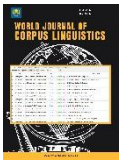


World Journal of Corpus Linguistics

Vol. 2, No. 1 – 2025, Page. 6-14| Available online at
https://pusatpublikasi.com/index.php/corpus/index

LEXICAL BUNDLES FOUND ON JOURNAL OF MEDICAL ARTICLES REGARDING CORONAVIRUS DISEASE

Indra Desman Telaumbanua
Magister Ilmu Linguistik Universitas Warmadewa
indradesman1297@gmail.com

Abstrac
The goal of this research was to assess the utilization of lexical bundles of three and four words in medical papers about coronavirus diseases. Using a corpus-based analytic approach, this research created a tiny, specialized corpus of articles with a medical emphasis. The computational tool used to produce the commonly occurring list of sequences was AntConc 3.5.8 by Laurent Anthony. The data revealed that in terms of the pandemic of co-occurrence, the most frequent phrase was the function of three-word bundles and WHO being clinically handled as four-word bundles were detected.

Keywords: Antconc, Corona Virus, and Lexical Bundles.

Abstrak
Tujuan dari penelitian ini adalah untuk menilai penggunaan kumpulan leksikal tiga dan empat kata dalam makalah medis tentang penyakit coronavirus. Dengan menggunakan pendekatan analitik berbasis korpus, penelitian ini menciptakan korpus kecil artikel khusus dengan penekanan medis. Alat komputasi yang digunakan untuk menghasilkan daftar urutan yang umum terjadi adalah AntConc 3.5.8 oleh Laurent Anthony. Data mengungkapkan bahwa dalam hal pandemi kejadian bersama, frase yang paling sering adalah fungsi bundel tiga kata dan WHO secara klinis ditangani saat bundel empat kata terdeteksi.

Kata Kunci**: Antconc, Virus Corona, dan Lexical Bundles**.

I. INTRODUCTION

These components may also be referred to as "prefabs," "parts," "patterns," or "multi-word sequences" (Biber, Contes, & Cortes, 2004). The most common lexical bundle in any given text corpus is the one containing the most words (p. 150). The importance of lexical aggregates in written or oral discourse has been underlined by numerous corpus linguistic studies and publications, including those by Biber & Barbieri (2007), Biber, D. et al. (2004), Biber, Johansen, Leech, Conrad, & Finegan (1999), Cortes (2004), 2006, Hyland (2008)a, 2008b, Wray & Perkins (1999), and others (2000).

Much research has been done to reveal the corpora concept to describe and classify speech perception processes in various fields.The frequency of lexical sets has formed the focus of various research projects in various contexts (eg, Cortes, 2004, 2006; Hyland, 2008a, 2008b). Current analysis of lexical sequences related to the coronavirus demonstrates this shift.

Many studies have used the corpora to analyze and describe the development of pronunciation across various themes. Lexical aggregates have been the subject of several studies in various fields (eg, Cortes, 2004, 2006; Hyland, 2008a, 2008b). This trend is being investigated by current research into the frequency with which lexical collations appear in scientific papers related to coronavirus disease. A term is only eligible for inclusion in the lexical bundle if it meets certain requirements.

The same concept is Biber et al.-esque word combinations must have some sort of statistical likelihood (1999). The principle of responsibility is used to describe the analysis of whether a person is responsible for his behavior or not. For study purposes, a lexical bundle is defined as a combination of

words that appear at least three times in different texts, with a word count of 40,000 or less being considered the minimum for launching a company. For this analysis, we modified the criteria from Biber and Barbieri to account for the smaller corpus (2007).

Classifying phrases from two to six words is a common use for corpus software (Ulfa & Muthalib, 2020). However, three-word strings can be found in all word sets and are recognized in almost all word sets (Mehta et al., 2021). The reason is that collocations are often specified as word pairs. Due to the fact that string length varies with the size of the corpus being studied, researchers can apply software for word set adjustment to account for this fact. The main aim of our study is to formulate three- and four-word string hypotheses about the phenomenon under study that can be tested.

II. METHOD

Word pairs can be classified into small groups of two to six using corpus software (Ulfa & Muthalib, 2020). However, groups of three words are common word groups found in all corpuses (Mehta et al., 2021). This occurs when two words or phrases are used frequently together. Because the strings vary depending on the size of the corpus, researchers can use software to change the word cluster groupings for analysis. To facilitate further investigation of this topic, the main aim of this work is to generate explorable three- and four-word strings.

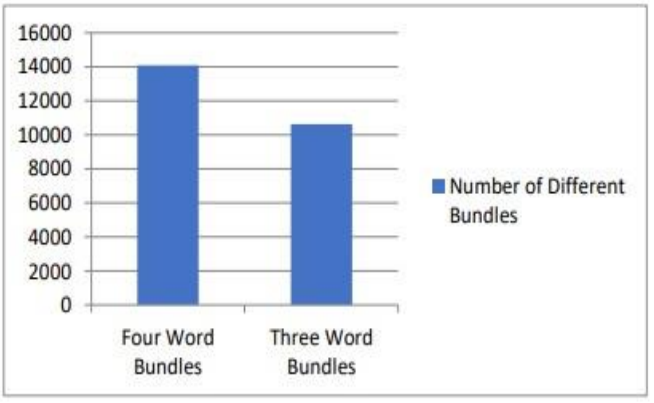
For discourse dataset analysis, there are several processors and tools (Al-Hamzi, 2020). But to find frequently occurring word combinations, we used the Windows Antconc 3.5.8 tool (available at http://www.laurenceanthony.net). It is very important to emphasize that AntConc is a free, cross-platform tool for corpus linguistic analysis and data-driven learning (Anthony, 2019).

We apply textual data as the main data source for our research. It involves buying and selling digital goods online. The corpus should be developed and organized in such a way as Biber's (2006, pp. 23-31). 204 articles were used to create a medical science-specific corpus for this study.

III. RESULT AND DISCUSSION

Below is the chart of the frequency of lexical bundle occurrences between three-word and four word bundles found in the corpus using AntConc software.

Figure 1: The frequency of three- and four-word clusters in the corpus



As shown by the figures in the section above, there were variations in the frequency of three-word and four-word bundles. The graph indicates that 33211 tokens contain over 10,000 unique four-word clusters (total number of words). Relatively, the 33215 tokens include more than 3000 distinct three-word bundle types. Not all bundle types were considered in this example since some did not meet the criteria set out by Biber et al (1999). The sequences must be avoided using exclusion criteria in order to filter the criteria. Exclusion criteria had to be employed in order to make administration easier since AntConc generated so many lexical bundles that needed to be investigated (as seen in Figure 1). The lexical bundles that AntConc produced after using the exclusion criteria are listed below (see Table 1). The lexical bundles found in the articles were listed from most frequent to least common.

Table1. The Top 10 Three Lexical Bundles “OF”

No.	List of Three Bundles	Frequency	Range	Article + Hits
1.	The epidemic of	1088	101	27
2.	Similarity of the	250	80	18
3.	Fatal case of	242	80	20
4.	immune monitoring of	227	70	25
5.	Potent neutralization of	160	58	19
6.	Structure of the	156	57	35
7.	the replication of	152	40	12
8.	International Journal of	141	27	12
9.	Clinical Conditions of	140	73	20
10.	zoonotic jump of	121	29	10

The bulk of three-word bundles in the corpus were noun phrases with embedded "of" fragments, as seen in Table 1. (7 instances over 7 articles). The pandemic of happened six times in six different publications, followed by Fatal case of and immunological surveillance of, which each occurred four times, to place the construction of prepositional phrase in second place in terms of frequency.

In the meanwhile, other expressions may be found in AntConc's findings for the four-word bundles. This is seen in Table 2 below:

Table 2. The Top 10 Four Lexical Bundles “OF”

No.	List of Four Bundles	Frequency	Range	Article + Hits
1.	WHO. Clinical management of	223	76	32
2.	Structure and genome of	112	45	30
3.	Laboratory confirmed cases of	88	34	15
4.	Characteristics of patients with	84	25	18
5.	of patients infected with	78	52	12
6.	The clinical characteristics of	61	38	30
7.	humidity for the survival of	55	36	20
8.	laboratory-confirmed cases of	50	35	13
9.	human transmission of a	48	38	27
10.	different regions of the	46	16	22

There were four-word bundles, but Table 2 revealed that they seldom ever occurred together in the corpus. As can be seen, the compositions in the corpus with the most distinctive features resemble noun phrases with embedded fragments (eg WHO clinical management, structure, and genome, occurring 6 and 4 times, respectively). The next phrase is "laboratory verification instances of," which appears six times and contains the copula form to + noun phrase or adjective phrase.

The previous discussion illustrated that depending on the type, scope, and list of studies, a different order of lexical bundles is used. The choice of intersection points for classifying lexical bundles also has an impact on the output of word combinations generated by the program.

A total of 10631 unique bundle kinds and 33215 tokens. A further 33211 tokens include 14077 distinct sorts of four-word bundles.

As stated by Biber and Barbieri (2007), who claimed that for a smaller sub-corpus of roughly 50,000 words, the word combinations must appear in three separate texts, the findings demonstrated that lexical sets were discovered in significant numbers as a consequence of employing tiny intersecting points as a consequen.

No	Lexical bundle	Type	Article & hits		Remarks (Tittle)
1	*"WHO. Clinical managemen t of"*	**Collocation:** Pre + N **Lexical bundle:** 4 words Bundle	10 Articles Total 8 hits	1.	Clinical features of patients infected with 2019 novel coronavirus in wuhan.
				2.	Clinical molecular and epidemological characterization of the sars-sov
				3.	Epidemiological and clinical characteristics of 99 case of 2019 novel coronavirus pneumonia in wuhan
				4.	Human Coronavirus Infection
				5.	Mysterious virus a review on bahavior and treatmen approaches of the novel coronavirus
				6.	1-s2.0-S0140673620301835-main 1 hits
				7.	1-s2.0-S0140673620302117-main 1 hits
				8.	1-s2.0-S0188440920305506-main 1 hits
				9.	1-s2.0-S0732889320304715-main 1 hits
				10.	3-s2.0-B9780128012383116344-main 3 hits
2	"*Structure and genome of*"	**Collocation:** Pre + N **Lexical bundle:** 4 words Bundle	4 Articles Total 8 hits	1	Clinical molecular and epidemological characterization of the sars-sov-
				2	Coronavirus disease 2019 current status and future perspectives.
				3	1-s2.0-S0732889320304715-main 1 hits
				4	1-s2.0-S0924857920301011-main 1 hits

				1	Human Coronavirus Infection
				2	3-s2.0-B9780128012383116344-main

3	"*Laboratory confirmed cases of*"	**Collocation:** Pre + N **Lexical bundle:** 4 words Bundle	12 Articles Total 8 hits	3	An overview of coronavirus including the SARS-2 coronavirus
				4	Coronavirus disease 2019
				5	Current epidemiological and clinical features of covid-19
				6	Current epidemiological and clinical features of covid-19
				7	-s2.0-S016344532030222X-main
				8	1-s2.0-S016344532030222X-main
				9	1-s2.0-S0887796320300146-main
				10	-s2.0-S0924857920300674-main.
				11	1-s2.0-S0924857920300674-main
				12	1-s2.0-S2352081720300398-main
4	"*Characteristics of patients with*"	**Collocation:** Pre + N **Lexical bundle:** 4 words Bundle	12 Articles Total 12 hits	1	An overview of coronavirus including the SARS-2 coronavirus
				2	Clinical molecular and epidemological characterization of the sars-sov-
				3	Comparison of hospitalized patients with ards caused by covid-19 and h1n1
				4	Correlation between viral RNA shedding and serum antibodies in covid-19 patients
				5	Correlation between viral RNA shedding and serum antibodies in covid-19 patients
				6	Therapeutic management of patients with COVID-19
				7	Therapeutic management of patients with COVID-19
				8	1-s2.0-S0012369220305584-main
				9	1-s2.0-S0732889320304715-main.
				10	1-s2.0-S2352081720300398-main
				11	1-s2.0-S2590088920300251-main
				12	1-s2.0-S2590088920300251-main
				1	A review on possible modes of action of chloroquine

5	"*of patients infected with*"	**Collocation:** Pre + N **Lexical bundle:** 4 words Bundle	15 Articles Total 108 hits	2	Clinical course of severe and critical covid-19 in hospitalized pregnancies
				3	Clinical features of patients infected with 2019 novel coronavirus in wuhan.
				4	Clinical molecular and epidemological characterization of the sars-sov-2
				5	Coagulation disorders in coronavirus infected
				6	omparison of hospitalized patients with ards caused by covid-19 and h1n1
				7	omputational screening of antagonists againt the sars-cov-2
				8	Consistency analysis of covid-19 nucleic acids test
				9	Coronavirus disease 2019 and pregnancy
				10	Coronavirus disease 2019 current status and future perspectives
				11	Coronavirus disease 2019
				12	COVID-19 A critical care perspective informed by lessons
				13	COVID-19 and SARS-Cov-2 infection
				14	Covid-19, sars and mers.
				15	Current epidemiological and clinical features of covid-19
6	"*Similarity of the*"	**Collocation : adj+ n Lexical bundle:** 3 words bundle	6 Articles Total 6 hits	1	ecent progress in understanding 2019 novel coronavirus
				2	omputers and viral disease
				3	1-s2.0-S092485792030100X-main
				4	1-s2.0-S0010482520300627-main
				5	Coronavirus disease 2019 and pregnancy.
				6	1-s2.0-S0002937820301976-main
	"*immune*	**Collocation : adj+ n Lexical**	4 Articles	1	Immunology of covid-19 current state of the science
				2	1-s2.0-S1074761320301837-main

7	*monitoring of*"	**bundle:** 3 words bundle	Total 6 hits	3	A review on possible modes of action of chloroquine
				4	1-s2.0-S0924857920301916-main
8	"*Potent neutralization of*"	**Collocation : adj+ n Lexical bundle:** 3 words bundle	13 Articles Total 16 hits	1	Immunology of covid-19 current state of the science.
				2	Coronavirus pandemics
				3	Coronavirus disease 2019 current status and future perspectives
				4	Coronavirus pandemics
				5	1-s2.0-S1074761320301837-main
				6	1-s2.0-S0024320520305865-main
				7	1-s2.0-S0024320520305865-main
				8	1-s2.0-S1074761320301837-main
				9	-s2.0-S0024320520305865-main
				10	1-s2.0-S1074761320301837-main
				11	1-s2.0-S0924857920301011-mai
				12	-s2.0-S1074761320301837-main.
				13	1-s2.0-S1074761320301837-main
9	"*zoonotic jump of*"	**Collocation : adj+ n Lexical bundle:** 3 words bundle	6 Articles Total 6 hits	1	Genotyping coronavirus sars-cov-2
				2	Comparison of hospitalized patients with ards caused by covid-19 and h1n1
				3	Coagulation disorders in coronavirus infected
				4	1-s2.0-S0888754320303189-main
					Clinical molecular and epidemological
				5	characterization of the sars-sov-2
				6	Clinical features of patients infected with 2019 novel coronavirus in wuhan
10	"*Clinical Conditions*	**Collocation**	4	1	Clinical molecular and epidemological characterization of the sars-sov-2.
	of"	**: adj+ n Lexical bundle:** 3 words	Articles Total 6 hits	2	covid-19 transmission prevention and potential therapeutic opportunities
				4	1-s2.0-S0732889320304715-main
				5	Global coronavirus disease 2019

		bundle			

The statement made by Ucar (2017) that these articles often contain shorter lexical bundles than longer ones is supported by this. Additionally, lexical bundles are organized in keeping to how frequently words appear together after the application of the aforementioned exclusion criteria (Salazar, 2011). The most common three-word clusters were caused by (6 occurrences in 6 article in the corpus). There were a number of future clause fragments as well as verb/adjective + clause fragments. The four-word phrase "the benefits of being," it was found, was used the most often.

The fullness or incompleteness of lexical bundles depends on how they are structured. The weak structure is metaphorized by clauses and phrases. This inadequate design occurs 38,749 times, or 78.7% of the time, within the bundles. There are entire structures as well as partial ones, and the latter often take the form of phrases. Noun, preposition, verb, adjective, and clause-based bundles are the five main types of lexical bundles. In academic publications, the bulk of them (49.2%) are lexical bundles based on clauses. The use of passive verbs and sentence fragments is a feature of the genre.

IV. CONCLUSION AND SUGGESTION

From the discussions in the sections above, it can be determined that each register, whether spoken or written, has unique features with respect to the use of certain combinations. The corpus, which contains 204 medical articles and more than 60,000 words related to linguistics, was used to produce tens of thousands of bundles. Exclusion criteria were then used to filter these bundles. This investigation's primary goal was to establish the frequency of occurrences. Frequency was used to determine a corpus's propensity to employ lexical bundles. Four-word bundles were much more common than three-word bundles, according to the data. The exclusion criteria were applied to the list bundles generated by AntConc, and the sequence with the greatest frequency in terms of three-word bundles and WHO was the pandemic of. employing four-word clinical management statements. Each consisted of a noun phrase and a sentence fragment. In order to understand how lexical bundle clusters are employed in connection to their particular roles, students must also be be able to recognize them. Teachers in this case must be aware of this problem and include corpora and corpus-based learning into their lesson plans in order for students to identify clusters that are unique to the corpora being utilized.

The study's scope was narrow since it only focused on three- and four-word word bundles that were identified in a smaller corpus. Furthermore, because this study only looked at the corpus linked to linguistics, it is predicted that future research would be able to look at a wide range of subjects to provide insight on the lexical bundles utilized in many academic areas.

REFERENCES

Al-Hamzi, A. M. S. (2020). Corpus Linguistics and Corpus-Based Research and its Implication in Applied Linguistics: A Systematic Review. *Journal of Linguistics and Education*, *10*(2), 176–181.

Anthony, L. (2019). *AntConc (Version 3.5.8) [Computer Software]*. Waseda University.

Biber, D.,], Conrad, S., & Cortes, V. (2004). Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371– 405.

Biber, D., & Barbieri, F. (2007). 'Lexical bundles in university spoken and written registers.' *English for Specific Purpose*, *26*, 263–286.

Kashiha, H., & Chan, S. H. (2013). An Exploration of Lexical Bundles in Academic Lectures: Examples from Hard and Soft Sciences. *Journal of Asia*, *10*(4), 133–161.

Mehta, V., Bawa, S., & Singh, J. (2021). Weclustering: Word Embeddings Based Text Clustering Technique for Large Datasets. *Springer*, *7*, 3211–3224.

Meunier, F., & Granger, S. (2007). *Phraseology in Foreign Language Learning and Teaching*. John Benjamins Publishing.

Salazar, D. J. L. (2011). *Lexical Bundles in Scientific English: A Corpusbased Study of Native and Non-Native Writing (Doctoral Dissertation)*. Universitat de Barcelona.

Ucar, S. (2017). A Corpus-Based Study on The Use of Three-Word Lexical Bundles in The Academic Writing by Native English and Turkish Nonnative. *English Language Teaching*, *10*(12).

Ulfa, N., & Muthalib, K. A. (2020). Writing, Lexical Bundles in Students' Essay. *English Education Journal (EEJ)*, *11*(3), 367–379.