World Journal of Corpus Linguistics

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

IDENTIFYING NEW VOCABULARIES AND ITS MEANING IN COVID-19 TEXTS THROUGH ANTCONTC 3.5.8 (IN THE PERSPECTIVE OF CORPUS LINGUISTICS)

Kiki Nurwahyuni
Magister Ilmu Linguistik, Warmadewa University
kikinw4@gmail.com

Abstract

One of the problems that have arisen by all people around the world is the spread of Covid-19 cases which have occurred more than recently. With this case, the presence of new vocabularies that are mostly not implemented by the wider community requires a special method to explain these words and their meanings. In linguistics perspective, this case is a advantage in the use of words and adding words in the dictionary because it is able to increase knowledge related to language for the community. The aim of the study is to identif new vocabularies related to Covid-19 and their meanings and classify words based on their class types. 204 articles regarding Covid-19 were analyzed using the Antcontc 3.5. application through the corpus linguistics approach. The results of this study indicates that there are several new vocabularies that are not known by the public because they are not listed in the dictionary.

Keywords: Covid-19, Corpus Linguistics, New Vocabularies

Abstrak

Salah satu masalah yang menjadi perhatian seluruh masyarakat di seluruh dunia adalah penyebaran kasus Covid-19 yang semakin marak terjadi akhir-akhir ini. Dengan adanya kasus ini, hadirnya kosakata-kosakata baru yang sebagian besar belum diimplementasikan oleh masyarakat luas membutuhkan metode khusus untuk menjelaskan kata-kata tersebut beserta maknanya. Dalam perspektif linguistik, kasus ini merupakan sebuah keuntungan dalam penggunaan kata dan penambahan kata dalam kamus karena mampu menambah pengetahuan terkait bahasa bagi masyarakat. Tujuan dari penelitian ini adalah untuk mengidentifikasi kosakata baru yang berhubungan dengan Covid-19 beserta maknanya dan mengklasifikasikan kata berdasarkan jenis kelasnya. Sebanyak 204 artikel mengenai Covid-19 dianalisis menggunakan aplikasi Antcontc 3.5. melalui pendekatan linguistik korpus. Hasil dari penelitian ini menunjukkan bahwa terdapat beberapa kosakata baru yang belum diketahui oleh masyarakat karena tidak tercantum dalam kamus.

***Kata kunci**:* Covid-19, Linguistik Korpus, Kosakata Baru

INTRODUCTION

The presence of Covid-19 not only affects the economy, health and education but also in terms of language. At present, the emergence of new vocabularies related to Covid-19 is something that must be followed up because it is able to enrich linguistics. The use of this new vocabularies can be used according to certain conditions or situations. It is very important to know the meaning of vocabularies to improve knowledge in writing, reading and speaking skills. Tarigan (2011) states that the quality of a person's language skills clearly depends on the quality of his vocabulary. While Kridalaksana (2003), vocabulary is (1) a language component that contains all information about the meaning and its use in language; (2) the richness of words that are owned by the speaker or writer of a language. Vocabulary is all the words that have been heard that belong to a speaker who has arranged like a dictionary and equipped with a short and complete explanation so that it is easily read by readers (Hariyanto, 2010). Moreover, it can be said that the main characteristic of person who have a lot of knowledge, such as flexibility in speaking and writing is that the person has and mastered more vocabularies.

Currently, along with the development of new vocabularies, there is sub-disciplines of linguistics, namely corpus linguistics which is able to collect data (vocabularies) quickly without having to collect data one by one. This corpus linguistics has a role to facilitate human work related to language, especially

in making dictionary. The corpus is defined based on its form and purpose (Hunston, 2002). Regarding form, linguists describe a corpus as a a collection of examples of naturally occurring language, consisting of everything that is relating to text, ranging from a few sentences to a set of written texts or cassette tapes, collected for linguistic studies. Meanwhile, With regard to objectives, the corpus is planned and designed for several purposes linguistics. The specific purpose of the corpus design determines the text selection by consider the principles of representation and balance (Reppen & Simpson-Vlach, 2020).

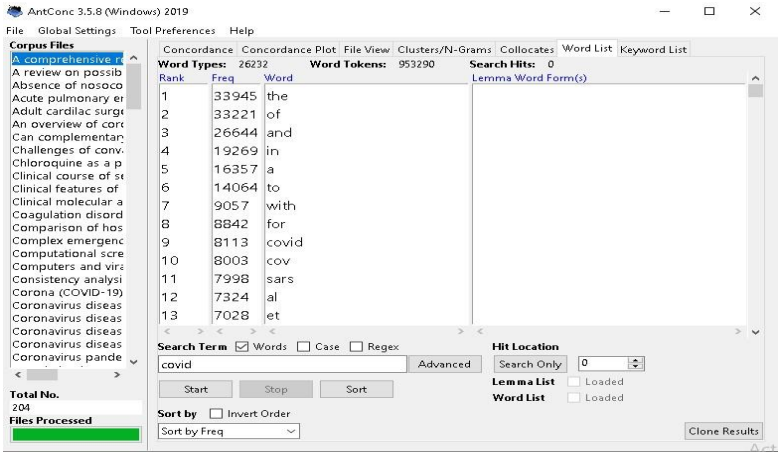
Corpus linguistics is a relatively recent branch of linguistics, made popular since the advent of personal computers in the 1990s. Put simply, corpus linguistics is the study of language based on examples of real life language use' (McEnery and Wilson, 1996). The word corpus comes from the Latin word for body; its plural is corpora (although corpuses is perfectly acceptable, if a little more difficult to pronounce). A corpus is therefore a 'body' of language, or more specifically, a (usually) very large collection of naturally occurring language, stored as computer files (Paul Baker, 2010). A corpus (plural corpora) is a body of material (textual, graphic, audio, and/or video) upon which some analysis is based. Several disciplines make use of corpora: linguistics of course, but also literature, philosophy, art, and science. A corpus is not just a collection of linguistically relevant material. For that collection to count as a corpus, it has to meet a number of criteria: sampling, balance, representativeness, comparability, and naturalness (Guillaume Desagulier, 2017). To maximize performance in data processing, an application is needed, namely Antcontc.AntConc has quickly become popular among researchers, teachers, and language learners due to easy-to-use interface, its rich set of features, multiplatform support, and the fact that it is freeware. AntConc is as frequently used to analyze the corpora as is WordSmith Tools. Both programs offer access to raw data to create corpora. Moreover, many corpus linguists have a great deal of confidence in Antconc for their work (Anthony, 2013). The aim of the study is to identif new vocabularies related to Covid-19 and their meanings and classify words based on their class types.

METHOD

This research used a qualitative descriptive approach using linguistic corpus theory, namely a problem-solving procedure by describing the state of the research object based on the facts that appear as they are (Sugiyono, 2014: 205). The data source used in the form is 204 texts related to Covid-19. With the aim of identifying new vocabularies related to the current situation, namely the Covid-19 pandemic and its meaning as well as classifying vocabularies based on the word class in order to put in the dictionary. It also has relation to lexicography. Lexicography typically makes use of the ostensive approach in the sense that inclusion in a dictionary provides an incremental defining of the area. This is particularly the case with the glossary or encyclopedic type of dictionary which describes the key terminology in an area of interest (for example politics, biology, applied linguistics) and by doing so defines it.

RESULT AND DISCUSSION

Before analyzing data in the form of new vocabulary about covid 19. The first thing to do is look at the number of words with a data source in the form of 204 articles.



Picture 1. The result of number of words

The image above is a display using the antcontc application to find out the number of words in 204 articles about Covid-19. In analyzing corpus data, first we need to know what words, how many words, and how often the words appear in the text, this is one of the functions of the Antcontc application. Word Type is 26232 words, with the word "the" ranking the highest because it is the word used most often in the text. We can also see the number of word tokens or the total words, which is 953290 words.

Table 1. Eleven vocabularies related to Covid-19

No.	New Vocabularies	Concordance Hits	Collocate Tokens
1.	Outbreak	913	9182
2.	Pandemic	750	7500
3.	Epidemic	456	4560
4.	Symptoms	988	9880
5.	Suspect	4	40
6.	Lockdown	56	560
7.	Droplet	68	680
8.	Incubation	173	1730
9.	Spike	808	8080
10.	Cluster	115	1150
11.	Asymptomatic	317	3170

Data 1. Vocabularies with one cluster

Table 2. Nine vocabularies related to Covid-19

No.	New Vocabularies	Concordance Hits	Collocate Tokens
12.	Community Spread	10	100
13.	Social Distancing	67	670
14.	Contact Tracing	26	260
15.	Self-quarantine	8	80
16.	Physical Distancing	4	40
17.	Swab Test	4	40
18.	Rapid Test	10	100
19.	Local Transmission	10	100
20.	Imported Case	4	40

Data 2. *Vocabularies with two clusters*

Table 3. Four vocabularies related to Covid-19

No	New Vocabularies	Concordance Hits	Collocate Tokens
21	SARS-COV-2	4269	39318
22	Mers	2057	20570
23	Covid-19	7903	75116
24	PCR	533	5330

Data 3. *Acronym*

Based on the results of the analysis using the Antcontc application, there are 24 new vocabulares related to Covid-19. The first data contained 11 vocabularies with one cluster. In this analysis, the word with the highest frequency is "Symptoms" with 988 concordance hits. The second data contained 9 vocabularies with two clusters. In this analysis, the word with the highest frequency is "Social Distancing" with 67 concordance hits. The third data is the acronym there are four vocabularies and the most frequent word is "Covid-19" with 7903 concordance hits.

From the data above it can be concluded that from all the vocabularies that is appeared, the highest concordance is Covid-19 with 7903 concordance hits and 75116 collocate tokens.

Identifying the meaning and word class

No.	New Vocabularies	Word Class	Definition
1.	Outbreak	(n)	*a sudden increase in the spread of the disease*
2.	Pandemic	(adj) (n)	*widespread (about disease) in an area, continent, or around the world widespread epidemic disease*
3.	Epidemic	(adj)	*a condition that arises when the spread of disease only occurs in one particular area for a long time and is constant*
4.	Symptoms	(n)	*changes or special conditions in the body that indicate a disease; symptoms of the disease:*
5.	Suspect	(n)	*people who are suspected of having symptoms of certain diseases*
6.	Lockdown	(adv)	*a situation that prohibits residents from entering a place due to an emergency*
7.	Droplet	(n)	*liquid or drool that a person secretes from the nose or mouth when sneezing, coughing, or even speaking.*
8.	Incubation	(n)	*the period from when the cause of the disease enters the body (the time of infection) to the time the disease occurs; budding period*
9.	Spike	(n)	*the form of a bulge in the Corona virus*
10.	Cluster	(n)	*one group with one health event.*
11.	Asymptomatic	(adv)	*how the corona virus is transmitted by people who have no symptoms (people without symptoms) and will never get symptoms from the infection.*
12.	Community Spread	(adv)	*the spread of the disease in an area it is not yet known exactly when and how a person becomes infected.*
13.	Social Distancing	(v)	*an action that aims to prevent sick people from coming into close contact with other people to reduce the chance of spreading the virus.*
14.	Contact Tracing	(v)	*the concept of detecting people with a high potential for contracting the virus from Covid-19 positive patients.*
15.	Self-quarantine	(v)	*activities to stay in a place that are carried out for 14 days to see someone potentially sick or can transmit the COVID-19 virus to other people*

16.	Physical Distancing	(v)	*recommended physical restrictions to prevent the spread of the Corona virus.*
17.	Swab Test	(v)	*a method for detecting the corona virus by wiping the nasopharyngeal cavity and / or oropharynx using a tool such as a special cotton swab.*
18.	Rapid Test	(v)	*method of examination or test results quickly obtained. This examination uses a cartridge device to see the presence of antibodies in the body when there is a viral infection.*
19.	Local Transmission	(adj)	*someone who is positive of corona virus but has no history of traveling to an affected country and has no contact with a positive patient.*
20.	Imported Case	(adj)	*cases of the spread of the virus that occur while abroad, or when someone travels to a country that is currently infected with the corona virus.*
21.	SARS-COV-2	(n)	*the virus that causes COVID-19*
22.	Mers	(n)	*respiratory disease that occurs due to the Corona virus*
23.	Covid-19	(n)	*illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus*
24.	PCR	(v)	*SARS Co-2 virus examination method by detecting viral DNA*

Table 4. Meaning and Word Class through Covid-19 Case.

In the table above, there are 24 new vocabularies related to Covid-19 which are divided into four types of word classes, namely noun, adverb, adjective and verb. Based on the results of the highest frequency analysis by nouns, namely 11 new vocabularies verb amounting to 7 vocabularies, adjective amounting to 4 vocabularies and the last is adverb which amounting to 3 vocabularies. But on the other hand, the word "Pandemic" has two types of word classes, namely adjective and noun.

CONCLUSION

In conclusion, this study has demonstrated the utility of corpus linguistics in analyzing the use of conjunctions in a "social" text. The findings reveal that conjunctions such as "but," "when," "for," "or," "because," and "yet" play significant roles in structuring the text and conveying relationships between ideas. The most frequently used conjunction, "but," was primarily employed to introduce contrasts, while "when" was used to indicate time or condition. Other conjunctions served various functions, such as presenting alternatives, indicating reasons, or introducing contrasts.

The analysis also highlighted the importance of understanding the contextual usage of conjunctions, as their functions can vary depending on the surrounding text. By examining concordance lines and clusters, this study was able to provide a detailed understanding of how conjunctions contribute to the coherence and cohesion of the "social" text.

Overall, this research underscores the value of corpus linguistics in linguistic analysis, offering a systematic approach to understanding the role of conjunctions in text construction. Future studies could expand on this research by examining conjunctions in different types of texts or by exploring other linguistic features using corpus-based methods.

REFERENCES

Anthony, L. (2013). Developing AntConc for a new generation of corpus linguists. Proceedings of the Corpus Linguistics Conference (CL 2013) July 22-26. Lancaster University, UK, 14-16.

Guillaume Desagulier. (2017). Corpus Linguistics and Statistics with R Introduction to Quantitative Methods in Linguistics. Springer International Publishing Press. Hariyanto, B. (2010). Istilah-Istilah dalam Chatting (Sebuah Analisis Sosiopragmatik). Jurnal Adabiyyat, 9(2), 294-326.

Hunston, S. (2002). Corpora in Appllied Linguistics. hlm. 1–24.

Kridalaksana, H. (1993). Kamus Linguistik. Jakarta: Gramedia Pustaka Utama. McEnery, T. and Wilson, A. (1996). Corpus Linguistics. Edinburgh: Edinburgh University.

Paul Baker. (2010). Sociolinguistics and Corpus Linguistics. Edinburgh University Press Ltd.

Reppen, R., Simpson-Vlach, R. (2020). Corpus Linguistics. Dalam N. Schmitt & M. P. H. Rodgers (Eds.), An Introduction to Applied Linguistics (Edisi ke-3) hlm. 91-108. New York: Routledge

Sugiyono. (2014). Metode Penelitian Kuantitatif, Kualitatif. R&D. Bandung:Alfabeta.

Tangan, H.G.(2011). Pengajaran Kosakata. Bandung: Angkasa.