the table, these bundles appear embedded in structures that serve the roles of adverbial and postnominal modifier to similar extents in both corpora.

TABLE 6
Distribution of Syntactic Roles of PP-based Bundles in the BBC and TED Corpora

Syntactic role	BBC	TED
adverbial	309 (76.7%)	511 (75.4%)
postnominal modifier	94 (23.3%)	167 (24.6%)
Total	403 (100%)	678 (100%)

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

The following examples illustrate the adverbial use of the same PP-based bundle, *around the world*, in the BBC corpus (28) and the TED corpus (29).

- (28) But it did note there are risks, some of which are set out more fully in two reports just out, one on the stability of the financial system and the other focusing specifically on government finances **around the world**. (BBC)
- (29) I mean, you know, it’s not just in our own country, but really we’re exporting the system **around the world**. (TED)

The following examples show the use of another shared PP bundle, *of the world*, as a postnominal modifier in each corpus. The bundle modifies the preceding NPs (underlined): *the rest* in the former example and *the areas* in the latter. Note that two bundles are used in a row in (31).

- (30) But there is little doubt that the ease, comfort, and generosity of the rest of the world was oversold. (BBC)
- (31) And actually, it looks like the areas of the world that are hungry are also the areas of the world that are pretty insecure. (TED)

The study of Y. K. Shin (2018b), which examined the syntactic roles of LBs used by native and nonnative freshman students in argumentative essays, found that their PP-based bundles almost always functioned as adverbials, and rarely as postnominal modifiers. That study differs from the current study in the novice/expert status of the writers/speakers, as well as in the genre and semantic domain of the texts. While the earlier study’s data were essays written by incoming college students who had yet to receive formal instruction in academic prose at the university level, the present study’s data are productions. Therefore, the current finding of relatively more use of PP-based bundles in the role of postmodifiers is in line with the literature that has shown that academic writing relies heavily on phrasal bundles, often containing prepositional phrase(s) as postnominal modifiers (e.g., *with half of the subjects in each age/instructional condition receiving each form*; Biber et al., 2011, p. 31).

3) VP-based LBs

VP-based bundles in the two corpora were first categorized depending on whether they occur in a main clause or a dependent clause. Those in the latter group were subcat-

egorized by their syntactic roles in context. Table 7 shows the syntactic roles of the VP-based bundles occurring in dependent clauses in each corpus, along with the results of the log-likelihood tests.

TABLE 7
Distribution of Syntactic Roles of VP-based Bundles in the BBC and TED Corpora

Syntactic role	BBC	TED
finite complement clause (CC) controlled by V****	3 (5.4%)	53 (12.7%)
finite CC controlled by N****	0 (0%)	33 (7.9%)
which relative clause****	4 (7.1%)	0 (0%)
who relative clause****	0 (0%)	42 (10.1%)
what relative clause****	0 (0%)	70 (16.8%)
when relative clause****	0 (0%)	30 (7.2%)
where relative clause****	4 (7.1%)	22 (5.3%)
how relative clause****	0 (0%)	5 (1.2%)
finite adverbial clause****	17 (30.3%)	60 (14.4%)
nonfinite CC controlled by V****	28 (50%)	48 (11.5%)
nonfinite CC controlled by N****	0 (0%)	28 (6.7%)
nonfinite adverbial clause****	0 (0%)	25 (6%)
Total	56 (100%)	416 (100%)

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$

Overall, the VP-based bundles that occur in dependent clauses show distinctive patterns related to levels of grammatical complexity. Eight of the 12 bundle types were more frequently used in the TED corpus than in the BBC corpus. For one, VP-based LBs in structures taking the role of finite complement clause preceded by both a verb and a noun occur much more often in the TED corpus (BBC: 3 tokens, TED: 53 tokens). This finding corresponds to the prior research that has consistently shown that nonfinite structures are more typical of spoken registers (Biber et al., 1999). The following examples illustrate the use of VP-based bundles: combined with a nonfinite clause (*to*-infinitive) from the BBC news corpus (32) and combined with a finite clause from the TED talk corpus (33).

- (32) Dutch authorities and businesses have already made extensive preparations **to ensure that** any disruption that Brexit might cause is kept to a minimum. (BBC)
- (33) But what **I want to** talk about is what I think is the most fundamental problem. (TED)

In addition, VP-based bundles in *wh*-clauses appear more frequently in the TED corpus. Together, all types of *wh*-clauses comprise over 40% of the VP-based bundles in the TED corpus. For *which*-clauses in particular, however, the BBC corpus contains more than the TED corpus (BBC: 7.1%, TED: 0%). The following examples show the use of an LB in a *which*-clause (34) taken from the BBC corpus, and an LB in a *who*-clause (35) from the TED corpus.

- (34) Picton Yard, which **will be a** mix of homes, shops, and offices, is unique due to all of the elements included in the plans. (BBC)

- (35) We've heard extraordinary activities of civil society groups who **are engaging in** local and global collective action, and this is leading to digital protest and real revolution. (TED)

Prior corpus-based research has argued that while *wh*-relative clauses as postnominal modifiers are strongly associated with academic prose, the use of specific *wh*-clauses differs across spoken and written genres, because *who*-clauses occur only with animate (usually personal/human) head nouns and *which*-clauses usually occur with inanimate head nouns (Biber et al., 1999, pp. 609-611). Such an account corresponds to the findings of the present study; the head nouns in the examples above are *Picton Yard* in (34) and *civil society groups* in (35). A communicative focus on humans, which is common in TED talks, results in the frequent use of relative clauses with the pronoun *who*. In contrast, in the BBC news corpus, where *which*-clauses are the most frequently used *wh*-clause type, animate references might be expected to be less frequent.

V. DISCUSSION AND CONCLUSION

The present study compared the LBs identified in the BBC corpus and the TED corpus in terms of their discourse functions, grammatical structures, and syntactic roles in context. Overall, the results show different usage features in each corpus in the same semantic domain. The BBC corpus displays specific, known characteristics of written genres, including referential expressions and phrasal bundles embedded in PPs as postnominal modifiers. In contrast, the TED corpus shows a heavy reliance on known features of spoken genres such as stance expressions and clausal bundles, often with personal pronouns. At the same time, both corpora share some features typical of academic written genre, in line with the accumulated findings in the literature, including NPs with *of*-phrases and NPs functioning for quantity specification and place/time/text-deixis. In addition, as the current study's corpora were closely matched for semantic domain (e.g., global issues related to international trade and business), many of the most frequent LBs identified in the corpora pertain to that domain, including the nine shared bundles found in both corpora (e.g., *the impact of, the European Union, around the world*).

Furthermore, this study's scrutiny of the syntactic roles of the structures in which frequently occurring LBs are embedded demonstrates that even LBs of the same structure type play different roles across genres. For one, the NP-based bundles in the BBC corpus are predominantly embedded in the initial position and serve as a subject (48.1% of all NP-bundle uses), while those in the TED corpus often function as a constituent in prepositional phrases (41.8% of all PP-based bundles were used in adverbials and complements of *of*-phrases). Regarding VP-based bundles, the TED corpus shows significantly more

use of them in complement clauses and relative clauses (over 60% together), whereas the BBC corpus used VP-based bundles much more often in nonfinite complement clauses and adverbial clauses (over 80% together). On the other hand, the two corpora present similar pictures of the use of PP-based bundles.

The results of the study could inform pedagogical advice for all novice academic writers (native and non-native alike), and thus be useful for EAP practitioners. Instructional materials and activities using the sets of bundles in each genre (see Appendix) could enable students to become familiar with formulaic language in a given genre – an essential skill for communicative fluency in language production and successful language learning (e.g., Arnon & Snider, 2010; Salazar, 2014;). Specifically, sets of bundles identified in the same semantic domain in different genres can be used in teaching materials, combined with ‘input enhancement’ and ‘noticing’ activities. For example, student activities could include highlighting such frequent sequences in provided handouts of BBC news articles and TED talk scripts. Class discussions centring on how LBs are similar/different in terms of grammatical structure (e.g., clausal vs. phrasal) can be followed by activities leading the students to discover specific functional types typical of each genre. These activities would also show students that even the same type of LB can serve different functions according to where it is placed in context. Materials with LBs in different positions highlighted using input enhancement would enable students to *notice* such usage differences (e.g., Y. K. Shin, 2019).

Another pedagogical implication relates to the instruction of different genres within academic prose. As noted earlier, previous studies have shown that English learners tend to produce spoken language features in their academic writing (e.g., Ädel & Erman, 2012; Bychkovska & Lee, 2017; Chen & Baker, 2010, 2016; Y. K. Shin, 2019). These studies suggest that learners must be taught to distinguish genres. The findings of the current study can contribute to our understanding of how to teach different genres in academic contexts.

This study has several limitations. One major limitation is that its findings are based on small corpora (approximately 61,200 words each), which is due to limiting the data to a single semantic domain. The small size of the corpora means this is an exploratory study, and its findings cannot be generalized. One of its strengths, however, is that the exploration it describes demonstrates the value of strictly controlling semantic domains when comparing different genres. Another possible limitation of this study is that it does not address language variety as a potential factor. Some of the differences it found may be related to the fact that the BBC corpus is a corpus of British English whereas the English of the TED corpus is more varied. Future research using much larger corpora of different genres, rigorously controlled for semantic domain and language variety, should provide a more concrete picture of the usage and features of bundles.

On a final note, the current findings illustrate very spe-

cific LB uses in a particular domain of academic prose as used in different academic genres. It is hoped that the present study prompts more LB research across genres, controlled for semantic domain. The findings have the potential to be of great help in teaching EAP courses at the university level as well as second language English programs in general.

REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.
- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67-82.
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- 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., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9, 2-20.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.
- Bychkovska, T., & Lee, J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38-52.
- Cadwalladr, C. (2019, April). *Facebook's role in BREXIT and the threat to democracy* [Video]. TED conferences. Retrieved from https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.
- Chen, Y., & Baker, P. (2016). Investigating critical discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849-880.

- Cortes, V., & Hardy, J. (2013). Analyzing the semantic prosody and semantic preference of lexical bundles. In D. Belcher & G. Nelson (Eds.), *Critical and corpus-based approaches to intercultural rhetoric* (pp. 180-201). Ann Arbor: University of Michigan Press.
- Edgington, T. (2020, March 2). *Brexit: What trade deals has the UK done so far?* BBC. Retrieved from <https://www.bbc.com/news/uk-47213842>
- Harrison, V. (2019, October 4). *Japanese car sales sink in South Korea amid trade rift.* BBC. Retrieved from <https://www.bbc.com/news/business-49929745>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Kim, Eun-Soo. (2019). Lexical bundles in applied linguistics research article English abstracts. *Korean Journal of Applied Linguistics*, 35(2), 3-28.
- Kim, Junghyun. (2018). The effects of learning lexical bundles on the college students in an English composition class. *Korean Journal of Applied Linguistics*, 34(1), 3-27.
- Li, L., Franken, M., & Wu, S. (2018). Chinese postgraduates' explanation of the sources of sentence initial bundles in their thesis writing. *RELC Journal*, 50(1), 37-52.
- Nutt, S. (2015, November). *The real harm of the global arms trade* [Video]. TED conferences. Retrieved from https://www.ted.com/talks/samantha_nutt_the_real_harm_of_the_global_arms_trade
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Park, Eunjeong. (2019). The effectiveness of corpus-aided instruction to improve second language college students' academic writing. *Modern English Education*, 20(3), 15-25.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94.
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231.
- Reppen, R., & Chen, M. (2019). A comparison of lexical bundles in spoken courtroom language across time, registers, and varieties. In T. Fanego & P. Rodriguez-Puente (Eds.), *Corpus-based research on variation in English legal discourse* (pp. 105-122). Amsterdam: John Benjamins.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Philadelphia, PA: John Benjamins.
- Shin, Yu Kyoung. (2018a). *Lexical bundles in argumentative essays by native and nonnative English-speaking novice academic writers* (Unpublished doctoral dissertation). Georgia State University, Atlanta.
- Shin, Yu Kyoung. (2018b). The construction of English lexical bundles by native and nonnative freshman university students. *English Teaching*, 73(3), 113-137.
- Shin, Yu Kyoung. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, 40, 1-14.
- Shin, Yu Kyoung, Choi, Huiseong, Kim, Donghwan, Ko, Seo-jeong, Yoo, Hyemin, Yoo, Hyungjoon, Yoon, Junsu, & Yoo, Isaiah WonHo. (2019). Syntactic complexity of recurrent multiword sequences in the writings of published authors and L1 and L2 English apprentice writers. *The Journal of Asia TEFL*, 16(2), 516-530.
- Shin, Yu Kyoung, & Kim, YouJin. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, 69, 79-91.
- Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Wang, Y. (2017). Lexical bundles in spoken academic ELF: Genre and disciplinary variation. *International Journal of Corpus Linguistics*, 22(2), 187-211.
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal*, 42(2), 155-166.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Yoo, Isaiah WonHo, & Shin, Yu Kyoung. (2020). Determiner use in English quantificational expressions: A corpus-based study. *TESOL Quarterly*, 54(1), 90-117.
- Yoon, Choongil, & Choi, Ji-Myoung. (2015). Lexical bundles in Korean university students' EFL compositions: A comparative study of register and use. *Modern English Education*, 16(3), 47-69.

APPENDIX

Distribution of LBs in the BBC and the TED Corpora

BBC (77 types, 1318 tokens)		TED (164 types, 2832 tokens)	
LBs	Tokens	LBs	Tokens
<i>trade between the</i>	42	<i>around the world</i>	90
<i>a no-deal Brexit</i>	36	<i>it is not</i>	63
<i>one of the</i>	34	<i>this is a</i>	57
<i>the US-China trade war</i>	33	<i>in the world</i>	54
<i>fee pinned post locator</i>	32	<i>one of the</i>	45
<i>read more about these</i>	32	<i>of the world</i>	52
<i>social ns source twitter</i>	30	<i>take a look at</i>	46
<i>the trade war</i>	30	<i>tell you that</i>	36
<i>about these links posted at</i>	29	<i>a lot of</i>	34
<i>as well as</i>	28	<i>are going to be</i>	32
<i>president Donald Trump</i>	28	<i>in war zones</i>	32
<i>that the us</i>	26	<i>we have to</i>	30
<i>the UK government</i>	26	<i>we need to</i>	30
<i>there is a</i>	24	<i>as a result</i>	29
<i>the no-deal tariffs</i>	23	<i>how to do</i>	28
<i>the EU and</i>	21	<i>I want to</i>	27
<i>the impact of</i>	20	<i>in our cities</i>	27
<i>a quick guide to</i>	20	<i>of the most</i>	27
<i>around the world</i>	20	<i>the European Union</i>	26
<i>deal with the</i>	20	<i>in the south</i>	22
<i>imposed tariffs on more</i>	19	<i>some of you</i>	22
<i>in order to</i>	19	<i>the number of</i>	22
<i>in the world</i>	19	<i>you think about</i>	22
<i>key points explained</i>	18	<i>all of the</i>	22
<i>manifesto key points</i>	18	<i>and in the process</i>	21
<i>over the past year</i>	18	<i>are dying in war</i>	21
<i>the end of the</i>	16	<i>are engaging in</i>	21
<i>with the us</i>	16	<i>away from this</i>	21
<i>according to the</i>	16	<i>be able to</i>	19
<i>at the time</i>	16	<i>because you see</i>	20
<i>between the two</i>	16	<i>by percent in</i>	20
<i>government says the</i>	16	<i>every single day</i>	20
<i>it is not</i>	16	<i>I've been working</i>	20
<i>members of the</i>	16	<i>in the global</i>	20
<i>of the country</i>	16	<i>in the last</i>	20
<i>of the trade</i>	15	<i>is where we</i>	20
<i>of the world</i>	15	<i>it is a</i>	20
<i>part of the</i>	15	<i>let's take a</i>	20
<i>two largest economies</i>	15	<i>of small arms</i>	20
<i>tariffs on more than</i>	15	<i>right here in</i>	20
<i>the European Union</i>	14	<i>thank you very much</i>	20
<i>the rest of the</i>	14	<i>the city of</i>	20
<i>there will be</i>	13	<i>the rise of</i>	20
<i>to ensure that</i>	13	<i>the vast majority</i>	20
<i>trade deal with</i>	13	<i>to do this</i>	20
<i>with the EU</i>	13	<i>what's more is</i>	19
<i>world's two largest</i>	13	<i>you know the</i>	19
<i>a number of</i>	13	<i>a little bit</i>	18
<i>billions of dollars 'worth</i>	13	<i>all around the world</i>	18
<i>by the end of</i>	13	<i>all of these</i>	18
<i>exports to the</i>	13	<i>all of us</i>	18
<i>growth in the</i>	12	<i>it did this by</i>	18
<i>impact on the</i>	12	<i>and this is a</i>	18
<i>in South Korea</i>	12	<i>the Atlantic slave trade</i>	18

<i>in the amazon</i>	12	<i>the European union and</i>	18
<i>in the second quarter</i>	12	<i>I'm going to</i>	17
it is a	12	<i>I think it's</i>	17
<i>left the EU</i>	12	<i>is going to be</i>	16
<i>of a no deal</i>	12	<i>is one of the</i>	16
<i>of the European</i>	12	<i>it's not just</i>	16
<i>of the largest</i>	12	<i>most of the</i>	16
<i>of the US</i>	12	<i>of people who</i>	16
<i>on climate change</i>	11	<i>percent of the</i>	16
<i>retaliated with tariffs on</i>	11	<i>that you can</i>	16
<i>says the deal</i>	11	<i>the food system</i>	16
<i>the amazon rainforest</i>	11	<i>the human race</i>	16
<i>the European commission</i>	11	<i>the world by</i>	16
<i>for the first time</i>	11	<i>the world who</i>	16
<i>the global economy</i>	11	<i>this is the</i>	16
<i>the IMF says</i>	10	<i>three quarters of</i>	16
<i>the trade secretary</i>	10	<i>to focus on</i>	16
<i>the US has imposed</i>	10	<i>to try to</i>	16
<i>at the same time</i>	10	to work with	16
to work with	10	<i>we can make</i>	16
we have to	10	<i>what do you</i>	16
<i>will be a</i>	10	<i>you know what</i>	15
in the last	10	<i>you might be</i>	15
		<i>you want to</i>	15
		<i>a billion people</i>	15
		<i>a way to</i>	13
		<i>about how to</i>	13
		<i>again and again</i>	13
		<i>and it is</i>	13
		<i>and now let's</i>	13
		<i>and others have shown</i>	13
		<i>are dying outside of</i>	13
		<i>are living in</i>	13
		<i>arms that were shipped</i>	13
		<i>arms trade treaty</i>	13
		<i>as a result of</i>	13
		<i>as human beings</i>	12
		<i>at the expense of</i>	12
		<i>called it the</i>	12
		<i>can tell you that</i>	12
		<i>circulation in the world</i>	12
		<i>countries that are</i>	12
		<i>cycle of violence</i>	12
		<i>define fragility in our</i>	12
		<i>dropped homicide by percent</i>	12
		<i>each and every single</i>	12
		<i>focus on those</i>	12
		<i>get access to</i>	12
		<i>happening in the</i>	12
		<i>I can tell you</i>	12
		<i>I know what it</i>	12
		<i>I think that</i>	12
		<i>I want to talk</i>	12
		<i>I want you to</i>	12
		<i>in armed conflict</i>	12
		<i>in circulation in the</i>	12
		<i>in fact that</i>	12
		<i>in our own</i>	12
		<i>in terms of</i>	12
		<i>in the world today</i>	12
		<i>in war zones the</i>	12
		<i>is actually at</i>	12

<i>it comes to cities</i>			12
<i>it dropped homicide by</i>			12
<i>might be saying</i>			12
<i>moment in time</i>			12
<i>most of the people</i>			12
<i>most of those</i>			12
<i>need to be</i>			12
<i>not going to</i>			12
<i>now some of you</i>			12
<i>number of people who</i>			12
<i>of all the</i>			12
<i>of doing this work</i>			12
<i>of the people who</i>			11
<i>of this is</i>			11
<i>of you may</i>			11
<i>people who are</i>			11
<i>people who have</i>			11
<i>rule the world</i>			11
<i>solve the problem</i>			11
<i>take away from this</i>			11
<i>take the case of</i>			11
<i>the anniversary of</i>			11
<i>thank you for</i>			11
<i>that were shipped to</i>			10
<i>the arms trade treaty</i>			10
<i>the countries that are</i>			10
<i>the decline in</i>			10
<i>the expense of</i>			10
<i>the hands of</i>			10
<i>the most dangerous</i>			10
<i>the most extraordinary</i>			10
<i>the people who are</i>			10
<i>the price of</i>			10
<i>the reality is</i>			10
<i>the vast majority of</i>			10
<i>the youth bulge</i>			10
<i>to be careful</i>			10
<i>want to talk</i>			10
<i>weapons in circulation</i>			10
<i>what it means</i>			10
<i>what we should</i>			10
<i>when it comes to</i>			10
<i>when you consider that</i>			10
<i>years of doing this</i>			10
<i>you might be saying</i>			10
<i>you will notice that</i>			10
<i>the impact of</i>			10
<i>the Brexit vote</i>			10
Total	1318	Total	2832

Note. The shared bundles (those found in both corpora) are indicated in bold.