The Use and Development of Lexical Bundles in Arab EFL Writing: A Corpus-driven Study

Abdulaziz B Sanosi ⓘ

Prince Sattam bin Abdulaziz University

Background: Lexical Bundles (LBs) have become the focus of many recent corpus linguistics studies. Research has found variable use of LBs in terms of quality and quantity pertaining to different linguistic groups or registers. Still, there is a paucity of research investigating Arab EFL writers' use and development of such a feature.

Purpose: This study investigates the 4-word LBs use and development by Arab EFL learners and expert writers in a corpus of 250000 words regarding their frequency, functions, and structure.

Methods: Two corpora were compiled for Arab learners and scholars. The LB use of both groups was compared to investigate the development of LB use. Further, the Arab corpus was analysed against a native reference corpus extracted from the British Academic Written English (BAWE) corpus to compare LB use across the two corpora.

Findings and Implications: The results imply that there is no noticeable effect of postgraduate education or professional practice on using LBs. The other results, however, are in-line with the previous literature in that native speakers' use of LBs varies in quantity and quality from non-natives'. The findings reveal that stance LBs are more frequent in the native corpus and that they tend to use more VP-based clausal LBs than their non-native counterparts. These findings offer empirical evidence that EFL writing quality is lower despite the current academic writing instruction they receive. They, therefore, indicate the need to foster academic writing instruction programs to include training on using LBs in learners' writing at both Bachelor and postgraduate levels. Also, the results are expected to raise teachers' awareness of how EFL learners use LBs to develop their writing quality and thus to adapt their teaching strategies accordingly. Moreover, Arab scholars are called to reconsider their use of effective writing techniques including LBs for more effective writing.

Keywords: lexical bundles, academic writing, corpus linguistics, Arab EFL, expert writers

INTRODUCTION

One of the core elements of higher education is the writing skill, since students aspire to be identified as proficient writers in their different fields (Kazemi et al., 2014). Scholars who intend to publish in English also aim to produce high-level academic writing which qualifies their research articles for acceptance and publication. Achieving these goals, however, calls for learners and scholars to follow specific patterns and techniques which are believed to produce a solid, comprehensible and cohesive piece of writing. Among these techniques is the use of Lexical Bundles (LBs) which are considered "an important component of fluent linguistic production and a key distinguishing feature of particular modes, registers and genres" (Hyland & Jiang, 2018, p. 383). Moreover, LBs are believed to facilitate both the production and comprehension of discourse (Gil & Caro, 2019). Writers "do not select single

words at a time but choose pre-constructed phrases to express a particular meaning" (Rezoug & Vincent, 2018, p. 48). Following this, "many of the items that have been identified as serving a signalling function in discourse are multi-word units rather than single words" (Nesi & Basturkmen, 2009, p. 24).

Due to the importance of LBs in academic writing, they have recently been studied from many perspectives. Certain studies have explored their use in different genres and registers e.g. (Biber et al., 2004), other studies have compared the use of LBs by native speakers to non-native use e.g. (Ädel & Erman, 2012), while other studies have explored factors which might affect the use of LBs such as L1 transfer e.g. (Dontcheva-Navratilova, 2012). The findings of most previous studies can be generalised in that the use of LBs differs according to the register e.g. (academic writing vs. spoken register) and the the linguistic background of the writer (native or non-native). It is not clear, however, whether scholars and expert writers use LBs better than learners or in a similar way to native speakers. In other words, no sufficient clues are available that further studies and professional practice can develop the use of LBs. Thus more research is needed first to investigate EFL writers' use of LBs and to check if this use is developed overtime; and secondly to compare it to native speakers' use of such a feature. This study investigates the use and development of LBs by Arab EFL learners from these perpectives.

Multi-word Units

The term *Multi-word Units* (MWUs) (Biber et al., 2004; Granger, 2018; Moon, 1998) is an umbrella term encompassing a range of sequences of lexical structures, the meaning of which can be comprehended not only by applying syntactic or semantic conventions, but also by other measures such as their frequency of use, idiomaticity, or pragmatic functions. This general term includes other terms which pertain to different types of MWUs and denote more restricted descriptions. Examples of these terms are *collocations* (Nattinger & DeCarrico, 1992; Salazar, 2014). They are referred to as vocabulary items which tend to co-occur with other specific items, sharing syntactic relations and some degree of semantic opacity (Granger, 2018) such as in *take a break* or *break a record*. These items, as Nattinger and DeCarrico (1992) noted, "should occur at a frequency greater than the chance would predict" (p. 20). Multi-word units can also be represented by *idioms* which refer to invariable expressions which can subsume a large number of multi-word items whether semantically opaque or not (Moon, 1998). Examples of idioms can range from two-words sequences such as *hot potatoes* to a complete sentence such as *Don't put all your eggs in one basket*.

A third example of MWUs, investigated by the current study, is *Lexical Bundles (LBs).* LBs are defined by Biber et al. (1999) as "sequences of word forms that commonly go together in natural discourse" (p. 990). In this respect, LBs are "extended collocations" (Bychkovska & Lee, 2017, p. 39) since they share the feature of co-occurrence of different words. Moreover, while some LBs can represent complete grammatical structures such as *on the other hand,* and *at the same time,* most of them are incomplete grammatical units such as *in the form of* and *as can be seen.* Since LBs are identified "solely on frequency of occurrence and breadth of use" (Hyland & Jiang, 2018, p. 386), corpus linguistics is used to extract LBs based on specific frequency and dispersion criteria. The analysis of lexical bundles has only been made possible by advances in corpus analysis tools. To this end, a considerable body of research has been conducted recently investigating LBs (Allen, 2010)

Researchers tend to investigate the occurrence of three to four-word LBs (Biber et al., 1999; Güngör & Uysal, 2016, and Rezoug & Vincent, 2018). Two-word bundles are not usually investigated, since there are too many of them. On the other hand, five-word or six-word bundles are far less common in different registers. Biber and Barbieri, (2007) stated that the frequency cut-off is normally around 40 times in one-million words, while Hyland and Jiang (2018) specified a lower threshold of 20 times per one-million words. In general, it should be recognised that "the higher the frequency cut-off is, the more representative the lexical bundles are and thus have greater significance for investigation" (Yang, 2017, pp. 58-59). Moreover, recurrent LBs are generally distributed among different texts within a corpus, helping to avoid idiosyncrasies from individual writers/speakers (Chen, 2009). The dispersion criteria of LBs are specified by researchers according to the total number of texts in their corpus (Hyland & Jiang, 2018). However, a common threshold is at least five different texts, as set by Biber et al. (1999).

Functions of Lexical Bundles

Lexical bundles serve important discourse functions in both spoken and written texts (Biber & Barbieri, 2007). Biber (2006) identified three major types of LBs according to the functions they accomplish: (1) stance bundles, (2) discourse organisers, and (3) referential expressions. Stance bundles are used to express attitudes or assessment whether personal or impersonal. The second type of LBs, i.e. discourse organisers, are used to indicate the relationship between different segments of a discourse. They can introduce a new topic or elaborate on a previous one. Referential bundles make direct reference to physical or abstract entities or to the context itself. They may identify or focus on something, specify attributes or express time, place or text reference.

Biber et al. (2004) provided a comprehensive classification of the functional types of LBs in academic prose, as shown in Table 1.

There is a general similarity between discourse organisers and referential bundles. For example, the two LBs are syn-

Table1

Functional Classification of Lexical Bundles in Academic Prose

Type / subtype of Bundle	Example
Stance Bundle	
Epistemic Stance	*the fact that the*
Attitudinal/Modality	*it is important to*
Ability	*it is possible to*
Discourse Organisers	
Topic Introduction	*to look at the*
Topic Elaboration	*on the other hand,*
Referential Bundles	
Identification/Focus	*one of the most*
Specification of Attribute	
Quantity Specification	*the rest of the*
Tangible Framing Attribute	*as a result of*
Intangible Framing Attribute	*in the form of*
Time/place/text reference	
Place Reference	*in the united states*
Time Reference	*at the same time*
Text Dixies	*as shown in figure/table*
Multifunctional Reference	*at the end of*

Note. Adapted from "If you look at…: Lexical bundles in university teaching and textbooks" by D. Biber, S. Conrad, an V. Cortes, 2004, Applied Linguistics, 25(3), 371-405. https://doi.org/10.1093/applin/25.3.371

tactically identical, but serve two different functions according to the above taxonomy. Furthermore, potential confusion between subtypes of bundles could arise even within the same category. For instance, *one of the most* which is used above as an example of identification/focus referential bundle can also serve as a quantity specification referential bundle. Another example is *to look at the* albeit a clear discourse organiser, for a not completely explicit reason is considered a topic introduction bundle rather than a topic elaboration one. These concerns are addressed by the authors who acknowledge that "a single bundle serves different functions depending on the context" (Biber et al., 2004, p. 384). They state that they classify the bundles according to their typical meaning and use of each one of them. For the present study, this is a potential limitation, since the researcher did not manually examine each LB to ascertain what its function was according to the surrounding context, since the context of all the texts are the same and this leads to limited functions across the corpora.

Structure of Lexical Bundles

Lexical bundles do not always represent complete structural units, and they are normally used to bridge phrases or clauses (Gil & Caro, 2019). However, previous studies attempted to categorise them according to their basic gram-

matical constituents. In a detailed classification, Biber, et al. (1999) grouped LBs in academic prose into 12 major categories which are outlined in Table 2:

In a revised classification, Biber (2006) identified three groups of LBs which are: Np/PP-based bundles; VP-based bundles; and Dependant clause bundles. The former classification is adopted by many recent studies e.g. Yang (2017) and Gil and Caro (2019). However, for the current study, a combined structural scheme is adopted, in which the located LBs are firstly classified in the light of Biber's 2006 taxonomy and then for the sake of a finer analysis, their structural category according to Biber et al. (1999) will be reported. For example, the LB *at the end of the* is classified generally as an NP-based bundle and described in detail as NP + *of* phrase fragment.

Previous Studies on Lexical Bundles

Many studies have investigated the use of specific lexical aspects in academic written and, to a lesser extent, spoken discourse. Biber et al. (2004) investigated multi-word sequences in two different university registers: textbooks and university teaching. Comparing the LB used in the two registers, they found that classroom teaching uses more stance and discourse organising bundles than conversation and

Table 2

Structural Categories of LBs in Academic Prose

	Example
noun phrase with *an of*-phrase fragment	*the end of the*
noun phrase with other post-modifier fragments	*the extent to which*
PP with an embedded of-phrase fragment	*as a result of*
other PP (fragments)	*at the same time*
anticipatory *it* + VP/adjective phrase	*it was found that /it is important to*
passive verb + prepositional phrase fragment	*can be found in*
copula *be* + noun phrase/adjective phrase	*is one of the – is similar to that*
(verb phrase +) *that-clause* fragment	*should be noted that*
(verb/adjective +) to-clause fragment	*can be used to / May be able to*
adverbial clause fragment	*as shown in figure*
pronoun/noun phrase + *be* (+ . . .)	*This is not to – there are a number*
other expressions	*as well as the*

Note. Adapted from "Longman grammar of spoken and written English", by D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, (1999), pp. 1014-1024. Copyright 1999 by Pearson Education Limited.)

that more referential bundles are used in academic prose. This point entails a further argument about the quantity of the LBs used in a specific register as this may mean either LB tokens (the number of LBs used in a text or a corpus) or LB type (the number of unique instances of LBs used in a text or a corpus). This distinction is also referred to as bundle density versus bundle diversity (Granger, 2018; Lehmann, 2013). The LB literature suggests that while more LB tokens are used in speech, more LB types are used in writing.

Other studies have compared the use of LBs by non-native writers to their use by native counterparts. Ädel and Erman (2012) investigated the use of 4-word bundles by Swedish university students writing in English in comparison to their native peers. Their results showed that native speakers' use of LBs was more varied and frequent than non-native use. Analogous findings were also reported in many other studies e.g. (Chen, 2009; Amirian et al., 2013; Bychkovska & Lee, 2017; Shin, 2019). Research on native vs. non-native use of LBs was not limited to learners' writing. Other studies (e.g. Salazar, 2014; Güngör & Uysal, 2016; Ucar, 2017) investigated advanced non-native writers' use of LBs and compared it to native scholars. The results, as might be expected, showed different patterns of LBs by non-native writers.

Dontcheva-Navratilova (2012) suggested an effect of language transfer on the use and structure of LBs by non-native speakers of English. These results are supported by Paquot (2013) who investigated French EFL use of LBs. She found a significant L1 effect on their LB use that she traced back to "various properties of French words, including their collocational use, lexico-grammatical patterns, function, discourse conventions, and frequency of use" (p. 391).

Despite this variety of perspectives in addressing LBs, there are still few studies regarding EFL learners' use and development of LBs. For example, few studies have investigated the use of LB by different groups of EFL learners or users who share the same L1 background. An example of these studies is (Johnston, 2017) who compares the LB use by Chinese intermediate learners and professional writers, determining that professionals use LBs differently in terms of form, function, and frequency. Another study was conducted by Zhang et al. (2021), who also found considerable structural and functional differences between Chinese students and expert writers. These studies, however, focused on differences in terms of discipline variation (Johnston, 2017) and analysed the structural and functional differences of the LBs used (Zhang et al., 2021). Investigating the overall LB use in one discipline and by two groups of writers of the same linguistic background and two different proficiency levels is still a research gap. Moreover, research on the use of LBs by the Arab EFL learners is limited and has been directed to analysing the use of LBs by EFL learners in different registers e.g. (Alhusban and Vijayakumar, 2021). Other studies analysed the use of LB in specific areas. For example, Alamri (2021) conducted a genre analysis of the research articles written by Saudi writers to identify LBs associated with patterns of moves in research articles. Conversely, as far as the author is aware, there is no research that has investigated the use and development of Arab EFL writers of LBs. The research gap discussed above motivated the production of the cur-

rent study which investigates LB use by Arab learners and experts through a two-phase analysis. In the first phase, LB use by non-native speakers was investigated to explore the effect of professional experience and postgraduate studies. The use of both non-native groups was then compared to the use of native writers to investigate any variation in frequency, functions or structure of LBs.

To explore the use of Arab EFL learners and scholars of LBs and the effect of professional practice and post graduate studies on such use. Thus, the current research attempts to answer the following questions: (1) What is the difference between the use of lexical bundles by Arab EFL learners and scholars? (2) What is the difference between the use of lexical bundles by Arab EFL writers and native speakers?

METHOD

The current study used the corpus linguistics method, aiming to describe language use through analysing samples of texts written by Arab EFL writers. This aim could be achieved by investigating the frequency distribution of the specific linguistic structure under study i.e. lexical bundles. Studies of such a type normally adopt the corpus method, since this method "aims to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context." (Tognini-Bonelli, 2001, p. 87). Since the present study did not adopt pre-hypothesised LBs as used by the participants but depended on the corpora to inform such findings, it is a corpus-driven study.

The Corpora

This study used two corpora: a non-native writing corpus and a reference one. The non-native corpus is entitled the *Arab EFL Writing Corpus* (AEWC). It incorporates over 250.000 tokens and it is composed of two sub-corpora. The first sub-corpus, *Arab Learner English Corpus* (ALEC), was manually compiled from research articles and reports written by senior EFL students at Prince Sattam ibn Abdulaziz University, Saudi Arabia. Some of these articles were graduation projects, while others were regular writing assignments. The other subcorpus of AEWC, however, was compiled from research articles written by Saudi Arabian scholars who

have published research in international journals in the field of EFL and Applied linguistics. Thus, it was labelled *Arab Scholar English Corpus* (ASEC). The texts were extracted from the Saudi Digital Library, the national online library which incorporates much research in different scientific specialisations. The scholars are PhD holders in the fields of Applied Linguistics and TESOL. While they are believed to possess a high level of fluency in English language, they are non-native speakers. As such it cannot be postulated that they have a native-level competence of the English language. All the topics of the papers were related to Applied Linguistics and TESOL. The reference corpus, on the other hand, consisted of texts taken from the British Academic Written English (BAWE)[1] corpus, selected to match the learner corpus in quantity and quality. Detailed information about the two corpora is displayed in Table3.

As reported in Table 3, the average text-length of ASEC is higher than its counterparts in ALEC and BAWE. This variance influenced the total number of texts in ASEC to become less than those in ALEC and BAWE, and attempts were made to balance the total word-count of the two sub-corpora. No significance variation occurs between other statistics of the corpus.

Procedure

All the texts were converted into *txt* format using Anthony[2] AntFileConverter software. Further, the texts were processed using EmEditor Professional[3] software. Using the Regex feature, all the noise including numbers, mathematical symbols, university and authors' names were deleted. Moreover, extra spaces, line breaks and other formatting characters resulting from the converting process were removed.

Regarding the BAWE sub-corpus, the researcher firstly informed the project owner of his intent to use parts of the corpus and confirming his consent to the conditions[4], then he selected 113 texts according to these criteria: (1) L1: English. (2) Discipline: Arts and Humanities (AH) including English and Linguistics. (3) Genre family: essay; Macro type: simple assignment (4) Courses: BA English and Linguistics and MA TESOL.

---

[1] BAWE was developed at the Universities of Warwick, Reading and Oxford Brookes, under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800). *Source: The University of Warwick.* https://warwick.ac.uk/fac/soc/al/research/collections/bawe/how_to_cite_bawe/. 21st April. 2020

[2] Anthony, L. (2017). AntFileConverter (version 1.2.1) [Computer Software]. Waseda University. http://www.antlab.sci.waseda.ac.jp/

[3] Emurasoft, Inc. (2019). *EmEditor Professional* (Version 19.3.2) [Computer Software]. Filepuma. https://www.filepuma.com/download/emeditor_professional_64bit_19.3.2-23779/

[4] BAWE corpus is available free of charge for research purpose at: https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2539. Certain conditions need to be met before using it.

After building the corpora, the major step of the analysis was to extract the LBs. Using the *Ngram* feature, the researcher utilised *LancsBox* Software[5] to extract all the LBs in the corpora. Then the frequency and dispersion criteria mentioned above were applied, i.e. only 4-word bundles were considered for analysis, since the 4-word bundles represent the optimal structure of LBs (Biber et al., 1999). Moreover, LBs that occurred at least 40 times per million words in at least four texts were elected as data for the study. Since each corpus word count is around 250.000 words, the formula (40 ÷ 1.000.000 × 250.000) was applied and LBs occurring 10 times, or more were investigated. Topic and context-related bundles were excluded. LBs such as *English as a foreign language, Kingdom of Saudi Arabia* were manually excluded, since they might distort the data. Further, overlapping LBs were merged for the same reason, e.g. the two LBs *there are many ways* occurred (18) times and *are many ways to* recurred (14) times. Therefore, the two LBS were merged into one: *[there] are many ways (to)* and the higher frequency (18) was assigned to it. A manual check using the concordance feature of *LancsBox* was conducted, in order to confirm that there are no LBs with *are many ways* followed by a different preposition.

The resultant data was labelled according to the combined structural scheme. First, a general description of the LB was made according to Biber (2006) classification. In this stage each LB was marked as NP-based, VP-based, or PP-based. Further, each LB was marked with a detailed label following Biber et al. (1999) classification. The researchers asked three referees specialised in English language and linguistics to revise the labelling according to the scheme and made a few changes. The labelled LBs according to the above conditions represent the findings of the study and they are presented and discussed below.

RESULTS

In order to respond to the first research question which is: *What is the difference between lexical bundle use by Arab EFL learners and scholars?* the lexical bundles used by each group were compared. The analysis of the AEWC revealed that Arab learners and scholars used LBs in similar ways in terms of the number of LBs used across the two subcorpora. In ALEC, 21 four-word bundles were identified, whereas in ASEC 28 LBs were identified. Table 4 summarises the findings.

Table 4 shows that the amount of use of LBs by learners and scholars is almost identical. Although the 4-word LBs in ASEC outnumber those in ALEC. The overall LB tokens are approximately the same: 376 versus 389 tokens. Moreover, the variation of the LBs used across the two sub-corpora is also approximately the same. This result is inferred from the convergent type/token ratios of the two sub-corpora i.e. 0.06 versus 0.07.

When computing the percentage of LB tokens from the overall token count of the two sub-corpora, the analysis revealed 49 unique LB types in the corpus (21 in the ALEC and

---

5   Brezina, V., Timperley, M., & McEnery, T. (2018). #LancsBox (Version 4.5) [Computer Software]. Lancaster Univeristy. http://corpora.lancs.ac.uk/lancsbox

Table 3

The Corpora Statistics

Corpus	Non-native Corpus (AEWC)			Reference Corpus (BAWE)	TOTAL
Sub-corpus	ALEC	ASEC	Sum		
Token	125040	125608	250648	250278	500926
Type	7576	8258	11919*	19735	35569
Texts	57	33	90	113	203
Average text-length	2193	3806	2785	2166	3104

Note. This is the actual number of types in the whole corpus AEWC. It does not represent the sum of types in ALEC and ASEC as many types are shared by the two corpora.

Table 4

Summary of LBs in AEWC

	ALEC	ASEC
Total Corpus Tokens	125040	125608
4-word LBs (Type)	21	28
LB Tokens	376	389
Type/Token ratio	0.06	0.07

28 in the ASEC), with five LB types are shared by the two sub-corpora. Table 5 displays the shared LBs. All the LBs in the two sub-corpora will be presented in Appendix A.

The 44 types distinct to ALEC or ASEC were compared with those in the reference corpus. With regard to the functional distribution of the LBs in both sub-corpora, it was found that referential bundles were used more than other types in both sub-corpora when compared to the other types. Table 6 summarises the functional distribution of LB types across AEWC.

A final point of the first phase of the analysis was to investigate the structural types of the LBs used across the AEWC. Following Biber, et al. (1999), Table 7 provides a detailed overview of the structural categories of LB types in the two sub-corpora according to the frequency of occurrence.

Table 7 shows that the two corpora included similar LBs in terms of their structural types. Most of the LBs in both corpora are NP-based. Interestingly, the proportion of NP-based LBs in the two corpora are analogous (38 % in the ALEC and 39 % in the ASEC). The second type of LBs used by learners and scholars is the PP-based structural type (29 % in ALEC VS. 32 % in ASEC). The least-used type, however, is the VP-based one (representing 24 % in ALEC and 25 % in ASEC). Other types not classified by Biber et al. (1999) are rarely used in the two corpora.

In answer to the second research question i.e. *What is the difference between lexical bundle use by Arab EFL writers and native speakers?* the use of LBs in the AEWC corpus (after incorporating the results of both ALEC and ASEC and merging the shared LBs) was compared to the reference corpus selected from BAWE. The applied method was to compare the overall LB types in the non-native learner and scholar corpus to the reference corpus. The same criteria of LB size, frequency and dispersion were used to extract LBs types and tokens in the BAWE sub-corpus. The number of LBs generated by these criteria was 68 LB types in the BAWE sub-corpus vis-a-vis 44 LB types in AEWC. In addition to the difference in the number of LBs used by the native speakers, there was also a divergence in the types of LBs used. Table 8 displays a comparison between the LB frequency and functional types in the two corpora:

Because of the difference in the numbers of LBs of the two corpora, it might be more proper to represent the percentage of each LB class in the whole corpus. Figure 1 summarises the main findings in terms of the percentages of each type of LB in the two corpora.

Figure 1 shows that Arab writers used referential bundles more frequently than the British writers. Moreover, they used stance bundles far less than their British counterparts. Despite the difference in the functional types of LBs, there are still shared LBs between the two corpora. These are presented in Table 9. It should be mentioned that the comparison in Table 9 is between AEWC corpus as a whole (44 LBs) and the BAWE sub-corpus (68 LBs). This comparison addressed the 13 LBs that occur simultaneously in BAWE and in either one sub-corpus of AEWC i.e. ALEC or ASEC or in

Table 5

Shared Four-word bundles in ALEC and ASEC

Bundle	ALEC		ASEC	
	Rank	Freq*	Rank	Freq.*
one of the most	1	49	15	13
is one of the	2	37	21	11
on the other hand	4	27	6	16
the results of the	16	12	8	15
as well as the	20	10	24	10

Note. Freq. = Raw frequency.

Table 6

Functional Distribution of LBs in ALEC and ASEC

Type	ALEC		ASEC	
	Number	Percentage	Number	Percentage
Referential	15	71.4 %	24	85.7 %
Discourse Organizer	3	14.3 %	3	10.7 %
Stance	3	14.3 %	1	3.6%

Figure 1

LBs functional type percentage across the two corpora
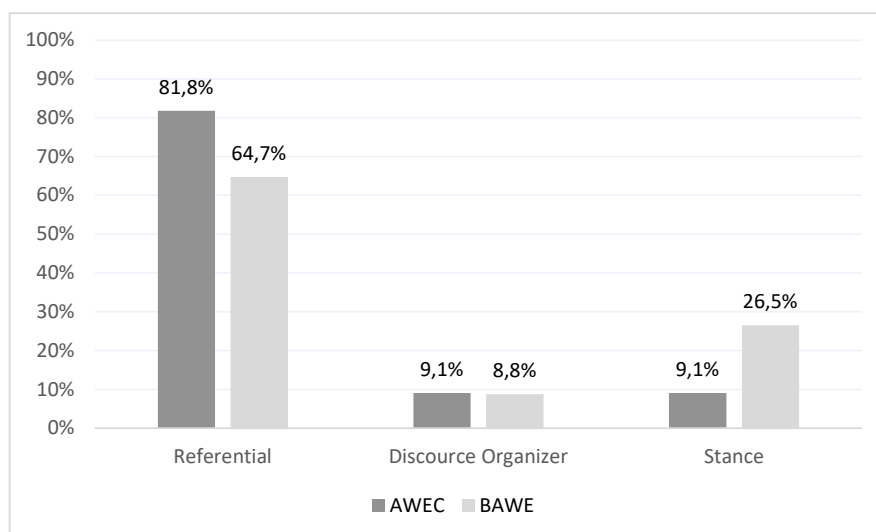
Table7

Structural Types of LBs in ALEC and ASEC

LB structure	ALEC			ASEC		
	Freq.	**example**	**percent**	**Freq.**	**example**	**percent**
NP-Based	8	-	38 %	11	-	39 %
Np + *of* phrase fragment	6	*one of the most*		9	*the finding of the*	
Np + other post modifier fragment	2	an important role in		2	the partici-pants in the	
PP-based	6	-	29 %	9	-	32 %
PP with an embedded *of-* phrase fragment	3	*at the end of*		5	*in the field of*	
Other PP (fragments)	3	*of the most important*		4	with regard to the	
VP-based	5	-	24 %	7	-	25 %
Anticipatory *it* + VP/adjec-tive phrase	1	*it is important to*		1	*it was found that*	
Copula *be* + NP	2	*is one of the*		1	*is one of the*	
That-clause fragment	-	-		3	*that there is a*	
Adverbial clause fragment	-	-		1	*as shown in table*	
To-clause	1	*to deal with the*		1	*to be the most*	
Pronoun + be + NP	1	*[there] are many ways to*		-	-	
Other Expressions	2	*when it comes to*	9 %	1	*as well as the*	4 %
Total	**21**			**28**		

both of them. Interestingly, 4 of the shared LBs in AEWC are also existed in BAWE, while the other 9 LBs are found in BAWE and in one of the sub-corpora of AEWC.

Most of the shared LBs are clearly referential. This is mainly due to the fact that most of the LBs used in AEWC are referential. This suggests that the distinction in functional types between the two corpora is in discourse and stance bundle, as evidently used more in the corpus of the native speakers

i.e. BAWE. Regarding the structural distribution of the LBs in the BAWE sub-corpus, there is a clear difference in structural type preferences, as presented in Table 10.

Table 10 indicates that while British writers tend to use more verb-based LBs, Arab learners and scholars use more NP-based LBs. A detailed classification of the structural types used in the BAWE sub-corpus is presented in Appendix C.

Table8
Frequency and functional types of LBs in AEWC and BAWE

Type	AEWC	BAWE
Referential	36	44
Discourse Organizer	4	6
Stance	4	18
Total	**44**	**68**

Table 9
LBs occurring in both BAWE and AEWC

No.	LB	Structure	Function	Frequency	
				AEWC	BAWE
1	one of the most	NP-based	Referential	62	20
2	is one of the	VP-based	Referential	48	15
3	that there is a	VP-based	Referential	16	16
4	the use of the	NP-based	Referential	13	36
5	as a result of	VP-based	Referential	12	14
6	at the end of	PP-based	Referential	12	40
7	the rest of the	NP-based	Referential	12	23
8	it was found that	VP-based	Referential	11	10
9	at the beginning of	PP-based	Referential	11	26
10	the meaning of the	NP-based	Referential	10	17
11	on the other hand	PP-based	Discourse Organizer	43	57
12	as well as the	other	Discourse Organizer	20	13
13	it is important to	VP-based	Stance	19	24

Table 10
Structural types of LBs in both corpora

Structure type	BAWE	per cent	AEWC	per cent
NP-Based	22	32.4 %	18	40.9 %
VP-Based	27	39.7 %	9	20.5 %
PP-Based	18	26.5 %	15	34.1 %
Other	1	1.5 %	2	4.5 %
Total	**68**	**100**	**44**	**100**

DISCUSSION

The results provided by the first phase of the analysis revealed that Arab EFL learners and scholars employed comparable LBs in terms of both quantity and function. This is an improbable result when compared to the literature which identifies LBs as a feature that marks advanced and fluent writing (Biber & Barbieri, 2007; Allen, 2010; Hyland & Jiang, 2018; and Granger, 2018). Following this, it was expected that scholars would use more LBs in terms of both quantity and quality i.e. it was expected that scholars' use of LBs would be different in token and type when compared to learners. This could be represented, for example, in their use of more structural types known to mark advanced proficient levels e.g. more use of clausal fragments or NP-based bundles. This presumption was postulated in the light of many factors, including scholars' study, experience and practice. Their failure to achieve this, however, suggests two possibilities: (1) experience and post-graduate education have no effect on the use of LBs; and (2) being a non-native English speaker is a strong factor that prevails over other factors governing the use of LB. It can thus be hypothesised that non-native speaker use of LBs is not as effective as that of native speakers. This hypothesis can be traced back to L1 interference as found by Paquot (2013) and Granger (2013). According to this hypothesis, L1 collocational use and lexico-grammatical properties affect the choice and use of foreign language LBs by learners. Another reason to justify the different use of LBs is the use of traditional language teaching methods based on word-level on their language description (Salazar, 2014) rather than on the discourse level. This type of teaching might affect learners use of LBs.

Three of the shared LBs were used much more by the Arab learners than by the Arab scholars i.e. *one of the most (49 vs. 13); is one of the (37 vs. 11),* and *on the other hand (27 vs. 16);* whereas only one bundle was used slightly more by the scholars i.e. *the results of the (15 vs. 12).* This finding supports the conclusion that the scholars in this study did not outperform the learners at any level. Although this finding was unlikely at the beginning, it is in line with a few previous studies, for example, (Gil & Caro, 2019) which found a high level of resemblance between LBs used by L1 Spanish learners and expert writers of English. Whatever the reason, further research is needed to prove or refute this finding since research-related factors might lead to this result.

In terms of functional analysis, it was found that most of the bundles used in both sub-corpora were referential bundles of different types. These LBs are utilised to identify something e.g. *is one of the,* specify quantity e.g. *the majority of the*, or refer to a specific place in the text e.g. *as shown in table*. This result confirms what was suggested by Biber et al. (2004), i.e. that referential bundles are usually the most common type in academic writing. Arab learners, however, express their attitudes to the text more than Arab scholars,

with 3 of the identified LBs being stance bundles. Another finding to note is that two of the discourse organisers are common in both sub-corpora i.e. *on the other hand* and *as well as the.* The highly frequent use of referential bundles across the two sub-corpora indicated the focus of the writers of the two sub-corpora to reflect on their own text, since most of these referential bundles are related to information and data presented earlier in their texts.

The structural types of the LBs identified in the two sub-corpora were also similar in terms of both quantity and quality. Both learners and scholars tended to use NP-based LBs while unclassified bundles are not common in both of them. Moreover, the two sub-corpora used approximately the same structural types of LBs with slight differences. Scholars used two structural types that were absent in ALEC i.e. *that-clause fragment* and *adverbial clause fragment*. It is interesting that both types are in fact of clausal VP-based types. ALEC, on the other hand, made exclusive use of one structural type - phrasal i.e. *Pronoun + NP + be*. Thus, it can be said that the only difference between the corpora is in that Arab EFL scholars use more clausal LBs than learners. This difference, though trivial, suggests more sophisticated language use by the Arab scholars.

When the overall use of the AEWC corpus is compared to the reference corpus of the native speakers, similar results to the previous literature were found. The overall use of LBs in the selected corpus from the BAWE outnumbers the LBs used in the AEWC i.e. 68 versus 44. This finding coincides with the results of many previous studies e.g. (Ädel & Erman, 2012; Bychkovska & Lee, 2017; Esfandiari & Barbary, 2017; Salazar, 2014). The less frequent use of LBs by non-native writers can be justified by the consideration that the optimal use of formulaic expressions, in general, is much more difficult for non-native speakers, since it is more related to intuitive aspects that only native speakers possess. When it comes to academic writing, it was previously found that low usage of LBs is a mark of non-native speakers writing (Güngör & Uysal, 2016), while the use of MWUs including LBs is a feature which marks sophisticated native-like academic writing (Salazar, 2014). This low usage can be traced to the methods of EFL teaching which focus on single-word structure. Focusing on MWUs can make using LBs structures more sophisticated and native-like than using teaching techniques based on single-word structures

The variance in the use of LB across the two corpora is not only in the frequency level, there is an explicit variation in the functional and structural types used across the two corpora. As noted previously by Bychkovska and Lee (2017) and Güngör and Uysal, (2016), native speakers use more stance bundles than non-native speakers. In the BAWE sub-corpus, the writers employed epistemic e.g. *the fact that the,* attitudinal e.g. *it is important to, and* ability stance bundles *e.g. it is possible to.* Considering that stance bundles are used to express a writer's level of certainty about the subject and/or

his/her attitudes towards what s/he is writing, this finding suggests that Arab EFL writers are satisfied with projecting other people's or general viewpoints without reflecting their own or evaluating what they are writing about. This finding is also indicative of the higher quality of LBs used by native speakers, since stance bundles are found to mark higher proficiency levels (Granger, 2018), at least in certain registers such as academic writing. Stance bundles are considered a sign of higher linguistic and thinking skills because they express an assessment of what has been previously written, which is an advanced feature of academic writing. The implication that Arab learners do not utilise this feature and merely present their own ideas is further supported by the finding that more referential bundles are used in AEWC. It is also known that referential bundles are used to identify an entity or its attribute, meaning that writers are more neutral when using such type of LBs.

Another aspect of the difference between the LBs used in the two corpora is represented in the structural types of most of the bundles used in each of the corpora. While native speakers tended to use clausal VP-based bundles, most of the LBs used by the Arab learners and scholars were phrasal NP-based. In fact, around 40% of the LBs detected in the BAWE sub-corpus were clausal VP-based bundles while approximately the same percentage of AEWC LBs were of an NP-based phrasal type. There was no consensus in the structural types preferred by non-native speakers. Therefore, while this finding was provided by certain previous studies such as Salazar (2014) and Bychkovska and Lee(2017), other studies found that native writers use more NP-phrasal LBs than non-native or equal to them. This can be traced back to the nature of the L1 of the writers, since these studies were conducted to investigate texts by writers with different L1s e.g. Persian (Amirian et al., 2013) and Turkish (Güngör & Uysal, 2016). Moreover, VP-clausal bundles are not only used less in AEWC than in BAWE sub-corpus, but it is also the least used type in the corpus, since in addition to NP-based phrasal LBs non-native speakers also use PP-based bundles which are used moderately in the BAWE sub-corpus. Since research on the LBs in Standard Arabic is still rare, no definitive comparison can be stated at this stage. However, this is a potentially rich area of study for further research.

CONCLUSION

This study has aimed to investigate the use of lexical bundles by Arab learners and scholars in their academic writing. In order to achieve such an aim, quantitative and functional analyses were performed on the learner and reference corpora. First, a general comparison was made regarding the use of LBs by learners and scholars. The findings revealed no evident difference between the frequency, function or types of LBs used by learners and scholars. This leads to the hypothesis that there is no significant effect of experience or higher studies on the use of LBs by Arab EFL writ-

ers. However, the second phase of the analysis shows that native writers of English use LBs differently in terms of frequency, type and structure.

Further research is required to explore this finding and to extend the investigation, in order to incorporate a deeper analysis of the probable reasons for this variance. These reasons could be attributed to potential limitations of the current research: i.e. that this study utilised texts of single discipline, genre and context. Further studies that take this into consideration could yield more reliable results to support or refute the present findings.

The results of the present study and proposed further research may lead to better implementation of LB instruction programs which teach Arab EFL learners academic writing skills at the phrasal level, not only at the vocabulary level. These proposed programs may consider the LBs which have been proved to be preferable by scholars and native speakers in the field of academic writing in applied linguistics and/or humanities in general. Moreover, these programs should include the targeted bundles in context-based teaching materials which enhance learners' competence in acquiring and using LBs rather than focusing on the LBs as isolated units. Achieving this, the researcher believes, will lead to better, more robust writing by Arab EFL learners and scholars.

DECLARATION OF COMPETING INTEREST

None declared ∎

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, 2*(31), 81-92. https://doi.org/10.1016/j.esp.2011.08.004

Alamri, B. (2020). A comparative study of Saudi and international journals of applied linguistics: The move–bundle connection approach. *Journal of Language and Education*, *6*(2), 9-30. https://doi.org/10.17323/jle.2020.10531

Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education*, *1*, 105-127.

Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The use of lexical bundles in native and non-native post-graduate writing: The case of applied linguistics MA theses. *Journal of English Language Teaching and Learning, 11*, 1-29.

Biber, D. (2006). *University Language: A Corpus-based study of spoken and written registers.* John Benjamins.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. English for Specific Purposes, *26*(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Pearson Education Limited.

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52. https://doi.org/10.1016/j.jeap.2017.10.008

Chen, Y. H. (2009). *Investigating lexical bundles across learner writing development* [Unpublished doctoral dissertation]. Lancaster University.

Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica, 20*, 56-71. https://doi.org/10.1515/9783484604674.56

Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English, 38*(2), 37-58, https://doi.org/10.5817/BSE2012-2-3

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21–42. https://doi.org/10.1016/j.jeap.2017.09.002

Gil, N. N., & Caro, E. M. (2019). Lexical bundles in learner and expert academic writing. *Bellaterra Journal of Teaching & Learning Language and Literature, 12*(1), 65-90. https://doi.org/10.5565/rev/jtl3.794

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocation and lexical phrases. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Clarendon Press.

Granger, S. (2013). A lexical bundle approach to comparing languages: organisational and stance markers in English and French. *LSB2013 conference - Genre- and Register-related Text and Discourse Features in Multilingual Corpora.* Institut libre Marie Haps.

Granger, S. (2018). Formulaic sequences in learner corpora: Collocations and lexical bundles. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 228-247). Routledge.

Güngör, F., & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and nonnative scholars. *English Language Teaching, 9*(6), 176-188. http://dx.doi.org/10.5539/elt.v9n6p176

Hyland, K., & Jiang, K. F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics, 23*(4), 383-407. https://doi.org/10.1075/ijcl.17080.hyl

Johnston, K. M. (2017). Lexical bundles in applied linguistics and literature writing: A comparison of intermediate English learners and professionals. *Dissertations and Theses, Paper 3482*. Portland State University.

Kazemi, M., Katiraei, S., & Rasekh, A. E. (2014). The impact of teaching lexical bundles on improving iranian EFL students' writing skill. *Procedia - Social and Behavioral Sciences, 98*, 864-869.

Lehmann, M. (2013). The Use of Lexical Bundles in EFL Academic Writing Tasks. In J. Djigunović & M. Krajnović (Eds.), *Empirical studies in English applied linguistics* (pp. 131-141). FF Press.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice.* Cambridge University Press.

Moon, r. (1998). *Fixed expressions and idioms in English.* Clarendon press.

Nattinger, J. R & DeCarrico, J.S. (1992). *Lexical phrases and language teaching*. Oxford University Press.

Nesi, H., & Basturkmen, H. (2009). Lexical bundles and discourse signalling in academic lectures. In J. Flowerdew & M. Mahlberg (Eds.), *Lexical cohesion and corpus Linguistics* (pp. 23-44). John Benjamins.

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics, 18*(3), 391-417. https://doi.org/10.1075/ijcl.18.3.06paq

Rezoug, F., & Vincent, B. (2018). Exploring lexical bundles in the Algerian corpus of engineering. *Arab Journal of Applied Linguistics, 3*(1), 47-77.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching.* John Benjamins.

Shin, Y. K. (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. https://doi.org/10.1016/j.jeap.2019.04.004

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* John Benjamins.

Ucar, S. (2017). A corpus-based study on the use of three-word lexical bundles in the academic writing by native English and Turkish non-native writers. *English Language Teaching, 10*(12), 1-28. https://doi.org/10.5539/elt.v10n12p28

Yang, Y. (2017). Lexical bundles in argumentative and narrative writings by Chinese EFL learners. *International Journal of English Linguistics, 7*(3), 58-69. https://doi.org/10.5539/ijel.v7n3p58

Zhang, S., Yu, H., & Zhang, J. (2021). Understanding the Sustainable Growth of EFL Students' Writing Skills: Differences between Novice and Expert Writers in Their Use of Lexical Bundles in Academic Writing. *Sustainability. 13*(10), 1-17. https://doi.org/10.3390/su13105553

APPENDIX A

Four-word bundles in AEWC:

Sub-corpus	ALEC		ASEC		
Rank	Freq*	Bundle	Rank	Freq.*	Bundle
1	49	one of the most	1	23	in the current study
2	37	is one of the	2	21	the effect of the
3	27	to deal with the	3	20	as shown in table
4	27	on the other hand	4	20	in the use of
5	22	of the most important	5	17	the total number of
6	20	the best way to	6	16	on the other hand
7	19	it is important to	7	16	that there is a
8	18	[there] are many ways	8	15	the results of the
9	14	are many ways to	9	15	the first of these
10	14	at an early age	10	15	with regard to the
11	14	the end of the	11	14	of the present study
12	13	an important role in	12	14	the majority of the
13	12	as a result of	13	14	the participants in the
14	12	at the end of	14	13	in the field of
15	12	the rest of the	15	13	one of the most
16	12	the results of the	16	13	the use of the
17	12	when it comes to	17	12	of the importance of
18	11	at the beginning of	18	12	that most of the
19	11	the development of the	19	12	the findings of the
20	10	as well as the	20	11	a high level of
21	10	the meaning of the	21	11	is one of the
			22	11	it was found that
			23	11	on the use of
			24	10	as well as the
			25	10	in the process of
			26	10	significant difference between the
			27	10	that the use of
			28	10	to be the most
Total Hits	376		Total Hits	389	

Note. Freq raw frequency*

APPENDIX B

List of the LBs in the selected BAWE sub-corpus.

Rank	Freq.*	Range	LBs	Structure	Function
1	57	30	on the other hand	PP Fragment	Discourse Organizer
2	50	20	the way in which	Np +post modifier fragment	Referential
3	40	29	at the end of	pp + embedded *of*-phrase	Referential
4	37	30	the end of the	NP + *of* phrase fragment	Referential
5	36	22	the use of the	NP + *of* phrase fragment	Referential
6	33	15	it is possible to	anticipatory *it* + adjective phrase	Stance
7	26	20	at the beginning of	pp + embedded *of*-phrase	Referential
8	24	20	at the same time	PP fragment	Referential
9	24	19	the beginning of the	NP + *of* phrase fragment	Referential
10	24	14	it is important to	anticipatory it + AdjP	Stance
11	24	17	the fact that the	NP + other post-modifier fragments	Stance
12	23	17	the rest of the	NP + *of* phrase fragment	Referential
13	22	16	in the form of	pp + embedded *of*-phrase	Referential
14	21	11	it could be argued [that]	anticipatory *it* + VP	Stance
15	20	16	that there is no	that clause fragment	Referential
16	20	13	it is interesting that	anticipatory *it* + AdjP	Stance
17	20	14	one of the most	NP + of phrase fragment	Referential
18	19	13	it is said that	anticipatory *it* + VP	Stance
19	17	12	the meaning of the	NP + *of* phrase fragment	Referential
20	17	12	through the use of	pp + embedded *of*-phrase	Referential
21	17	12	way in which the	Np +post modifier fragment	Referential
22	16	14	it seems to be	anticipatory *it* + VP	stance
23	16	13	that there is a	that-clause fragment	Referential
24	15	13	by the use of	pp + embedded *of*-phrase	Referential
25	15	10	is one of the	VP + NP	Referential
26	15	8	the extent to which	Np +post modifier fragment	Referential
27	15	11	the repetition of the	NP + *of* phrase fragment	Referential
28	15	8	the ways in which	Np +post modifier fragment	Referential
29	15	11	to the fact that	PP fragment	Referential
30	14	10	as a result of	pp + embedded *of*-phrase	Referential
31	14	10	that it is a	that-clause fragment	Referential
32	14	11	it is interesting to	anticipatory it + adjective phrase	stance
33	14	11	in this way the	PP fragment	Discourse Organizer

Rank	Freq.*	Range	LBs	Structure	Function
34	14	10	is an example of	copula *be* + noun phrase	Referential
35	14	10	the image of the	NP + *of* phrase fragment	Referential
36	13	8	in contrast to the	PP fragment	Discourse Organizer
37	13	11	as well as the	other expressions	Discourse Organizer
38	13	7	can be seen in	passive verb + PP fragment	Stance
39	13	5	to make sense of	To clause	Stance
40	13	4	to be able to	To clause	Stance
41	12	6	are more likely to	copula be + AdjP	Stance
42	12	6	as can be seen	adverbial clause fragment	Stance
43	12	10	can be seen as	passive verb + PP fragment	Stance
44	12	9	in the case of	pp + embedded *of*-phrase	Referential
45	12	10	it is clear that	anticipatory *it* + adjective phrase	Stance
46	12	10	the nature of the	NP + *of* phrase fragment	Referential
47	12	8	the form of a	NP + of phrase fragment	Referential
48	11	9	it is necessary to	anticipatory *it* + AdjP	Stance
49	11	8	in order to make	PP fragment	Discourse Organizer
50	11	8	to focus on the	to-clause	Referential
51	11	6	with the help of	PP fragment	Referential
52	11	9	be read as a	passive verb + PP fragment	Stance
53	11	5	could be read as	passive verb + PP fragment	Referential
54	11	7	due to the fact	PP fragment	Referential
55	10	9	with the use of	pp + embedded *of*-phrase	Referential
56	10	5	a part of the	NP + *of* phrase fragment	Referential
57	10	9	allows the reader to	VP + to-clause	Discourse Organizer
58	10	10	example of this is	NP + *of* phrase fragment	Referential
59	10	9	in the middle of	pp + embedded *of*-phrase	Referential
60	10	9	it is difficult to	anticipatory *it* + adjective phrase	Referential
61	10	7	it was found that	anticipatory it + VP	Referential
62	10	9	of the text is	PP fragment	Referential
63	10	6	the context of the	NP + *of* phrase fragment	Referential
64	10	8	the idea of the	NP + *of* phrase fragment	Referential
65	10	9	the importance of the	NP + *of* phrase fragment	Stance
66	10	9	the role of the	NP + *of* phrase fragment	Referential
67	10	9	the structure of the	NP + *of* phrase fragment	Referential
68	10	6	is likely to be	copula *be* +AdjP	Stance

Note. Freq raw frequency*

APPENDIX C

Structural Types of LBs in the BAWE sub-corpus*

LB structure	Frequency	example
VP-based		
Anticipatory *it* + VP/adjective phrase	11	it is possible to
Copula *be* + NP/AdjP	4	is an example of
That-clause fragment	3	that it is a
Adverbial clause fragment	1	as can be see
VP + to-clause	1	allows the reader to
To-clause	3	to be able to
passive verb + PP fragment	4	can be seen in
NP-Based		
Np + *of* phrase fragment	17	the end of the
Np + other post modifier fragment	5	the way in which
PP-based		
PP with embedded *of*-phrase fragment	9	at the end of
Other PP (fragments)	9	at the same time
Other Expressions	1	as well as the
Total	**68**	

Note. Structure classification as suggested by (Biber, Johansson, Leech, Conrad, & Finegan, 1999, pp. 1014-1024)