International Journal of Arts, Humanities and Social Studies
Website: https://www.ijahss.in/
ISSN(*Online*): 2582-3647
Volume 4; Issue 1; Jan-Feb 2022; Page No. 332-338
Open Access
Original Paper

Analysis of Lexical Bundles in Published writing: Education and geology: A Corpus- Based Study

Arwa Suliman Alhwimel
Lecturer of English Language; Mustaqbal University

ABSTRACT
This paper presents the findings of a study that analyzed the use of lexical bundles in two registers of research articles: education and geology. The corpus (Ant Conk) consisted of education articles, and geology articles. The most frequent four-word combinations were selected as lexical bundles and classified functionally in each register. The analyses showed that the bundles identified in each discipline are different to some extent. Finally, a functional classification showed that some bundles in both disciplines shared functions connected to the specific sub- function, while some of them are connected to the register's interest.

Keywords: Lexical Bundles (Lbs), Education, Geology
***Citation*:** Arwa Suliman Alhwimel (2022). Analysis of Lexical Bundles in Published writing: Education and geology: A Corpus-Based Study. *International Journal of Arts, Humanities and Social Studies*, 4(1), 332-338.

INTRODUCTION

Word combinations have become a significant analysis of academic discourse. Frequent occurrence words in different written disciplines play an importance role to show the writers' competence. Knowing these lexical bundles and using them in language production is an important aspect to English Language Learners (ELLs) in order to place the individual words within a larger context in different disciplines. In fact, the use of collocations is considered a proficient ability in language use on particular registers [1]. High proficiency students and authors are expected to write in a native-like proficiency language to be able to reach to their audience' satisfaction. They need to be familiar with academic writing features and the usage of lexical bundles. Lexical bundles have specific features distinguishing them from other types of multi-word expressions such as collocations and idioms. According to Biber and Barbieri [2], "most lexical bundles are not idiomatic in meaning and not perceptually salient" as cited in [3]. In other words, a lexical bundle is easily can be understood by looking at it as individual items, whereas idioms are more deeply embedded in the literal meaning.

Statement of Problems

By reviewing studies completed on LBs in research articles, I found that studies present lexical bundles and word lists in a variety of registers such as Cortes [1,4] in history and biology registers, Herbel- Eisenmann, Wagner, and Cortes [5] in mathematic registers, Sánchez in biology register as well, P. [6]. However, no study appears on lexical bundles in education and geology academic research articles. Therefore, the purpose of this study is to identify and functionally analyze the most frequent lexical bundles in education and geology research articles using comparable corpora to discover some common ground in the use of these linguistic features.

Research questions:

1) What are the most frequent lexical bundles in published education and geology research articles?
2) What are the functions of the lexical bundles among these disciplines?
3) Are there any significant differences in the functions of LBs between these disciplines?

Limitation

There is a considerable limitation in this study which is the size of the corpora. These corpora are considered a small size of corpus that may affect the reliability of the study's results. Therefore, a further study is needed to be conducted on large size of corpora in education and geology research articles.

Literature Review

There have been many studies have focused on academic writing and they indicated that "language in use is characterized by repetition of fixed and semi-fixed multi-word combinations and by use of formulaic patterns" [7] and one type of these multi-word combinations that is extensively studied in the literature is "lexical bundles" [3]. Lexical bundles (LBs) are defined by Biber, Johansson, Leech, Conrad and Finegan [8] as the most frequent recurring multi-

words in a discipline regardless of their idiomatic and their structural status [1]. According to Biber et al. [8], lexical bundles first appeared in the Longman Grammar of Spoken and Written English. They are sequences of words that have various length typically, four-word sequences that are repeated at least 20 times in a million words as cited in [9]. Some sequences appeared in many registers, while others are specific to particular registers such as, "What do you mean?" or "I don't know why," which appear frequently in daily conversations, and occur more often in academic prose [10]. In series of lexical bundle studies conducted by Biber and others [9] investigated that conversation and academic prose reveal different distribution patterns of LBs as cited in [11].

Furthermore, multi-word expressions in academic prose, according to Biber et al. [9], often performed to link two phrases; for example, a prepositional phrase in the case of or a noun phrase the base of the, to function as scaffolding for new knowledge. Thus, nemours researchers like Hyland [12], Neely and Cortes [13], and Hyland and Tse [14] claim that LBs are agents that determine the familiarity of writers in a particular discourse [15]. Therefore, LBs require understanding the context in academic prose in order to be produced appropriately. Many studies have examined clusters in both non-native academic writing [16,17], and general academic writing such as Byrd & Coxhead [7].

METHODOLOGY

The present study quantitatively adopts a corpus-driven approach [18] based on analysis of the use of four-word lexical bundles by Ant Conc program designed by Wright [19]. The four- words combinations are chosen to be analyzed because they are the most studied length of words in much research and considered to be manageable in size [11]. They are also "over 10 times more frequent than five-word sequences and offer a wider variety of structures and functions to analyze" [20].

Corpus Collection

The collection of the study's texts are two different academic research articles corpora. One is academic research articles in education and the other one is academic research articles as well in geology (see table 1 for the corpora details).

Table 1: *Corpus of Published research articles in education and geology*

Disciplines	**Education**	**Geology**
Journals	American Educational Research Journal. American Secondary Education.	Geology Geochimica & Cosmochica Acts.
Number of texts	15	15
Number of words	122,932	80,242

These disciplines are selected because they represent different usage of lexical bundles and obviously function differently in these registers due to the variety of their topic-focus. In addition, they are chosen because none of any previous corpus-based studies discussed them. Moreover, the texts are chosen randomly from high ranked journals published in 2012. In fact, theses corpora are created by Heidi Wright, a professional researcher in corpus-based studies in different genres and various focused topics. She created them for her study "Stance features within stand-alone literature reviews and research articles: An interdisciplinary register analysis, 2014". With all appreciation, she shared the education and geology research articles corpora with me to look at the functions of lexical bundles in each one of them, which is completely different focus than her study in 2014.

The present study used Ant Conc computer program that released in 2002 to analyze the registers. It designed specifically for cluster and lexical bundle analysis, and a word distribution criterion [21]. In this study, all the corpora inserted in Ant Conc and "N.grams" is chosen to release the data.

RESULTS

The most frequent four-word combinations that match the cut-off point were chosen in this study to be lexical bundles. The cut off of the LB was calculated based on their occurrence of 50 times per 100,000 words. Lexical bundles are also classified in this study functionally according to the taxonomy designed by Cortes [1] as mentioned in Biber, Conrad, and Cortes [1], which refer to the purposes of the LBs in the disciplines according to situations or contexts [22]. The three core categories in this taxonomy used in the current study are taken from Cortes [1], referential bundles, text organizers and stance bundles.

Lexical bundles (LBs) in education and their functions.

Table 1: Frequent Lexical bundles in education research articles

Frequent	Range	Lexical Bundles
21	6	are more likely to
18	4	the extent to which
15	7	it is important to
12	6	as a result of
10	6	at the beginning of
10	7	at the same time
9	6	as well as the
9	4	it is possible that
9	3	the effect of the
9	4	were more likely to
8	4	at the time of
8	4	in the context of
8	3	in the same school
8	4	in the united states
8	5	more likely to be
8	6	on the other hand
8	5	the end of the
8	4	the time of the
7	5	At the end of
7	3	In this study we
7	6	of high school students
7	3	out of high school
7	4	the percentage of student
7	6	to be able to
6	4	findings of this study
6	3	in the development of
6	3	schools are more likely
6	4	students were asked to
6	3	the administration of the
6	5	the beginning of the
6	3	the findings of this
6	4	the impact of the
6	3	who participated in the

The above table reveals the lexical bundles list that identified in education corpus. The result indicates that all the functions are identified and used in the education corpus (as can be seen in Table. 2 below). Particularly, according to the mentioned taxonomy, one of the identified functions are referential bundles that used as time and place markers, as in the following examples:

We also included class measures collected (*at the beginning of*) the school year. *(At the same time)*, 10 interviewees stated that at least some students did not Text organizers in education also used different functions, as contras bundles, inferential bundles, focus bundles, and framing bundles, to organize logical development of the context and to "interpret prior or forthcoming discourse" [1] as in the following examples:

Thus, we have reservations about the argument that a hands-on approach is better understood as a strategy of Whiteness. On the other hand, we agree that it is important to contextualize the hands-on approach.

Final function in this register is stance bundles as a category of expressions [1], and it used to convey a degree of probability, possibility, and ability. An example is:

Teachers *are more likely to* implement innovations in their classroom. *It is possible that* the portions of the course that focused on having teachers analyze classroom video in very specific ways were among the most important for helping teachers use these practices in their classrooms.

Table 2: *Lexical bundles in education articles*

Category	**Sub -category**	**Education**
Referential bundles	Time markers	at the beginning of, at the same time, at the time of, the end of the, the time of the, at the end of, the beginning of the
	Place markers	in the context of, in the same school, in the united states, in this study we
Text organizers	Contrast	on the other hand
	Inferential	as a result of, findings of this study, the findings of this, the effect of the, the impact of the
	Focus	it is important to
	Framing	the extent to which as well as the
Stance bundles	Possible/ probable	are more likely to, it is possible that, were more likely to, more likely to be, schools are more likely
	ability	to be able to
Other bundles identify the register Context.		of high school students, out of high school, the percentage of students, students were asked to

Lexical bundles (LBs) in geology and their functions.

On the other hand, with a fewer number of bundles, geology writers feature much fewer uses of LB than educational writers (see Table 3 below).

Table 3: *Lexical bundles in geology*

Frequency	range	Lexical bundles

10	3	in the water column
10	6	the end of the
8	3	as a function of
8	3	in the deep sea
7	4	as a result of
7	5	at the end of
7	4	similar to that of
6	4	the center of the
6	4	is consistent with the
6	4	it is likely that
6	4	the base of the

The results show some of the discussed functions used (see table 4 below).

Actually, there is a slight functional difference in this corpus than in the education corpus, which elaborates that the corpora are linguistically similar in language use with absolute differences in their focus. The first identified function is referential bundles, time markers and places, as in the following examples:

Table 4: *Lexical bundles in geology*

Category	**Sub-category**	**Geology**
Referential bundles	Time markers	the end of the, at the end of
	Place markers	the center of the
Text organizers		similar to that of
	Compare	the base of the, as a result of
	Inferential	is consistent with the, as a function of,
	Focus	

Stance bundles
it is likely that
Possible/ probable
Other bundles
identify the
in the water column, in the deep sea
register context.

The initial amount of liquid H2O2 put in the desiccators ensured that H2O2 vapor remains available until *the end of the* experiment.

Text organizers also recognized differently in than education functions. In geology, the authors mostly use inferential and focus functions of the lexical bundles more than educational authors. Example of them is as following:

Because strong weathering *as a result of* terrestrial ecosystem destruction did not fertilize the oceans until the Early Triassic, end-Permian conditions are better represented by the experiments with no or little nutrient increase.

Also, substance bundles are the less function used in this corpus. An example of probability which is the single sub-function of stance bundles has used:

it is likely that only a small portion of Fe(II) would be lost back to the water column, with a larger fraction being absorbed onto Fe(III) oxyhydroxides or precipitated as Fe carbonates or Fe silicates.

Unlike the education corpus, framing and contrast are two sub-categories that do not appear in the geology corpus for this study.

Eventually, I found that there are other LBs that are not categorized functionally, however, Cortes [1] described them as subject-specific, but I think they would be described as disciplined-focus interests because most of them have a word related to the discourse. For example, in education, we notice words like "school," and "students" as in the following: "of high school students," or "the percentage of students." In geology, we found words like "water," and "sea" as in the following phrases: "in the water column," or "in the deep sea."

Final findings compare the two corpora in terms of different LB functions. There are no remarkable differences between these corpora in lexical bundle functions. They exhibit differences in the number of LBs that appear in this study. Education corpus reveals more bundles than geology corpus. This difference is a problematic due to the different size between these corpora. The education corpus contains more words than the geology corpus as shown in (Table 1.) in order to compare the two corpora in equal size, normed frequency is used for the present study to show the exact number of lexical bundles in per 100,000 words. In comparing these corpora in normed frequency, the results show that there are several lexical bundles appear in both corpora and used in the same functions.

However, they differ in their frequent occurrence in each register per 100,000 words. For example, (*at the end of)* occur in both corpora, and it occurs in education research articles 5.7 times per 100,000 words, whereas it occurs to double in geology articles 10.0 times.

DISCUSSION

Based on the results of the present study, the answer of the first research question is that, as shown in Tables 1 and 2, the most frequent lexical bundles in current education corpus is (*are more likely to)* in normed of 16.3 per 100,000 words; whereas in geology *in the water column* is the most frequent lexical bundles in normed of 5.0. Interestingly, in geology corpus, (*in the water column)* and (*at the end of the)* appear frequently equal in the corpus with different ranges. However, (*the end of the)* disappear in the normed frequency count which elaborates that the bundle *in the water column* is more used than (*the end of the)* in geology registers.

For the second research question regarding the functions of LBs, both registers use the lexical bundles in the same functions with slight difference in their sub- category functions. Despite they share the same functions, they do differentiate in which function is mostly used in each discipline. The findings show that educational register mainly focuses on referential bundles function as time and place markers because I believe that the nature of schoolwork require explicitly stated steps and their exact position on marked time for example, writing a lesson plan requires timed steps, and a school's schedule also requires time and place markers, as well. On the other hand, the text organizers function is the function mostly used in geology register due to their writing style that required analysis and explanations that require organizing thoughts.

Moreover, these results, along with some prior research by Biber, et al [9], Cortes [1] and Cosmay [10], indicate that different LBs were employed in different registers with similar functions. Consequently, there are not any significate functional differences between the disciplines of using the word combinations which answer the third research question.

CONCLUSION

The main goal of this research is to identify and analyze the functions of four words lexical bundles in both education and geology research articles applying a corpus-driven approach. After inserting the two corpora in the computer program AntConc, the results show that there is initial difference in the actual bundles that appeared in each one; however, they share the same functions of the LBs in the texts.

The current study supports the findings of other studies conducted by Cortes [1] and Biber [2] and indicate considerable differences in the frequency terms of forms, structures, and functions across sorts of academic writing.

REFERENCES

1. Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*, 397–423.

2. Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263–286.
3. Öztürk, Y., & Köse, G. D. (2016). Turkish and Native English Academic Writers' Use of Lexical Bundles. *Journal Of Language & Linguistics Studies*, *12*(1), 149-165.
4. Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3*(1), 43-57. doi:10.3366/E174950320800006
5. Esimenmann,B.,Wagner,D.,Cortes,V.(2010). Lexical bundle analysis in mathematics classroom discourse: the significance of stance. *UNB Scholar Research Repository.* doi: 10.1007/s10649-010-9253-6
6. Sánchez, P. (2014). Lexical bundles in Biology Differences between textbooks and research articles. *Revista Española De Lingüística Aplicada (John Benjamins Publishing Co.)*, *27*(2), 492-513. doi: 10.1075/resla.27.2.12san
7. Byrd, P. , & Coxhead, A. (2010). On the other hand: lexical bundles in academic Writing and in the teaching of EAP. *University of Sydney Papers in TESOL*. ISSN: 1834-3198
8. Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman
9. Biber, D., Conrad, S., & Cortes, V. (2004). "If you look at" ...: lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405.
10. Csomay, E. (2010). Lexical bundles in discourse structure: a corpus-based study of classroom discourse. *Applied Linguistics*, 34(3), 369-388.
11. Chen, Y. & Baker, P. (2010). Lexical bundles in l1 and l2 academic writing. *Language Learning & Technology.*
12. Hyland, Ken., (2008). As can be seen lexical bundles and disciplinary variation. *The American University.* doi: 10.1016/j.esp.2007.06.001
13. Neely.E, Cortes.V.(2009). A little bit about: Analyzing and teaching lexical bundles in academic lectures. *Language Value. ISSN 1989-7103*
14. Hyland,K., Tes,P.(2009). Academic Lexis and Disciplinary Practice: Corpus Evidence for Specificity. *International Journal of English Studies.* doi: 10.6018/ijes.9.2.90781
15. Christina Ong Sook, B., & Yuen Chee, K. (2015). Functional types of lexical bundles in reading texts of malaysian university english test: a corpus study. *GEMA Online Journal Of Language Studies*, *15*(1), 77-90.
16. Wei,Y.,Lei,L. (2011). Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners. *RELC Journal. doi.org/10.1177/0033688211407295*
17. Hyland,K.(2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied linguistics.* doi.org/10.1111/j.1473-4192.2008.00178.x
18. Biber, D., & Gray,B. (2015). Phraseology. *The Cambridge Handbook of English Corpus Linguistics.* Cambridge University Press.
19. Wright, H.R. (2014). Stance features within stand-alone literature reviews and research articles: An interdisciplinary register analysis. *Paper presented at the conference of the American Association for Corpus Linguistics (AACL), Flagstaff, Arizona*.
20. Hyland,K.(2012). Disciplinary identity. *Cambridge University Press.*
21. Anthony, L. (2005). Antconc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. IEEE International Professional Communication Conference Proceedings.
22. Ranjbar, N., Pazhakh, A., & Gorjian, B. (2012). The effect of lexical bundles on Iranian efl learners' linguistic production fluency. *International Education Studies*, *5*(4), 243-251.