LEXICAL BUNDLES IN JOURNAL ARTICLES ACROSS ACADEMIC DISCIPLINES

Deny Arnos Kwary[1]
Dewantoro Ratri[2]
Almira F. Artha[3]
Universitas Airlangga
kwary@yahoo.com[1], dewa.ratri@gmail.com[2], airafidel@gmail.com[3]

Abstract
This study focuses on the use of lexical bundles (LBs), their structural forms, and their functional classifications in journal articles of four academic disciplines: Health sciences, Life sciences, Physical sciences, and Social sciences. The corpus comprises 2,937,431 words derived from 400 journal articles which were equally distributed in the four disciplines. The results show that Physical sciences feature the most number of lexical bundles, while Health sciences comprise the least. When we pair-up the disciplines, we found that Physical sciences and Social sciences shared the most number of LBs. We also found that there were no LBs shared between Health sciences and Physical sciences, and neither between Health sciences and Social sciences. For the distribution of the structural forms, we found that the prepositional-based and the verb-based bundles were the most frequent forms (each of them accounts for 37.1% of the LBs, making a total of 74.2%). Within the verb-based bundles, the passive form can be found in 12 out of 23 LB types. Finally, for the functional classifications, the number of referential expressions (40 LBs) is a lot higher than those of discourse organizers (12 LBs) and stance expressions (10 LBs). The high frequency of LBs in the referential expressions can be related to the needs to refer to theories, concepts, data and findings of the study.

Keywords: lexical bundle; health sciences; life sciences; physical sciences; social sciences

Learning English as a foreign language can be a strenuous practice, particularly for adult learners, as they may find difficulties on where to start or what to learn. Several researchers have suggested that vocabulary should be the basis in learning foreign languages. Wilkins (1972) stated that without vocabulary, nothing could be conveyed. Harmer (1991) considered vocabulary as "the vital organs and the flesh" of a language (p. 153). In addition, Lewis (1993) suggested a lexical approach because language actually consists of grammaticalized lexis, rather than lexicalized grammar. However, the notion of whether one is required to master all words or lexemes in a dictionary is purely whimsical.

In further developments, a number of researchers have turned their attention into groups of words instead of individual words. The groups of words studied are usually larger than two-word sequences. In studying the groups of words, several researchers refer to them with different terms. Pawley and Syder (1983) use the term "lexicalized sentence stem" (p. 216). Scott (1997) uses the term "cluster" (p. 234). On the other hand, Stubbs (2005) prefers the term "N-grams". Nevertheless, they basically refer to the similar object of study. Biber, Johansson, Leech, and Finegan (1999) define a lexical bundle as a recurrent sequence of words which appears across texts in the same register and help shape distinctiveness of the register. In a further publication, Biber (2006) simply defines lexical bundles as the most frequent recurring sequences of words in a given register.

Several studies have shown that the knowledge of these lexical bundles marks a higher level of proficiency than the knowledge of individual words. Vidacovic and Barker (2010) found that learners with lower proficiency relied more on individual words than conventionalized multi-word sequences or lexical bundles. Likewise, Chen and Baker (2010) also found that the frequency of lexical bundles increased as the language proficiency grew. Consequently, several researchers have ventured into the comparison of the lexical bundles used by L1 and L2 writers. Pang (2010) conducted research on lexical bundles to demonstrate the essential role of lexical bundles in academic writing and to explore strategies in second language (L2) learning. Salazar (2013) compiled lexical bundles from published articles on biology and biochemistry of native and non-native English speakers to compare their use of lexical bundles. Staples, Egbert, Biber, and McClair (2013) related the use of lexical bundles and the proficiency levels of L2-English writers. Pan, Reppen, and Biber (2016) compared the use of lexical bundles by L1-English versus L2-English academic professionals in Telecommunications journals.

Nonetheless, no study has compared the lexical bundles of academic articles from different disciplines. Hence, to fill the void, this study aims at comparing the lexical bundles in the four broad classifications of academic disciplines, i.e. Health Sciences, Life Sciences, Physical Sciences, and Social Sciences. These broad classifications are applied in the Scopus database which is an abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings (https://www.elsevier.com/solutions/scopus).

Previous studies on lexical bundles mostly focus on one discipline. For example, the data of the study conducted by Salazar (2013) were only from biology and biochemistry published articles. Qin (2014) studied the academic papers and published articles in the field of applied linguistics. Grabowski (2015) analysed lexical bundles using the corpus of written English pharmaceutical discourse. Pan et al. (2016) focused on the lexical bundles used in Telecommunications journals. The study made by Pérez-Llantada (2014) actually included research articles from various disciplines, i.e. twelve different disciplines. However, the focus was still on the comparison of the L1 and L2 use of lexical bundles.

This study seeks to fill in the gap of the research of lexical bundles by focusing on the comparison of lexical bundles in the four broad classifications of academic disciplines. Biber (2006) has already mentioned that the lexical bundles occurring and applied in one discipline may be distinctive from those of other disciplines and thus, presumably, they may show distinctive uses of lexical bundles which may vary in terms of forms, structures and functions. Qin (2014) found that in terms of the structural forms of the lexical bundles, non-native graduate writers at the higher levels of study applied more forms of academic writing, such as noun phrases with post-modifier fragments than those of lower levels. Grabowski (2015) found that the discourse functions of the most frequent lexical bundles vary across pharmaceutical text types due to situational contexts, functions and target users of patient information. Several researchers (Biber, Conrad, & Cortes, 2004; Biber, 2006; Hyland, 2008) have also shown that lexical bundles vary in their discourse functions (e.g., expressing stance, discourse organization, or referential meanings).

Considering that functional differences are as important as structural differences in studying lexical bundles, this study aims at comparing both the structures and the functions of lexical bundles across the different academic disciplines.

Based on the explanation given in the Introduction section, the objectives of the present study are as follows: (i) to compare the use of lexical bundles across the different academic disciplines; (ii) to determine the distribution of the structural forms of the lexical bundles across the different academic disciplines; and (iii) to determine the distribution of the functional classifications of the lexical bundles across the different academic disciplines.

METHOD

The corpus of this study is derived from journal articles. The journal articles are classified into four disciplines following the classification found in Scopus which is one of the most extensive databases in the academic community to publish numerous papers in different scientific fields (Chadegani, et al., 2013). The four disciplines stated in Scopus are: (1) Health Sciences, (2) Life Sciences, (3) Physical Sciences and (4) Social Sciences (Elsevier, 2016). Based on statistics, Elsevier is the publisher with the highest coverage of Scopus. There were 10% of nearly 5,000 journals provided in Scopus were published by Elsevier (2016).

There are four criteria set for selecting the articles to be used in building the corpus. The first criterion is that the articles must be open access articles, meaning that the journal articles can be freely accessed and downloaded. Second, only journal articles written in English, as an international language, were selected in this study. The third criterion is that the article must be written by at least one native speaker of English, which can be checked from the names of the authors. The final criterion is that the journals chosen must be the journals with the highest 5 years' impact factors

Since we selected 100 articles for each discipline, the corpus comprises 400 journal articles, with a total of 2,937,431 running words. The details of the number of words for each discipline are shown in Table 1.

Table 1. The size of the corpus

No	Disciplines	Number of articles	Number of words
1	Life Sciences	100	667.055
2	Heath Sciences	100	460.621
3	Physical Sciences	100	884.578
4	Social Sciences	100	925.177
	Total	400	2.937.431

From the corpus data shown in Table 1, we then extracted the lexical bundles using AntConc 3.4.4 (Anthony, 2014). Following Biber (2006), the lexical bundles chosen are those that consist of four

words. In addition, according to Hyland (2008, p. 8) four-word sequences are far more common than five-word strings and offer a clearer range of structures and functions than 3-word bundles. In

selecting the lexical bundles, the frequency and range should also be taken into account. Biber et al. (1999) identified the most common lexical bundles as sequences of words that occurred at least 10 times per million words in the target register and distributed across at least five different texts. In this study, since the corpus is more specialized, the sequence of words is identified as a lexical bundle when it has a minimum frequency of 20, and a minimum range of 20 as well, i.e. it can be found in 20 different texts.

FINDINGS

The use of lexical bundles across the academic disciplines

Based on the method mentioned in the previous section for extracting the lexical bundles from the corpus, we found 62 lexical bundles that fulfil the criteria (i.e. a minimum range and frequency of 20). Table 2 presents the 62 lexical bundles and their distributions in each discipline.

As shown in Table 2, there were no lexical bundles used in all four disciplines, which implies that there are differences in the use of lexical bundles across the different disciplines. Physical sciences feature the most number of lexical bundles, i.e. 43. Social sciences and Life sciences have 27 and 12 lexical bundles, respectively. Health sciences have the least number of lexical bundles, which is only 3. If we take the lexical bundles which can be found in at least three disciplines, we will also see that Health sciences are not included in them. There are six lexical bundles found in all of the three disciplines, i.e. Life sciences, Physical sciences and Social sciences. These six lexical bundles are "in the case of", "in the context of", "in addition to the", "as a result of", "as well as the", and "can be used to".

If we pair-up the disciplines, we will be able to see which disciplines shared the most number of lexical bundles. For Health sciences and Life sciences, the shared lexical bundle is only "has/have been shown to". There are no lexical bundles shared between Health sciences and Physical sciences, and neither is shared between Health sciences and Social sciences. This means that Health sciences share a similarity, though very small, with Life sciences, but none are shared between Physical sciences and Social sciences.

The pair-up of Life sciences and Physical sciences show eight shared lexical bundles, i.e. "as a result of", "as well as the", "can be used to", "in addition to the", "in the absence of", "in the case of", "in the context of", and "in the presence of". The number is slightly bigger than that between Life sciences and Social sciences, i.e. six, namely: "as a result of", "as well as the", "can be used to", "in addition to the", "in the case of", and "in the context of". Finally, the pair-up of Physical sciences and Social sciences show the highest number of shared lexical bundles, i.e. 14. This implies that Physical

sciences are closest to Social sciences in terms of the use of lexical bundles.

The structural forms of the lexical bundles across the academic disciplines

After knowing the use of lexical bundles across the academic disciplines, it will also be interesting to know the distribution of the structural forms of the lexical bundles. The structural forms used in this study are based on Biber et al. (1999), who divided the forms into three main categories: noun-based, preposition-based, and verb-based bundles. The tabulation of the structural forms is presented in Table 3.

The data in Table 3 indicate that in terms of structural forms, lexical bundles in the academic journal articles are mostly prepositional-based and verb-based bundles, i.e. each of them covers 37.1%, making a total of 74.2% for both structural forms. The lowest number of structural forms is the noun-based, which is only approximately a quarter of the total (i.e. 25.81%). These results are different from those shown by Pan et al. (2016) who found that the noun-based form is the most frequent one, i.e. 36.4% (20 LB types out of 55 LB types) in their corpus of telecommunication journals. The prepositional-based and verb-based forms are 32.7% (18 out of 55) and 25.4% (14 out of 55), respectively. The results of the current study are similar to those made by Qin (2014) who analysed applied linguistics journal articles. Qin (2014) found that the prepositional-based form covers 41.46%, the verb-based form includes 33.46%, and the noun-based form only contributes 12.27%. These studies show that in telecommunication journals, the noun-based and the prepositional-based are the most frequent forms; whereas, in applied linguistics journals, the prepositional-based and the verb-based are the most frequent ones. In this current study, which includes journal articles from all major disciplines, the prepositional-based and the verb-based are the most frequent ones.

If we see further into the classifications of the structural forms, the results of this study are more similar to those of Pan et al. (2016) than those of Qin (2014). Within the verb-based form, Pan et al. (2016) found that the passive verb is the most frequent one (7 LB types out of 14 LB types), whereas in Qin (2014) the passive voice only covers 2.23% of the LB types. This means that applied linguistics journals do not incorporate the passive form as often as telecommunication journals do. For the data across different disciplines, as shown in this current study, journal articles still have the tendency to employ the passive form quite often. This study shows that the passive form can be found in 12 out of 23 LB types. This may indicate that journal articles across different major academic disciplines are in favour of using the passive form, because by using a passive form, the sentence focuses on what is done, rather than who does it (Hinkel 2004).

Table 2. The use of lexical bundles across the academic disciplines

No.	Lexical Bundles	Health Sciences	Life Sciences	Physical Sciences	Social Sciences
1	a function of the			√	
2	a wide range of			√	√
3	an increase in the			√	
4	as a result of		√	√	√
5	as shown in figure			√	
6	as well as the		√	√	√
7	at the end of			√	√
8	at the same time				√
9	at the time of		√		
10	can be found in			√	
11	can be used to		√	√	√
12	figure legend the reader			√	
13	figure a and b (fig a and b)			√	
14	for each of the				√
15	for interpretation of the			√	
16	has/have been shown to	√	√		
17	in addition to the		√	√	√
18	in terms of the			√	√
19	in the absence of		√	√	
20	in the case of		√	√	√
21	in the context of		√	√	√
22	in the form of			√	√
23	in the presence of		√	√	
24	in the range of			√	
25	in this case the			√	
26	in this figure legend			√	
27	in this paper we				√
28	interpretation of the references			√	
29	is due to the			√	
30	is referred to the			√	
31	is/are likely to be				√
32	is/are more likely to be				√
33	is/are shown in figure			√	
34	it is clear that			√	
35	it is important to			√	√
36	it is possible that		√		
37	it is possible to		√		
38	legend the reader is			√	
39	of the references to			√	
40	of this study was	√			
41	on the basis of				√
42	on the other hand			√	√
43	one of the most				√
44	referred to the web			√	
45	that there is a				√
46	the effect of the			√	√
47	the extent to which				√
48	the nature of the				√
49	the reader is referred			√	
50	the rest of the				√
51	the result of the				√
52	the shape of the			√	
53	the size of the			√	
54	the web version of			√	
55	to the web version			√	
56	the fact that the			√	√
57	used in this study			√	
58	version of the article			√	
59	was found to be				√
60	web version of this			√	
61	were included in the	√			
62	with respect to the			√	
	Total	3	12	43	27

Table 3. The structural forms of the lexical bundles across the academic disciplines

Structural Forms		Types	% of types	Lexical bundles
Noun-based	Noun phrase with of-phrase fragment	13	20.97%	1. a function of the 2. a wide range of 3. interpretation of the references 4. one of the most 5. the effect of the 6. the nature of the 7. the rest of the 8. the result of the 9. the shape of the 10. the size of the 11. the web version of 12. version of the article 13. web version of this
	Noun phrase with other post-modifier fragment	3	4.84%	14. an increase in the 15. the extent to which 16. the fact that the
	Total	16	25.81%	
Prepositional-based	Prepositional-based with embedded -of phrase fragment	16	25.81%	17. as a result of 18. at the end of 19. at the time of 20. for each of the 21. for interpretation of the 22. in addition to the 23. in terms of the 24. in the absence of 25. in the case of 26. in the context of 27. in the form of 28. in the presence of 29. in the range of 30. in this case the 31. of this study was 32. on the basis of
	Other prepositional phrase segments	7	11.29%	33. at the same time 34. in this figure legend 35. in this paper we 36. of the references to 37. on the other hand 38. to the web version 39. with respect to the
	Total	23	37.10%	
Verb-based	Be+noun phrase/ adjective phrase	3	4.84%	40. is due to the 41. is/are likely to be 42. is/are more likely to be
	Passive verb	12	19.35%	43. as shown in figure 44. can be found in 45. can be used to 46. figure a and b (fig a and b) 47. has/have been shown to 48. is referred to the 49. is/are shown in figure 50. referred to the web 51. the reader is referred 52. used in this study 53. was found to be 54. were included in the
	Verb/adjective+that	2	3.23%	55. it is clear that 56. it is possible that
	Verb/adjective+to	2	3.23%	57. it is important to 58. it is possible to
	Verb phrase with active verb	1	1.61%	59. that there is a
	Adverbial clause	1	1.61%	60. as well as the
	Other expressions	2	3.23%	61. figure legend the reader 62. legend the reader is
	Total	23	37.10%	

The functional classifications of the lexical bundles across the academic disciplines

After discussing the distribution of the structural forms of the LBs, we also determine the distribution of the functional classifications of the LBs across the different academic disciplines. The data in Table 4 presents the tabulation of the LBs into their functional classifications along with their frequency information. The functional classifications are based on Biber (2006) who mentioned a functional taxonomy with three major categories: stance expressions, discourse organizers, and referential expressions. In terms of frequencies, each lexical bundle is categorized into four categories based on the number of occurrences in the corpus. In this case, a single asterisk (*) indicates that the frequency of the lexical bundle ranges from 21 to 30 occurrences, two asterisks (**) indicates 31 to 40 occurrences, three asterisks (***) refers to 41 to 50 occurrences; and finally, four asterisks (****) signifies more than 50 occurrences in the corpus.

Table 4. The functional classifications of the lexical bundles across the academic disciplines

Functional classifications	Health Sciences	Life Sciences	Physical Sciences	Social Sciences
I. STANCE EXPRESSIONS				
A. Epistemic stance – Impersonal				
is/are likely to be				****
is/are more likely to be				****
B. Attitudinal/Modality stance; Ability/effort – Impersonal:				
can be used to		*	**	*
C. Importance – Impersonal:				
it is important to			*	**
was found to be				*
has/have been shown to	*	***		
that there is a				*
of this study was	*			
D. Intention/prediction – Impersonal:				
it is possible that		*		
it is possible to		*		
II. DISCOURSE ORGANIZERS				
A. Topic introduction/focus				
in this paper we				*
B. Topic elaboration/clarification				
at the same time				**
on the other hand			**	****
as well as the		*	**	***
for interpretation of the			*	
in addition to the		*	*	*
is due to the			*	
it is clear that			*	
the effect of the			*	*
the result of the				*
C. Identification/focus				
one of the most				*
the fact that the			*	*
D. Conditions				
n/a				
III. REFERENTIAL EXPRESSIONS				
A. Specification of attributes/ Quantity specification			*	*
a wide range of			*	
an increase in the				*
for each of the			*	
the size of the				*
the rest of the			*	
the shape of the			*	
in the range of				
B. Tangible framing attributes				
a function of the			*	
as a result of		*	*	*
the extent to which				**
C. Intangible framing attributes				
in the case of		*	***	***

in the context of		*	*	***
in this case the			*	
on the basis of				*
the nature of the				*
in the absence of		**	*	
in terms of the			*	**
in the presence of		*	*	
in the form of			**	*
D. Time/place/text/other reference				
Place or institution reference				
is/are shown in figure			****	
as shown in figure			**	
can be found in			**	
figure a and b (fig a and b)			*	
referred to the web			*	
to the web version			*	
the web version of			*	
web version of this			*	
version of the article			*	
figure legend the reader			*	
in this figure legend			*	
used in this study			*	
legend the reader is			*	
the reader is referred			****	
E. Time reference				
at the time of		*		
F. Multi-functional reference				
at the end of			*	****
with respect to the			**	
of the references to			*	
were included in the	*			
is referred to the			*	
interpretation of the references			*	

The data in Table 4 shows that there are 10 LBs within the stance expressions, 12 LBs for the discourse organizers and 40 LBs in the referential expressions. These results are quite similar to those found by Biber and Barbieri (2007). In their study on LBs in academic prose, Biber and Barbieri (2007) found 14 LB types that are included as referential expressions, 3 LBs for stance expressions, and 2 LBs for the discourse organizers. These results, however, are different from those found by Grabowski (2015) who analysed the use of LBs in the selected chapters of academic textbooks on pharmacology. In Grabowski's study (2015), the highest number of LBs is found in the discourse organizers (26 LBs), and then the referential expressions (20 LBs), and finally the stance expressions with only 4 LBs. The high frequency of discourse organizers in the textbooks can be due to the function of discourse organizers to signal relationships between prior and coming discourse, i.e. to establish a firm relation among different parts of the textbooks.

For academic articles, as shown in this study, the highest frequency is found in the referential expressions which can be related to the fact that academic articles need to identify entities, e.g. to refer to theories, concepts, data and findings of the studies. Some examples of the LBs for this purpose are as follows: "on the basis of", "can be found in",

"an increase in the", and "as a result of". The low frequency of the stance expressions in academic articles can be related to the function of stance bundles, i.e. to express attitudes or assessments of certainty. This indicates that authors of academic articles do not like expressing attitudes and assessments too often in their articles.

If we see the data in Table 4 in more details, we can also find some differences in the use of LB functions across the different academic disciplines. For the stance expressions, Social Sciences articles use these expressions more frequently than the articles from other academic disciplines. The corpus shows a significant number of occurrences of the LBs that show epistemic stance, i.e. "is/are likely to be" and "is/are more likely to be" in Social Sciences, but not in the other academic disciplines. Epistemic stance usually functions to hedge claims rather than asserting certainty (Biber & Barbieri 2007: 276). This means that authors of academic articles in Social Sciences hedge their claims more often than those from other disciplines.

For the discourse organizer classification, there is no LB found in the articles of Health Sciences. It seems that in Health Sciences, it not necessary to indicate the overall structure of the articles or the elaborate a topic further. In the other disciplines, we can find the following LBs which are used for topic elaboration: "as well as the" and "in addition to

the". In Physical Sciences and Social Sciences, the LB of "on the other hand" is frequently used.

The use of LBs in the referential expression classification is quite similar to that in the discourse organizer classification. There is only one LB found in the articles of Health Sciences, i.e. "were included in the", whereas for the other academic disciplines, we can find several LBs within the referential expression classification, particularly within the framing attributes. LBs are frequently used in framing attributes because the authors would like to focus the readers on a given case (e.g. "in the case of" and "in this case the") and to state the conditions of an argument (e.g. "in the context of" and "in the absence of").

CONCLUSION

This study has shown similarities and differences in the use of the 62 LBs across the different academic disciplines. Physical sciences feature the most number of lexical bundles (i.e. 43), while Health sciences have the least number of lexical bundles (only 3). Social sciences and Life sciences have 27 and 12 lexical bundles, respectively. When we pair-up the disciplines, we found that Physical sciences and Social sciences shared the most number of LBs, i.e. 14, followed by Physical sciences and Life sciences with 8 LBs, Life sciences and Social sciences with 6 LBs, and Health sciences and Life sciences with only 1 LB. There are no LBs shared between Health sciences and Physical sciences, and neither between Health sciences and Social sciences. For the distribution of the structural forms, this study found that the prepositional-based form and the verb-based form are the most frequent structural forms (each of them accounts for 37.1% of the LBs, making a total of 74.2%). The noun-based bundles only account for 25.81% of the total number of LBs. Within the verb-based bundles, the passive form can be found in 12 out of 23 LB types, which highlight that the sentences in academic journal articles tend to focuses on what is done, rather than who does it. Finally, the distribution of the functional classifications of the LBs across the different academic disciplines shows that the referential expressions (40 LBs) outnumber the discourse organizers (12 LBs) and the stance expressions (10 LBs). The high frequency of LBs in the referential expression classification can be related to the fact that academic articles need to identify entities, e.g. to refer to theories, concepts, data and findings of the studies. A further analysis on the functional classifications across different disciplines shows that the stance expressions, particularly the epistemic stance which functions to hedge claims, are more frequently found in Social sciences than in the other academic disciplines. The data also show that Health sciences do not frequently use LBs for discourse organizers and referential expressions,

while the other academic disciplines (Life sciences, Physical sciences, and Social sciences) frequently employ the referential expressions, especially in framing attributes because the authors would like to focus the readers on a given case and to state the conditions of an argument they presented in the articles.

REFERENCES

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers.* Amsterdam/Philadelphia: John Benjamins.

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., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English.* Harlow, England: Pearson Education.

Chadegani, A, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science*, *9*(5), 18-26.

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

Elsevier. (2016). https://www.elsevier.com/solutions/scopus. Accessed 25.08.18

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23-33.

Harmer, J. (1991). *Teaching vocabulary: The practice of English language teaching*. London: Longman.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar.* New Jersey: Lawrence Erlbaum Associates.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

Lewis, M. (1993). *The lexical approach: the state of ELT and a way forward.* Hove, England: Language Teaching Publications.

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.

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal 47*(1), 30-43.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (eds). *Language and communication* (pp. 191-225).

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*(1), 220-231.

Salazar, D. (2013). *Biomedical English: A corpus-based approach*. Oxford: John Benjamins.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Scott, M. (1997). PC analysis of key words – and key key words. *System*, *25*(1), 233-245.

Stubbs, M (2005). The most natural thing in the world: quantitative data analysis on multi-word sequences in English. Paper presented at *Phraseology* 2005, 13–15 October 2005, Louvainla-Neuve.

Vidakovic, I., & Barker, F. (2010). Use of words and multi-word units in skills for life writing examinations. *Cambridge ESOL: Research Notes*, *41*, 7-14.

Wilkins, D.A. (1972). *Linguistics in language teaching*. London: Arnold.