



Lexical Bundles in Educational, Science and Political Articles in the International Journal Directory of Open Access Journals (DOAJ)

Sri Intan Widyaningrum^(✉), Ni Made Chandra Widayanti,
and Komang Elisa Ayumi Dewi

Master of Linguistics Program, Warmadewa University, Denpasar, Indonesia
sriintanwidyaningrum@gmail.com

Abstract. Lexical bundles are defined as sequences of three or more words that show a statistical tendency to occur together in a register. This study focuses on identifying the frequency, structure, and function of lexical clusters in three types of articles, namely Education, Science and Politics. This research is a corpus-based research with an inductive approach. Corpus of data. The results showed that science-themed articles had the highest percentage of lexical bundles reaching 65.22% compared to articles on education and government, which means science articles used the most lexical bundles in their writing. Followed by articles with the theme of Education with 26.88% and articles with the theme of government having the least percentage of lexical bundles, which is 8.7%. There are 2 categories of lexical bundles that can be identified in science-themed articles, namely noun phrases with the affix of, and prepositional phrases with the affix of, while 3 types of lexical bundles in articles with the theme of Education are prepositions with the affix of and others, and the category of lexical bundles “other”. And 2 types in the articles on government themes, namely prepositions with the affix of and the type of lexical bundle “other” or others. In this study, there were no lexical bundles with the structure of noun phrases with affixes other than of, and the lexical bundle was “verb-based”. So it is necessary to do further research with articles on the theme of Education, science and government.

Keywords: Lexical clusters · Lexical bundles articles Education · Science and politics

1 Introduction

Articles can be understood as a series or essays made based on facts and opinions to be published in the media, be it print media, online media, even now many articles are uploaded on social media. Mappatoto (1994) defines an article as a complete written work, non-fiction writing, and a written essay of indefinite length [1]. Article writing itself actually contains the aim of conveying an idea that contains data and facts. The

ideas in the article can ultimately educate, convince, and also become entertainment suggestions for readers. In addition, topics, themes, or ideas in an article can be presented in the form of an opinion. This opinion is usually used to respond to a problem that occurs in the community as well as provide a solution to the problem. The thing that needs to be considered in writing articles is that all opinions and solutions submitted must depart from accurate facts and data, so that articles can be tested for truth. In most articles, it was found that the use of lexical clusters (Lexical bundles) could not be separated from the Education, Science, and Politics Articles contained in the International Journal of the Directory of Open Access Journals (DOAJ). Lexical clusters are defined as sequences of three or more words that show a statistical tendency to occur together in a register.

Biber et al. (1999) define a lexical cluster as a repeated sequence of words that appears throughout the text in the same register and helps to form the distinctiveness of the register [2]. In a further publication, Biber (2004) simply defines a lexical cluster as the most frequently repeated word sequence in a given register [3]. This is seen as an important aspect that must be mastered to create a quality text or story. Several studies have shown that knowledge of this lexical bundle. The first is an article entitled *Lexical Bundles of Indonesian and English Research Articles: Frequency Analysis* by [4]. This study aims to obtain a list of common lexical bundles in articles of applied linguistics and describe the patterns of using bundles. The most frequently investigated lexical bundles based on the frequency criteria reflect the general pattern of bundle use in each corpus. The second article is a *Lexical Collection in Student Essay Writing* (2020) by Nadia Ulfa. This study uses a corpus-based analysis that composes a special small corpus consisting of 161 student essays on linguistic topics. And the third article is the *Indonesian Lexical Collection in Research Articles: Frequency, Structure, and Function* [5]. This study aims to determine the characteristics of the Indonesian lexical bundle which includes frequency, structure, and function in research articles. This study is different from the previous article because this article looks at the comparison of the use of lexical clusters that appear in the articles on Education, Science and Politics.

Because there are no previous studies that compare the collection of lexical clusters in articles of the three types, therefore, to fill the gap, this study aims to compare the lexical collections of the three types of articles, namely Education, Science, and Politics where these three types of articles are the most we often encounter in international journals. Based on the explanation given, the objectives of this research are as follows:

- a) The use of lexical groups in each article has a different frequency. There are articles that are often used and there may be infrequent ones. So the first objective of this research is to compare the use of lexical bundles in three different types of articles, namely Education, Science and Politics;
- b) Lexical clusters have a structural form so that the second objective in this study is to determine the distribution of the structural form of the lexical bundle in three different types of articles, namely Educational, Science and Politics Articles;

Lexical bundle, defined as ‘the lexical sequence that appears most frequently in the register’ Biber et al. (2004). Lexical bundles, also referred to as N-grams, often do not fit into traditionally idealized language units, but can traverse a number of structures e.g. In our research, it should be noted that. These bundles often have visible functions in certain registers. For example Cortes (2004) find that classroom teaching tends to feature a more personal set of attitudes (e.g., you should do it) than academic prose, which in turn displays a more impersonal set of attitudes (e.g., need) [6]. Also, the sets of organizing discourses are found to be different across the spoken and written registers: if you look at is the general set found in the former, while on the other hand, it is more often found in the latter. In terms of the preparation of the language corpus, there are now many models of the language corpus available. More specifically, judging from the type, there are several types of corpus that can be used depending on the type of analysis. Here are the kinds:

1. General/reference corpus, for example: British National Corpus (BNC) – “vs” special corpus, which is made to represent a language in general and comprehensive, consists of written and spoken language data which is then transcribed and sourced from various types of texts.
2. Historical corpus, eg Helsinki Corpus vs modern language corpus, which is structured to represent specific periods of the development of a language.
3. Regional corpus, such as the Wellington Corpus of Written New Zealand English (WCNZE), are structured to represent the language variations of a region.
4. A learning corpus, such as the International Corpus of Learner English (ICLE) vs. the corpus of native speakers, is created to represent the language that foreign learners produce for that language and are not native speakers.
5. Multilingual corpus vs monolingual corpus, which represents at least a phenomenon from two languages originating from the same variety of texts, is usually used for contrastive analysis purposes.
6. Oral vs. written corpus vs. oral-written mix, to represent the spoken language of a language-speaking community.

There is one other type of corpus whose preparation is not based on the source, but based on a stage in the initial processing process, namely the annotated corpus vs. orthographic corpus. In the annotated corpus, several types of linguistic analysis have usually been carried out on the text so that it contains annotations which are not actually part of the original text [7].

2 Method

Corpus linguistics is an empirical method of linguistic analysis and description to examine the language used by natural speakers [8]. This method is increasingly being used and is able to examine various linguistic problems in a broad scope and is even able to reveal interesting, fundamental, and even surprising things from the existing linguistic data. What kind of linguistic data can be analyzed linguistically with this method? Nesselhauf (2011: 2) states, in general, the data is divided into two large groups, namely

NO	Lexical Bundles	Pendidikan	Sains	Politik
1	In the study of	–	✓	–
2	In favor of a	–	✓	–
3	In which may have	✓	–	–
4	As well as its	–	–	✓
5	As a matter of the	–	✓	–
6	A source of	–	✓	–
7	As a part of the	–	✓	–
8	As the demand of	–	✓	–
9	As higher level of	✓	–	–
10	On the acceleration of	–	–	✓
11	On the mistakes of	✓	–	–
12	At the standard that	✓	–	–
13	At the same time	✓	–	–

In this section a collection of Lexical bundles is analyzed and classified according to their structural properties. According to Biber et al., (2004) although lexical bundles are not considered as complete structural units, it is possible to group them according to their grammatical correlation. In this study, lexical bundles were classified into 4 main structural categories. They are Noun Phrase (NP) (eg the edge of the), Verb Phrase (VP) (eg came out of the), Prepositional Phrase (PP) (eg: in the middle of) and others. The table below illustrates the distribution of the various types of lexical bundle structures found in this study.

Structure		Lexical Bundles	Frequency		
			Pendidikan	Sains	Pemerintahan
Noun-based	NP + of	<i>As a matter of, as a part of, as the demand of,</i>	0	5	0
	Noun phrase with other post-modifier fragment	–	–	–	–

(continued)

(continued)

Structure		Lexical Bundles	Frequency		
			Pendidikan	Sains	Pemerintahan
Prepositional-based	Prepositional Phrase + of	<i>In the study of, in which may have, in favor of a, as source of, on the acceleration of, on the mistakes of</i>	2	10	1
	Other Prepositional Phrase	<i>At the standard that, at the same time</i>	2	0	0
Verb-based	Be + noun phrase/adjective phrase	–	–	–	–
	Passive verb	–	–	–	–
	Verb/adjective + that	–	–	–	–
	Verb/adjective + to	–	–	–	–
	Adverbial clause	–	–	–	–
Others		<i>As higher level of, as well as its</i>	2	0	1

Based on the results of the analysis described in the table above, the authors can identify the lexical bundle which is then classified into four categories, namely Noun phrase + Of or (NP + of), Preposition + of (PP + of), Other prepositional phrase, and other or not included. Into another category. Based on the theory proposed by Bieber et al. (1999) noun phrases are further divided into two categories, namely Noun Phrase + of and Noun phrases with modifiers other than of. In this study, the authors found lexical bundles that fall into the noun phrase category, namely As a matter of, as a part of, as the demand of, all of which are part of science-themed articles. The lexicon “matter”, “part”, and “demand” are nouns and end in “of”. There were no lexical bundles for noun phrase categories with affixes after nouns other than of.

Similarly, prepositional phrases are also divided into two categories, namely with affixes after the preposition “of”, and prepositions with other affixes after them. In this study, it can be identified In the study of, in which may have, in favor of a, as source of, on the acceleration of, on the mistakes of which are found in the three articles. And prepositional phrases with other affixed words, namely At the standard that, at the same time., which can only be identified in articles with the theme of Education. As well as

lexical bundles for the “other” category, namely As higher level of, as well as its. Which is found in articles on education and government. The frequency of using lexical bundles in the three articles was 23 times where the highest category and frequency were found in science-themed articles.

4 Conclusion

Based on the results of the research that has been done, the authors can identify the similarities and differences in the use of lexical bundles in several articles with different themes. The results showed that science-themed articles had the highest percentage of lexical bundles reaching 65.22% compared to articles on education and government, which means science articles used the most lexical bundles in their writing. Followed by articles with the theme of Education with 26.88% and articles with the theme of government having the least percentage of lexical bundles, which is 8.7%. There are 2 categories of lexical bundles that can be identified in science-themed articles, namely noun phrases with the affix of, and prepositional phrases with the affix of, while 3 types of lexical bundles in articles with the theme of Education are prepositions with the affix of and others, and the category of lexical bundles “other” And 2 types in the articles on government themes, namely prepositions with the affix of and the type of lexical bundle “other” or others. In this study, there were no lexical bundles with the structure of noun phrases with affixes other than of, and the lexical bundle was “verb-based”. So it is necessary to do further research with articles on the theme of Education, science and government.

References

1. Andi Baso Mappatoto. (1994). *Teknik Penulisan Feature*. Gramedia Pustaka Utama.
2. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. England: Pearson Education Limited.
3. Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(35), 371–405.
4. Yuliawati, azka saeful haq; R. M. A. S. (2021). Lexical Bundles of Indonesian and English Research Articles: Frequency Analysis. *JALL (Journal of Applied Linguistics and Literacy)*, 5(1).
5. Adi Budiwiyanto, T. S. (2020). Frequency and structure of Indonesian lexical bundles on academic prose in legal studies: A driven-corpus approach. *Proceedings of the Proceedings of the Third International Seminar on Recent Language, Literature, and Local Culture Studies, BASA*, 20–21.
6. Cortes, V. (2004). Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology. *English for Specific Purposes*, 23, 397-423.
7. Nesselhauf, N. (2011). Collocations in a Learner Corpus. *Machine Translation*, 20(4), 3.
8. Zhu, Q., Yan, Y., Xu, C., & Jin, Y. (2012). Computer-Aided Value-Assessment Model: Review for Bilingual Teaching Courses Quantitative Analysis. *IERI Procedia*, 2, 2–7. <https://doi.org/10.1016/j.ieri.2012.06.042>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

