COMPOUND WORDS IN COVID-19 ARTICLES

David Perwira Nasrani Laia
Magister Ilmu Linguistik Universitas Warmadewa
pwrdave14@gmail.com

Abstract
Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. This research is motivated by the phenomena of Covid-19 on the language. Taking into the phenomena, this research aims to identify the compound words in Covid-19 articles. This research uses corpus linguistics approach. The method used is the qualitative research. The method of collection of data is observation. The research articles analyzed 71 research articles. The result shows that the researcher uses formal method. The open compound words are found on Community transmission, Mental health, Drug maker, Face mask, Physical distancing, Hand sanitizer, Social distancing, Health insurance, and Life science. While the closed compound words are found on Coronavirus, Lockdown, Droplet, Microbiology, Outbreak, Fallout, Sunbathing, Shortage, Handwash, Healthcare, Sunbathe, Sunlight, Worldwide, Lockdown, Coronacoma, and Coronaspeck. The last is hyphenated compound words which found on Self-isolation, Face-to-face, Well-being, Long-term, Self-isolate, Self-quarantine, Blood-brain-barrier, and Micro-needle.

Keywords: Compound words, Covid-19, Open compound words, Closed compound words, Hyphenated compound words

Abstrak
Penyakit coronavirus (COVID-19) adalah penyakit menular yang disebabkan oleh coronavirus yang baru ditemukan. Penelitian ini dilatarbelakangi oleh fenomena Covid-19 pada bahasa. Melihat fenomena tersebut, penelitian ini bertujuan untuk mengidentifikasi kata majemuk dalam artikel Covid-19. Penelitian ini menggunakan pendekatan linguistik korpus. Metode yang digunakan adalah penelitian kualitatif. Metode pengumpulan data adalah observasi. Artikel penelitian menganalisis 71 artikel penelitian. Hasil penelitian menunjukkan bahwa peneliti menggunakan metode formal. Kata majemuk terbuka ditemukan pada Community transmission, Mental health, Drug maker, Face mask, Physical distancing, Hand sanitizer, Social distancing, Health insurance, dan Life science. Sedangkan kata majemuk tertutup ditemukan pada Coronavirus, Lockdown, Droplet, Microbiology, Outbreak, Fallout, Sunbathing, Shortage, Handwash, Healthcare, Sunbathe, Sunlight, Worldwide, Lockdown, Coronacoma, dan Coronaspeck. Terakhir adalah kata majemuk dengan tanda penghubung yang terdapat pada Self-isolation, Face-to-face, Well-being, Long-term, Self-isolate, Self-quarantine, Blood-brain-barrier, dan Micro-needle.

Kata kunci*: Kata majemuk, Covid-19, Kata majemuk terbuka, Kata majemuk tertutup, Kata majemuk dengan tanda penghubung.*

I. INTRODUCTION

Currently, the world is facing and struggling with a new virus that has been announced as a global pandemic by the World Health Organization (WHO) on March 11, 2020 that is the coronavirus disease, COVID-19. After first found as an unknown pneumonia disease on 31 December 2020 in Wuhan, China, according to the Situation Report - 116 published by WHO on 15 May 2020, a total of 4.338.658 cases have been confirmed in more than 200 countries. As the virus progresses, every day, the world's community is introduced to terms that may sound unfamiliar. After going through the WHO official website, the terms found are commonly constructed in a form of compound words (Simatupang and Supri, 2020).

At the end of 2019, in December, to be precise, the world was shocked by the news of the emergence of a pneumonia outbreak with no known exact cause. Today the world is facing and struggling with a new virus that has been announced as a global pandemic by the World Health Organization (WHO) on March 11, 2020 that is the coronavirus disease, COVID-19. Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness (Li Q, Guan X, et al, 2020).

Commonly, a compound or compound word is regarded as a phrase that is constituted by the combining of two words. It is in line with McManis et al., (1987) stating that a compound is a word formed by the combination of two independent words. On the contrary, according to Spencer and Zwicky (2001), a compound is a word that consists of two or more words. Furthermore, Akmajian et al., (2001) explain that "compounds are not limited to two words". Similarly, Leech (1974) emphasizes a similar definition in which he states a compound is a word which contains two or more other words. In this regard, as this study progresses, the term compound is used to refer to the word combination that consists of two or more words.

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. The best way to prevent and slow down transmission is to be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol based rub frequently and not touching your face.

The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it's important that you also practice respiratory etiquette (for example, by coughing into a flexed elbow). In simple terms, corpus linguistics can be defined as the study of "the compilation and analysis of corpora" (Cheng 2012: 6), which are large collections of "naturally occurring language texts chosen to characterize a state or a variety of language" (Sinclair 1991: 171). According to Hunston (2002), even though corpus linguistics is a relatively new field, it has revolutionized language studies because it has provided new ways of analyzing and describing the use of language. The author emphasizes the fact that corpora consist of texts stored in an electronic format, which enables researchers to use special software (called concordancers) to conduct automatic searches and gain insights into the structure and regularity of naturally occurring language.

Corpus linguistics is any kind of language enquiry which relies on corpus data or corpus evidence to answer questions about language, including those related to the description of a language variety, theories of language, how language is learnt, and how language is used in specific contexts. Corpus linguists compile and investigate corpora, and so corpus linguistics is the compilation and analysis of corpora. In corpus linguistics, the term is used differently – it refers to a collection of samples of language use with the following properties:
  a. the instances of language use contained in it are authentic;
  b. the collection is representative of the language or language variety under investigation;
  c. the collection is large.

In addition, the texts in such a collection are often (but not always) annotated in order to enhance their potential for linguistic analysis. In particular, they may contain information about paralinguistic aspects of the original data (intonation, font style, etc.), linguistic properties of the utterances (parts of speech, syntactic structure), and demographic information about the speakers/writers.

Corpus linguists compile and investigate corpora, and so corpus linguistics is the compilation and analysis of corpora. Corpus linguistics allows researchers to process millions of words (Velzen, Nanetti, and Deyn 2013). A corpus (plural: corpora) is a collection of written or spoken material, occurring naturally, stored on computer, and typically used to carry out some kind of linguistic

analysis. A corpus is a collection of texts that has been compiled for a particular reason. In other words, a corpus is not a collection of texts regardless of the types of texts collected or, if a variety of text types (i.e., genres) are in the corpus, the relative weightings assigned to each text type.

The software programmes Antconc were used for the analysis of the corpus. AntConc as an example, a freeware multiplatform tool created by Laurence Anthony in 2004. According to Baker (2019) that AntConc is a freeware, standalone, multiplatform corpus toolkit developed by the author in collaboration with some of the leading corpus linguists in the world. A compound or compound word is regarded as a phrase that is constituted by the combining of two words. Compound words can be written in three ways: as open compounds (spelled as two words, e.g., ice cream), closed compounds (joined to form a single word, e.g., doorknob), or hyphenated compounds (two words joined by a hyphen, e.g., long-term). Sometimes, more than two words can form a compound (e.g., mother-in-law).

A corpus (plural: corpora) is a collection of written or spoken material, occurring naturally, stored on computer, and typically used to carry out some kind of linguistic analysis. A corpus has three main characteristics:

a. It consists of naturally occurring texts; that is, the texts are not written specially to go into the corpus, but are collected from sources such as newspapers, novels, blogs or (transcribed) conversations.

b. The texts are selected to represent a language or a variety of a language (Francis 1982; Nelson 2000). This might be a geographical variety such as 'the English spoken in East Africa', or a register such as 'the English spoken in academic conferences', or the language of a specific group of people such as 'English used by students in year 3 in secondary schools in Kenya'. The texts are chosen to be as representative as possible of the chosen variety.

c. It is relatively large. This means that there are so many texts in the corpus it would be difficult to

study them 'by hand and eye alone' (McEnery & Hardie 2012: 2). Instead, the corpus is stored on computer and accessed using specialized software. Most corpora consist of written texts, such as newspaper articles, academic journal articles, novels, blogs, text messages or student essays, and transcribed oral texts, such as speeches, conversations, or interviews. Increasingly, however, corpora incorporate visual material, such as pictures, photographs, or videos. This allows the range of languages studied to include signed languages used by deaf communities (Johnston & Schembri 2006); it also allows the study of written language to take account of accompanying pictures, and the study of oral languages to include the study of gesture (Adolphs & Carter 2013).

A corpus is a collection of texts that has been compiled for a particular reason. In other words, a corpus is not a collection of texts regardless of the types of texts collected or, if a variety of text types (i.e., genres) are in the corpus, the relative weightings assigned to each text type. A corpus, then, is a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative.

The term corpus has slightly different meanings in different academic disciplines. It generally refers to a collection of texts; in literature studies, this collection may consist of the works of a particular author (e.g. all plays by William Shakespeare) or a particular genre and period (e.g. all 18th century novels); in theology, it may be (a particular translation of) the Bible. In field linguistics, it refers to any collection of data (whether narrative texts or individual sentences) elicited for the purpose of linguistic research, frequently with a particular research question in mind (cf. Sebba & Fligelstone 1994: 769).

Corpora have usually been developed with the aim of electronically accessing linguistic forms in large language data bases. The software programmes Antconc were used for the analysis of the corpus. AntConc as an example, a freeware multiplatform tool created by Laurence Anthony in 2004. According to Baker (2019) that AntConc is a freeware, standalone, multiplatform corpus toolkit developed by the author in collaboration with some of the leading corpus linguists in the world. Unlike many of the standard tools, AntConc is under constant development, with new features and options added regularly in response to feedback and suggestions from some of the 70,000 or more users of the software. When two words are used together to yield a new meaning, a compound is formed. Compound words can be written in three ways: as open compounds (spelled as two words,

e.g., ice cream), closed compounds (joined to form a single word, e.g., doorknob), or hyphenated compounds (two words joined by a hyphen, e.g., long-term).

Sometimes, more than two words can form a compound (e.g., mother-in-law). The most common spelling quandary writers face is whether to write compounds as separate words, one word, or hyphenated words. An open compound word is created in cases when the modifying adjective is used with its noun to create a new noun. This isn't quite the same as a noun with a modifying adjective. We just use a space between the adjective and the noun, so sometimes it can be hard to identify as a compound; however, if the two words are commonly used together, it's considered to be a compound word. The examples of open compound words as follow: living room, full moon, real estate, dinner table, coffee mug. When adverbs ending in -ly combine with another word, the resulting compound is always spelled as two separate words; largely irrelevant, newly formed. Based on the phenomena, this research aims to identify the compound words in Covid-19 articles by applying the theories of corpus linguistics.

II. METHOD

This research consists of an approach and method of research. The approach used in this research is the approach of corpus linguistics. The qualitative research is the method used in this research. In addition, the method of collection of data is observation. Researcher observes the compound words in several research article which discuss about covid-19. The observation is done by using a software. The software is AntCont 3.5.8 which developed by Laurence Anthony. The research articles which analyzed are 71 research articles. The theory used in analysing the data is the theory of corpus linguistics. In presenting the research result, researcher uses formal method.

III. RESULT AND DISCUSSION

A compound or compound word is regarded as a phrase that is constituted by the combining of two words. There are three kinds of compound words found on the 71 research article of Covid-19, namely open compound words, close compound words, and hyphenated compound words. The compound words found on the research articles are classified as following.

Open Compound Words

An open compound word is created in cases when the modifying adjective is used with its noun to create a new noun. Based on the analysis, the open compound words found are presented below.

Table 1. Open Compound Words

MMNo	Compound words
Open compound words	Community transmission
	Mental health
	Drug maker
	Face mask
	Physical distancing
	Hand sanitizer
	Social distancing
	Health insurance
	Life science

From the research article of covid-19 which has been analysed, the open compound words which found consists of 9 (nine) compound words. The open compound words are Community transmission, Mental health, Drug maker, Face mask, Physical distancing, Hand sanitizer, Social distancing, Health insurance, and Life science.

Closed Compound Words

Closed compound words look like one word. Closed compound words are usually made up of only two words. Based on the analysis, the closed compound words found on the research article of Covid-19 are presented on this following.

Table 2. Closed Compound Words

Kinds of compounds	Compound words
Closed compound words	Coronavirus
	Lockdown
	Droplet
	Microbiology
	Outbreak
	Fallout
	Sunbathing
	Shortage
	Handwash
	Healthcare
	Sunbathe
	Sunlight
	Worldwide
	Lockdown
	Coronacoma
	Coronaspeck

Based on the analysis has been done, the closed compound words found consists of 16 (sixteen) closed compound words. The closed compound words are Coronavirus, Lockdown, Droplet, Microbiology, Outbreak, Fallout, Sunbathing, Shortage, Handwash, Healthcare, Sunbathe, Sunlight, Worldwide, Lockdown, Coronacoma, and Coronaspeck.

Hyphenated Compound Words
A compound adjective is hyphenated if placed before the noun it modifies, but not if placed after the noun. According to analysis which has been done on the several research articles of Covid-19, it found the hyphenated compound words as following.

Table 3. Hyphenated Compound Words

Kinds of compounds	Compound words
Hyphenated Compound Words	Self-isolation
	Face-to-face
	Well-being
	Long-term
	Self-isolate
	Self-quarantine
	Blood-brain-barrier
	Micro-needle

Based on the table (3) above, the hyphenated compound words which found on the research articles of Covid-19 are 8 (eight) hyphenated compound words such as Self-isolation, Face-to-face, Well-being, Long-term, Self-isolate, Self-quarantine, Blood-brain-barrier, and Micro-needle.

IV. CONCLUSION

Based on the research result above, it can be concluded that the compound words found on the several research article of covid-19 consists of three kinds of compound words. The compound words are open compound word, closed compound words, and hyphenated compound words. The open compound words are found on Community transmission, Mental health, Drug maker, Face mask, Physical distancing, Hand sanitizer, Social distancing, Health insurance, and Life science. While the closed compound words are found on Coronavirus, Lockdown, Droplet, Microbiology, Outbreak, Fallout, Sunbathing, Shortage, Handwash, Healthcare, Sunbathe, Sunlight, Worldwide, Lockdown, Coronacoma, and Coronaspeck. The last is hyphenated compound words which found on Self- isolation, Face-to-face, Well-being, Long-term, Self-isolate, Self-quarantine, Blood-brain-barrier, and Micro-needle.

REFERENCES

Akmajian, A., et al. (2001). An introduction to language and communication. MIT Press

Brezina, V. (2018). Statistics in Corpus Linguistics: A Practical Guide. New York: Cambridge University Press.

Cheng, W. (2012). Exploring Corpus Linguistics: Language in Action. Canada: Routledge.

Crawford, W. J., & Eniko, C. (2016). Doing Corpus Linguistics. New York: Routledge Taylor & Francis Group.

Deroey, K. L. B. (2014). "Anyway, the Point I"m Making Is: Lexicogrammatical Relevance Marking in Lectures." in Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora, edited by L. Vandelanotte, K. Davidse, C. Gentens, and D. Kimps. Amsterdam: Rodopi. Desagulier,

G. (2017). Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics. Switzerland: Springer International Publishing.

Hanks, P. (2012). "The Impact of Corpora on Dictionaries." in Contemporary Corpus Linguistics, edited by P. Baker. London: Continuum International Publishing Group.

Huntley, S. J., Michaela, M., Viola, W., Yves, V. G., Hui, Y., Rachel, S. D., & Marnie, L. B. (2018). "Analysing the Opinions of UK Veterinarians on Practice-Based Research Using Corpus Linguistic and Mathematical Methods." Preventive Veterinary Medicine 150, 60–69.

Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirusinfected pneumonia. N Engl J Med. 2020;382(13):1199–207.

Simatupang, E. C., & Supri, I. Z. (2020). Compound words that occur during the global pandemic covid-19: A morphosemantic study. English Review: Journal of English Education, 8(2), 291-298.

McManis, C., et al. (1987). Language files: Materials for an introduction to language. Advocate Pub. Group.

Spencer, A. (1999). Linguistics: An introduction. Cambridge: Cambridge University Press.