

Investigating the Use of Lexical Bundles and Keyness in B2 and C1 ESL Learners' Academic Writing

Hattan Hejazi ^[0000-0003-1204-8979]

University of Liverpool, Liverpool L69 3BX, United Kingdom

hattan@liverpool.ac.uk

Abstract. This research investigates whether there is a relationship between the use of three- and four-word Lexical bundles and language proficiency. The study conducts both quantitative and qualitative analyses to see whether learners from different CEFR levels groups exhibit the same behaviour in the use of Lexical bundles. Therefore, in the first stage, it compares between two different levels B2 and C1 in terms of frequency, structures and functions of Lexical bundles to give an overview of some of the linguistic features to differentiate between the levels. In the second stage, a longitudinal study investigated the development of ESL learners use of Lexical bundles across the levels to give a picture of the increases of the proficiency levels. A major finding from the analysis shows that generally, ESL learners favoured using more signalling bundles in their writing, three-word bundles turned out to be the most frequent bundles in ESL sub-corpora. Moreover, significant progress has been found in the variability of the structures and functions of Lexical bundles, C1 writers are found to have used various structures and functions as professional writers in their academic writing. For the development of Lexical bundles in relation to the CEFR levels, the findings clearly indicate that there is no significant relationship between the increased use of Lexical bundles and academic performance. However, multiple regression analysis revealed that there is a direct proportionality between variations of the use of Lexical bundles and the CEFR levels, as (C1) students act as professional writers and used variant structures and functions than (B2) Students.

Keywords: Academic writing, Lexical bundles, ESL Learners, Corpus-based.

Introduction

Lexical bundles are word combinations that can be defined as continuous multiword sequences that recur frequently to satisfy specified frequency and dispersion thresholds; for example, occurring at least 20-40 times per million words in five texts, or in at least 10% of texts [4, 8]. Lexical bundles have captured the attention of many linguists since Biber et al. (1999) first introduced the notion in Longman Grammar of Spoken and Written English. Considerable attention has been given to lexical bundles within the area of corpus linguistics, and interest has increased since being widely agreed that lexical bundles are widespread in spoken and written registers, serving a "building blocks of discourse," where "frequent use of these bundles is indicative

of fluency in linguistic production" [4]. These bundles have been found to be used by both native and non-native speakers of a language to fulfill specific discourse functions within a particular context [5, 9].

The bundles are important elements by which to measure learners' language development, and both native and non-native speakers indicate their language proficiency by using lexical bundles in their academic writing; the absence of these bundles signals a novice writer. This idea has been supported with empirical evidence showing that the competent use of lexical bundles contributes to fluent language production. [6, 12] For example, Biber et al. (1999) investigation of lexical bundles in conversation and academic prose found that bundles constituted approximately 21% of the written discourse. Cortes [11] agrees that using lexical bundles is an indication of a competent language user, and Ellis et al. 2008 argue that use of lexical bundles frequently results in native-like language use.

However, many studies have investigated the use of lexical bundles by non-native speakers of different levels across a range of registers and academic disciplines. According to the previous studies, although there has been an increase in the use of lexical bundles by non-native speakers, their use is limited to specific bundles causing them to overuse some expressions compared to others, making their writing appear non-native [17]. Some studies have argued that experts writers use lexical bundles in a way that is functionally different from novice authors and, in general, that lexical bundles are used much more frequently by experts than novice writers [1, 11]. Römer (2009) states that experts are more important than nativeness and the distinction between novices and experts is more important than L1 and L2 distinction. Similarly, Staples et al. (2013) [21] investigated idiomaticity through the use of lexical bundles in written response across three proficiency levels in the Test of English as a Foreign Language Internet-Based Test TOEFL iBT, in a controlled environment. The study found an increase in the number of lexical bundles used as proficiency level increased.

To the best of the researcher's knowledge, while most previous studies have paid considerable attention to the use of lexical bundles across different registers and a number of disciplines, little research has been done to investigate whether learners from different proficiency level groups exhibit the same behaviour in their use (or not) of lexical bundles. This research investigates whether there is a relationship between the use of three- and four-word lexical bundles and language competence. The study utilises both quantitative and qualitative analyses to determine whether learners from different CEFR (Common European Framework of Reference) level groups exhibit the same behaviour in the use of lexical bundles. Additionally, this study examines the development of lexical bundles across proficiency levels. Specifically, it compares between two different levels, B2 and C1, in terms of the frequency, structures, and functions of lexical bundles to give an overview of some of the linguistic features that differentiate between the levels. This study addresses the following questions:

- What are the most frequently used three- and four-word lexical bundles in the B2 and C1 sub-corpora?

- –What does a keyness analysis reveal about lexical bundles identified in the B2 and C1 sub-corpora?
- –How do lexical bundles in the B2 sub-corpus differ from C1 in terms of structure and function?
- Is there any growth in the lexical bundles identified in the study between B2 and C1 learners?

1 Methodology

1.1 Data

This study is first interested in the relationship between the use of lexical bundles and academic performance; thus, the author compared B2 and C1 sub-corpora (for the frequency, structures, and functions of lexical bundle) of ESL learners and then compared them with a reference corpus. The data used came from written essays equivalent to the IELTS test in terms of the title, written by 42 intermediate and advanced ESL learners from different mother tongue who have studied in the UK who contributing 130 essays. These learners write academic essays to test their progress and place them at new levels if they meet the requirements at the English Language Centre (ELC). Only argumentative or expository pieces written by L2 learners were chosen for the sub-corpora. The decision to use learners' sub-corpora was based on the assumption that they are useful to explore and identify the similarities and differences in the use of recurrent word combinations across L2 proficiencies of "actual language in use" [2].

The second stage of the study consisted of second language development research, which compares learners' language across proficiency levels (CEFR levels). A longitudinal study investigated the development over three months of two ESL learners use of lexical bundles in their academic essays across the levels to trace the increases in proficiency level. The participants were two ESL students (one male and one female) at the upper intermediate level that moved to advanced level after two months who contributed 36 essays to be used for the investigation.

1.2 Determination of CEFR levels

The procedure for determining the CEFR level originates from the manual for Relating Language Examinations to the CEFR for Languages [22]. Using the manual helps to choose the appropriate samples – for standardisation purposes – from the collected essays, which are considered representative of the B2 and C1 levels [9]. Three experienced examiners working at the British Council and teaching IELTS preparation were trained to rate the essays using a Writing Assessment Scale developed by the CEFR. The essays were marked by two raters independently; if any essays were given different scores, they were then re-rated by a third rater. Therefore, they received three ratings rather than two. If an essay received three different ratings, it was excluded. If raters agreed, the inter-rater reliability for the two raters was calculated to determine the percentage of agreement among the raters, following [18] which used by (Chen

and Baker [9] as a statistic to measure inter-rater reliability between the raters. After the rating step, the total number of words in the ESL learner's corpus forming 15488 in B1 sub-corpus and 12752 as described in Table 1.

Table 1. An overview of B1 and B2 ESL learners sub-corpora

Modules	B2 Corpus	C1 corpus
Number of texts	49	35
Average length of the essays	314	364
Total number of words	15488	12752
Total type of words	2723	2513

For the longitudinal study, 35 essays were rerated to be used in the investigation; 15 essays were incorporated into the B2 sub-corpus, totaling 5,007 words, by contrast, the C1 sub-corpus consisted of 20 essays totaling 10,597 words.

1.3 Reference corpus

The reference corpus used in this study was taken from the British Academic Writing English (BAWE) corpus, which contains 2,761 texts of proficient assessed academic works written at universities in the UK (6,506,995 words), ranging in length from around 500 words to approximately 5,000 words. However, since the target sub-corpora used argumentative essays (equivalent to the IELTS task 2) written by ESL learners, it was decided to use BAWE (linguistics and English disciplines) as a reference corpus to avoid skewing the sample heavily toward one discipline. These two disciplines are big enough to be used as a reference corpus as well as include relative language that ESL learners use in their academic essays, using other disciplines such as Philosophy or Biochemistry might effect the results. Therefore, linguistics and English disciplines are suited to the goal of this study as they provide a wide range of language representative of ESL students writing in an authentic academic context.

As stated by Leech [16] 'A Reference Corpus is designed to provide comprehensive information about the language which has to be a general Corpus of wide coverage of the language'. To ensure comparability, only 65 short texts of the BAWE corpus (linguistics and English disciplines) were selected for the investigation. This was sufficient number for a reference corpus and was used in this study, comprising 163,091 words – this is more than five times greater than the target sub-corpora (B2 and C1), having 15,488 and 12,752 words, respectively.

1.4 Analysis

The analysis used to answer the above research questions was carried out using Wordsmith computer software. [20] Due to the smaller sub-corpora size in this study, the low-frequency cut-off point of four times per 100,000 (40times per million words) was selected to include highly used lexical bundles in the analysis and eliminate low-frequency parameters. In addition to frequency cut-off, dispersion criteria were applied where a bundle had to be found in at least three to five texts [4, 8, 11]

or in at least 10% of the texts [12] to avoid focusing on idiosyncratic uses by the individual authors of the texts.

After retrieving the corpus and applying the frequency and distribution criteria, Wordsmith provided lists of three- and four-word lexical bundles for both sub-corpora. Hyponyms were checked and cleared from all the bundles found. In order to narrow down the included lexical bundles, all content-based bundles were discarded, as they do not reflect the use of general academic language, such as The United Kingdom or The University of Liverpool. In addition, overlapping bundles were combined as one bundle to avoid duplication in the counting of high-frequency bundles. For example, the bundle can be used to and it can be used to were counted as one bundle, adding a word between the brackets such as, (it) can be used to [8].

2 Finding and discussion

2.1 Frequency of lexical bundles

The results revealed that the B2 sub-corpus accounted for 102 (type) three- and four-word lexical bundles, which occurred 458 times, making up 9.2 % of the total number of words in the sub-corpus. The C1 essays contained 45 (type) three-and four-word lexical bundles, which occurred 204 times in the sub-corpus and made up 5 % of the total words in the sub-corpus. What stands out is that the lower-level students used a larger stock of lexical bundles than the higher-level students as presented in table 2.

Table 2. Bundles (type) and frequency (occurrences) found in the sub-corpora

Sub-corpora	Frequency	type	%	(Frequency) Per 100,000
B2 (3-word)	388	86	7.50%	2505
C1 (3-word)	169	38	4.00%	1325
BAWE (3-word)	4220	32	7.70%	2587
B2 (4-word)	66	16	1.70%	426
C1 (4-word)	35	07	1.00%	274
BAWE (4-word)	400	37	0.98%	245

In addition, the three-word bundles were revealed to be the most common bundles at both levels. Therefore, it can be concluded that ESL learners have a tendency to employ a higher number of three-word than four-word bundles with an increase in low-level students. A possible explanation might be related to the complexity of their production, which language learners avoid in their writing, as it requires more effort and time for students to produce longer sequences than shorter ones. The result was not surprising; Biber et al. [4] states that three-word Lexical bundles are extremely common because they are “a kind of extended collocational association”, while longer bundles are “more phrasal in nature and correspondingly less common”. Another finding to note is that the bundle *on the other hand* was the most frequently appearing bundle in the B2 and C1 sub-corpora. This bundle is common and important in aca-

democratic discourse; most ESL learners are familiar with it and know how to use it both structurally and functionally.

Surprisingly, few of the most frequent bundles in the BAWE corpus were found in the ESL learners' corpora: only eight out of the 50 most frequent lexical bundles in B2 and C1 sub-corpora were identified in the BAWE corpus. According to that, although the B2 level students used more lexical bundles than C1 students, certain bundles were new and used by only a few learners with repeated the same bundle more than once in their essays. For example, the bundle on the other was identified 19 times in the B2 sub-corpora (although one student used it three times in one text). A possible explanation for this might be that ESL learners tend to use certain lexical bundles more frequently to reflect a high level of formality and demonstrate their language competence; alternatively, they may still be in the process of learning additional lexical bundles. This result conflicts with those presented by Chen and Baker [7], who found many shared lexical bundles across both native and non-native academic writing.

2.2 Keyness analysis

To determine the 'key' bundles in B2 and C1, WordSmith software was used to generate a list of 'key' bundles that occur unusually frequently in the target sub-corpora when compared with a reference corpus (i.e. BAWE) by means of statistical tests (e.g. chi-square or log-likelihood). A 'keyness' value is given for each bundle that has statistically significant, the higher the keyness score, the more the key bundle is statistically significant. The *WordSmith* software provides a list of lexical bundles which are positively and negatively key. However, as the main focus only on the positive keyness, the *WordSmith* tool was sitting to ignore all the negative results as provided in Table 3 and 4.

The results provided some evidence for the common assertion in the previous studies that ESL learners favour particular bundles and overuse them in their writing [12, 15, 19].

Table 3. Key lexical bundles in the B2 sub-corpus with a significantly different frequency from those in the BAWE corpus

Lexical bundles	Frequency in B2	Frequency in RC	Keyness
Point is that	14	0	68.69
First of all	14	02	57.00
To sum up	07	0	34.34
On the other	19	16	34
I will discuss	06	0	29.44
Are very different	05	0	24.53
I believe that	05	0	24.53
Seem to be	05	0	24.53
On the other hand	17	25	31.22

Table 4. Key bundles in the C1 sub-corpus with a significantly different frequency from those in the BAWE corpus

Lexical bundles	Frequency in C1	Frequency in RC	Keyness
Can choose to	05	0	26.52
On the other (hand)	14	29	24.23

The keyness analysis of the sub-corpora revealed that L2 learners overuse some signaling words in their writing. In general, therefore, it seems that low-level students are more likely to rely on the use of lexical bundles than C1 students, and accounted for more instances: nine significant key bundles were identified at the B2 level, whereas only two key bundles were found in the C1 sub-corpus. This result might be affected by the corpus size for this study, as the C1 sub-corpus consisted of only 12752 words.

2.3 Structures and functions in B2 and C1 sub-corpora

Structurally, Biber et al. [6] structural taxonomies were adopted, which have been used in various research studies in this area [3, 10, 12]. However, they were modified and developed for this study, using Biber et al. (2004) classification to place the identified bundles that did not fall under Biber et al. [6] structural taxonomy, as provided in the table 5.

Table 5. Structural taxonomy of lexical bundle types (F= Frequency)

Types	Sub-types	B2		C1		BAWE	
		(F)	%	(F)	%	(F)	%
Verb-based	Anticipatory it + verb / adjective phrase	01	1.0	04	9.0	29	8.12
	be + noun /adjective phrase	03	3.0	02	4.4	09	2.52
	Pronoun/NP + be	13	12.7	08	17.8	18	5.04
	1st person pronoun + dependent clause	09	9.0	01	2.2	01	0.28
	(verb/adjective +) to-clause	12	11.8	09	20	36	10.08
	Other Verb-phrase	03	3.0	0	0	30	8.4
	Total verb-based bundles	41	40.5	24	53.4	123	34.44
Noun-based	NP with of-phrase	10	9.8	05	11.1	95	26.61
	NP with other post-modifier	06	5.5	03	6.66	18	5.04
	Other noun phrases	04	4.0	0	0	12	3.36
	Total noun-based bundles	20	19.3	08	17.8	125	35.01
Preposition-based	Prepositional phrase with embedded of-phrase	02	2.0	0	0	11	3.08
	Other prepositional phrase expressions	16	15.7	07	15.5	55	15.7
	Total preposition-based bundles	18	17.7	07	15.5	66	19.1
Other	Other structures	23	22.5	06	13.3	43	12.04
Overall	12	102	100	45	100	357	100

Although B2 and C1 writers showed variation in the use of lexical bundles according to the structural classification, there were differences in the use of lexical bundles between EFL sub-corpora and the RC.

The results showed that EFL learners used more phrase bundles than clausal bundles in their writing. More specifically, verb-based bundles were the most frequent three- and four-word bundles found in the B2 and C1 sub-corpora. Among the two CEFR levels, the C1 level had the highest proportions of verb-based bundles, at 53.4%, while the B2 level had a lower percentage, 40.5%. These results conflict with the idea of the rarity of verb-based bundles in academic discourse [6]. The results of the present study suggest that the language of EFL writing contains more conversational bundles. By contrast, the reference corpus clearly represents the formal writing genre, as it contains more noun-based bundles, which is a sign of academic writing.

It can be concluded that the three groups employed a different percentage of most of the structural sub-categories, except the 'preposition-based' category. The chi-square test results of the correlational analysis revealed a significant difference among the corpora. The standardised residuals in a chi-square contingency table for the distribution of structural types revealed that greater differences occurred in the 'verb-based', 'noun-based' and 'other' categories. For instance, the test shows that C1 writers overused verb-based bundles compared to B2 writers, which supports the idea that C1 students rely more on spoken language in their writing. In regard to Noun-based, it appeared that B2 students underused these bundles in their writing. On the other hand, B2 writers overused 'other' bundles not related to any sub-category (e.g., as adverbial or modal bundles).

As the standardised residuals in a chi-square did not show any significant difference in the use of 'prepositions-based' bundles, the result reflects the similarity of the proportion of preposition-based bundles in both levels and BAWE, at 15% of total bundles. The 'PP expressions' subcategory is typically used to show the logical relationship between prepositional elements, which means that EFL learners could use this type of lexical bundle to link between the ideas of the argumentation. The difference in frequency of the use of different structural categories across the levels suggests that as their level increases the students are able to recognise and use the adverbial meaning of the bundles.

Functionally, Hyland's taxonomy was adopted, since the data used in this study was mainly academic prose (see Table 6) [12].

In order to be able to classify bundles into the correct sub-categories, it was important to look at the concordance line to see the bundles in their context and to tackle the issue of multi-functionality of the target bundles. There was similarity in the use of functional categories between the levels. The most frequent functions of the identified bundles across the levels were research-oriented followed by participant-oriented, and then text-oriented. The increase in use of research-oriented bundles in the B2 and C1 sub-corpora might be due to the fact that in argumentative essays, students need to describe various aspects and provide different justifications of their ideas to the reader. Bundles of this function accounted for more than 40% of all bundles identified in the corpora. This result is similar to previous studies, which have found that academic writing is dominated by research-oriented bundles over other categories [6, 7, 14]. A consequence of the high proportion of research-oriented bundles might be a focus on describing the problems in the argumentative essay rather than its presentation.

Table 6. Distribution of the functional sub-categories across the groups

types	Sub-types	B2		C1		BAWE	
		(F)	%	(F)	%	(F)	%
Research-oriented	Location	07	6.8	03	7.0	18	5.0
	Procedure	05	5.0	03	7.0	18	5.0
	Quantification	18	17.6	09	20	29	8.0
	Description	19	18.6	06	13	91	25
	Topic	0	0	0	0	0	0
	Total	49	48.0	21	47	156	43
Text-oriented	Transition signals	07	6.8	05	11	12	3.0
	Regulative signals	02	2.0	01	2	13	4.0
	Structuring signals	12	11.7	05	11	0	0
	Framing signals	03	3.0	0	0	63	18
	Total text-oriented	24	23.5	11	24	88	25
Participant-oriented	Stance features	24	23.5	10	22	73	21
	Engagement features	05	5.0	03	7	40	11
	Total participant-oriented	29	28.5	13	29	113	32
Overall	10	102	100	45	100	357	100

In the comparison between the levels, it was seen that B2 writers used research-oriented bundles more often than C1 writers. By contrast, C1 writers employed more text-based and participant-based bundles than B2 writers. The study found a direct proportionality between the percentage of text-oriented and participant-oriented bundles as the level increased. In addition, chi-square unstandardized residuals statistical methods were used in the analysis of structural and functional type, to further support arguments in this study. Statistically, the study has failed to demonstrate any statistically significant difference in functional distributions between the levels.

2.4 Longitudinal study

For the development of lexical bundles across the levels in the second stage, the results were similar to the first stage, where three-word bundles were found to be the most frequent bundles in the EFL sub-corpora. However, the results provide some evidence that suggests there may be development of the use of lexical bundles across the levels, but not to a statistically significant degree. This might be due to the number of collected essays that made up the sub-corpus and the short period of time the learners were tracked over.

Structurally, there was much variability in terms of the structures and functions of lexical bundles across the levels. High-level EFL learners used a greater variety of structures and functions in their writing than low-level learners. The results showed that there were distinctive differences in terms of the greater use of 'noun-based', 'preposition-based' and 'verb-based' bundles by both levels and in the reference cor-

pus. It should be noted that, across the four categories, the percentage of three structural categories in the C1 level seem closer to those in the reference corpus than did those at the B2 level. The B2 levels students used six out of 12 subcategories, while C1 and reference corpus students used 10 out of 12 subcategories. The chi-square revealed significant differences between the levels and the reference corpus, and the standardised residuals (R), which compared observed and expected counts in each cell, showed that greater differences occurred in all the categories, as the C1 and reference corpus used significantly more ‘verb-based’ and ‘noun-based’ bundles and fewer ‘preposition-based’ bundles than B2, except in the ‘other’ category, which did not show any significant difference between the levels, which reflected the frequent use of bundles such as *I want to*, *a lot of*, *the fact that (the)*, and *the development of*.

By contrast, the overuse of preposition-based bundles in the B2 sub-corpus reflected the frequent use of bundles such as *in order to* and *as well as*. Functionally, while the density of text-oriented bundles appeared almost identical in the B2 sub-corpus, the use of research-oriented and participant-oriented bundles in the C1 sub-corpus seems to be more aligned with the reference corpus.

Further analysis of the functional sub-categories revealed the same results as for the structural sub-categories: the C1 level seemed closer to the reference corpus than the B2 level. The result of the chi-square test revealed a significant difference among the three groups. The standardised residuals (R), which compare observed and expected counts in each cell, showed that the greatest differences between the groups occurred in the ‘text-oriented’ and ‘participant-oriented’ categories, as the C1 and reference corpora used significantly more participant-oriented bundles but fewer text-oriented than the B2 level. This might be due to the wide range of topics that argumentative and expository essays covered.

3 Conclusion and Limitations

3.1 Summary of findings

A major finding from the analysis was that, generally, EFL learners favoured using more signaling bundles in their writing; three-word bundles were found to be the most frequent bundles in EFL sub-corpora. Moreover, significant progress was identified in terms of the variability of the structures and functions of lexical bundles, C1 writers were found to have used various structures and functions as professional writers in their academic writing. In terms of the development of lexical bundles in relation to CEFR level, the findings clearly indicated that there was no significant relationship between the increased use of lexical bundles and academic performance. However, multiple regression analysis revealed that there is a direct proportionality between variations in the use of lexical bundles and CEFR level, as higher-level students (C1) acted as professional writers and used more variant structures and functions than lower-level students (B2).

The results of this study show that there are specific lexical bundles that maybe considered to be the building blocks of ESL learners academic essays. These results might be interesting for English language teachers and instructors be-

cause it provides insights into the ESL learners community preferences in academic writing.

3.2 Limitation

Like many other studies, the present investigation has its limitations. One of which is the small corpora size. However, small corpora size can produce more lexical bundles than the big corpus. [13] To avoid biased results, the frequency cut-off point and dispersion criteria were set at 40 occurrences per million words to include highly used lexical bundles in the analysis and eliminate low-frequency parameters. In addition to frequency cut-off, dispersion criteria were also applying in at least three texts.

Acknowledgement. We are grateful to the three anonymous reviewers who provided insightful comments on earlier versions of this article.

References

1. Ädel, A. & Römer, U.: Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics*, 17, 3-34 (2012).
2. Adolphs, S.: *Introducing electronic text analysis: A practical guide for language and literary studies*, Routledge (2006).
3. Bal, B.: *Analysis of Four-word Lexical Bundles in Published Resesarch Articles Written by Turkish Scholars*. Georgia State University (2010).
4. Biber, D. & Barbieri, F.: Lexical bundles in university spoken and written registers. *English for specific purposes*, 26, 263-286 (2007).
5. Biber, D., Conard, S. & Cortes, V.: If you look at . . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405 (2004).
6. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. & Quirk, R.: *Longman grammar of spoken and written English*, MIT Press Cambridge, MA (1999).
7. Chen, Y.-H. & Baker, P.: Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14, 30-49 (2010).
8. Chen, Y.-H. & Baker, P.: Lexical bundles in L1 and L2 academic writing. 14, 30-49 (2010).
9. Chen, Y.-H. & Baker, P.: Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37, 849-880 (2016).
10. Cortes, V.: *Lexical bundles in freshman composition*, Amsterdam, John Benjamins Publishing Company (2002).
11. Cortes, V.: Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423 (2004).
12. Hyland, K.: Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62 (2008).
13. Hyland, K. J.: Bundles in academic discourse. 32, 150-169 (2012).
14. Jalali, H. & Zarei, G. R.: Academic writing revisited: A phraseological analysis of applied linguistics high-stake genres from the perspective of lexical bundles. *Journal of Teaching Language Skills*, 34, 87-114 (2016).

15. Lee, D. Y. & Chen, S. X.: Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18, 281-296 (2009).
16. Leech, G.: The importance of reference corpora. *Hizkuntza-corporak. Oraina eta geroa* (2002).
17. Li, J. & Schmitt, N.: The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102 (2009).
18. Mchugh, M. L.: Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22, 276-282 (2012).
19. Römer, U.: English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies*, 20(2), 89-100 (2009).
20. Scott, M.: *WordSmith Tools (Computer Software. Version 6.0)*. Liverpool: Lexical Analysis Software (2012).
21. Staples, S., Egbert, J., Biber, D. & McClair, A.: Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, 12(3), 214-225 (2013).
22. Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N. & North, B.: *Common European Framework of Reference for Languages: learning, teaching, assessment*, Cambridge University Press (2009).