Linguistic Corpus Based Analysis: Collocation of Word Meaning "Acute" In the Corpus Covid-19

Fitri Ana Rahmayani

Department of Linguistic, Airlangga University, Surabaya, Indonesia

fitri.ana.rahmayani-2019@fib.unair.ac.id

Abstract

This study presents a corporate-based analysis and aims to look for collocations that follow the word acute by using Word Sketch in the COVID-19 corpus through Sketch Engine software to find out what words follow the acute word and how it plays a role in using it. This study uses qualitative methods based on data collected through the Sketch Engine. It aims to find out the word acute based on its collection by taking data from the COVID-19 corpus. Data source from this research is acute word collocation through software from Sketch Engine. The stage for data collection is to enter the word acute in the concordance feature through the Sketch Engine software, then select the collocation icon and select span -3 + 3 (to the left -3 and to the right 3) so that the top 3 appear in the word of acute. The results obtained from this study are, side by side collocation of acute is respiratory, syndrome, and severe. In this case, the syndrome appeared as much as 19,938, respiratory appeared as much as 30,752, and severe appeared as much as 18,089. So, a more significant word that appears alongside the word acute is respiratory, which acts as a noun and its use to clarify the meaning of the SARS virus name.

Keywords: Acute, Covid-19, Collocation, Corpus, Corpus Linguistic

Introduction

One of the news media that is more rapid in conveying information is mass media and social media. Mass media is one of the basic needs of society. Mass media plays a very important role in conveying effective information, either directly or indirectly. With the mass media, people can choose and get clear information. However, currently the polarization of the media, both print and electronic, is very pronounced. In fact, this condition can affect communities and have the potential to divide them.

At this time, it is often found that news or media in conveying information is confusing and unclear, so that it creates controversy in the surrounding community.

The media that often appears is about COVID-19. COVID-19 stands for Coronaviruses desease-19. This virus is a virus that originated in China and over time, this virus has spread throughout the world including Indonesia, with this news creating a lot of new vocabulary so that the data that appears in the COVID-19 media has various variations of data. In this case, mass media coverage with a large amount of data is an interesting thing to study. We can see the tendency of discourse that appears in news by analyzing data extensively. So far, the discourse data on news has been analyzed with only one method, namely extensive methods such as corpus, or analyzed using intensive methods such as discourse analysis only. With extensive analysis, we can analyze data with a very broad scope without having to analyze the data individually. In this study, the news about COVID-19 will be analyzed using the linguistic corpus method using collocation data in discourse study theory.

Corpus linguistics is a method that can analyze large amounts of data using computer technology as a tool (McEnery, 2012 in (Waskita, 2018)). Corpus or corpora are very useful in the world of language, both in teaching and research materials. Some of the fields that rely on the corpus include lexicography (dictionary preparation), grammar, sociolinguistics, translation, teaching, stylistics (language style science), dialectology, discourse studies, and historical linguistics. However, to carry out an in-depth analysis of the data, this method alone is not sufficient. Another method is needed that can analyze the data in depth. One of the in-depth analysis methods is deep discourse analysis and critical discourse analysis. A discourse that appears in the news will have a certain meaning, in accordance with the way the discourse is constructed in continuous news, this is also conveyed by (Hardts-Mautner, 2009; Baker, 2010 in (Waskita, 2018)) that with data analysis broad and at the same time deep, we will find a way how a discourse is constructed through linguistic construction. In addition, we can see how a certain meaningful discourse is based on a lot of corpus data.

The term corpus refers to a collection of textual material arranged for a specific purpose. Apart from being in the form of software, corporations are also available online which can be accessed for free at very large sizes such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the Michigan Corpus of American Spoken English (MICASE) (McEnery, 2006). But over time, there is software that is more practical and in it there is a lot of data that is more complex and very easy for researchers to find data to research, the software is Sketch Engine which can be found via the web http://sketchengine.eu/

The definition of the corpus itself, according to Sinclair (2005) in (Suhardijanto, T., Dinakaramani, 2018), is "a collection of pieces of language text in electronic form, selected according to external criteris to represent, as for as possible, a language or language variety as a source of data for linguistic research ". Therefore,

there is an emphasis on the electronic form or machine machine readable form of the corpus in this modern era. In designing a corpus, there are several things that need to be considered, namely the selection criteria, size, corus, data authenticity, storage media, and data manipulation (Suhardijanto, T., Dinakaramani, 2018). The selection criteria are the main objective in the preparation of the corpus.

Kridaklaksana (2009) in (Khoiriyah, 2018) explains that collocation is a constant association between words and words that are side by side in a sentence. Thus, collocation is the tendency for words to join regularly in a language. Collocation is very important to obtain fluency and naturalness of language, because collocation can interpret the whole meaning of the individual constituent elements. According to the explanation from (McEnery, T. and Hardie, 2012) in (Saraswati, 2019) that collocation as an abstraction at the syntagmatic level. Thus, it can be understood that collocation is a habit of arising from a lexeme. There are two important terms in the language of collocation, namely collocate and node (Sinclair, 1991 in (Saraswati, 2019)). Node is the word that is learned (keywords), while collocate is used for words that appear around the node.

According to (Fan, 2009) in his research on collocation learning by English learners, it is concluded that foreign English speakers use a smaller number of collocations than native speakers and have difficulty using collocations appropriately. Previous research that is in line with this, namely Sugeha & Nurfaida (2016) revealed that the word mother has a positive meaning and is closer to family than the word mother and tends to be associated with unpleasant things such as death, illness, wrong, ordinary and so on. Research conducted by (Artha, 2018) reveals that the adjective collocation that follows the word "Indonesia" changes every period of government and this change is motivated by the context of the situation.

From the description above, the researcher will analyze the word acute using software collocation data in the Sketch Engine application through the Linguistic Corps method in discourse study theory. In this case, the researcher wanted to know the collocation of the word "acute" in the COVID-19 data as well as the form of the word and its use in the context of discourse.

Methods

This study uses a qualitative method based on data collected from a software called Sketch Engine on the online site http://www.sketchengine.eu/. This research was conducted with the aim of knowing the word acute based on its collocation by taking data from the COVID-19 corpus. According to McEnery & Wilson (2004) in Sugeha & Nurfarida (2016) Corpus comes from several texts as a source of language and literature research. In this qualitative method, data analysis regarding the

similarities, differences and the context of the situation that affects the appearance of acute word collocation which will be analyzed further.

The data source in this study was the acute word collocation that appeared in sketchengine. The Sketch Engine is one of the corpus analytical text tools and there are three (3) main functions in the Sketch Engine, namely 1) Concordance, which is the basis for using the corpus which shows raw data to find out more details about where and how a word is used by a concordance; 2) Word Sketch, which is a page summary about grammar and collocation between one word and another; 3) Thesaurus, which is a compilation list of common collocations where the words that appear from each entry are searched for their closest meaning. The data used in this study is collocation data that is in the concordance feature in the Sketch Engine software.

In the data collection step, it is necessary to classify the use of the Sketch Engine application. The first step is to enter the word acute into the concordance then select the collocation icon and select -3 to the left and 3 to the right then select "go" after it appears then click download on it to download it in excel which will analyze the data in the corpus.

Result and Discussion

Figure 1. Results of the acute word in collocation data through the Sketchengine application

Figure 1 shows the data for acute word collocation which is 111,813. The collocation of the word acute shows that the word respiratory appears more after the word acute and the word severe appears before the word acute and the word syndrome appears after the word respiratory. From this statement, it can be seen the number of co-occurrences in the acute word collocation below:

Figure 2. Collocation of acute words in the Sketch Engine application

		Word	Cooccurrences ?	Candidates ?	T-score	MI	↓ LogDice	
1	☐	syndrome	19,938	54,800	141.05	9.84	11.94	...
2	☐	respiratory	30,752	209,444	174.89	8.53	11.62	...
3	☐	severe	18,089	112,078	134.16	8.66	11.37	...
4	☐	Severe	4,413	7,881	66.38	10.46	10.24	...
5	☐	phase	4,969	44,842	70.24	8.12	10.02	...
6	☐	distress	3,387	9,761	58.13	9.77	9.83	...
7	☐	injury	3,891	30,313	62.18	8.33	9.81	...
8	☐	chronic	4,124	57,954	63.86	7.48	9.64	...
9	☐	Respiratory	2,932	13,583	54.05	9.08	9.58	...
10	☐	coronavirus	4,187	69,115	64.28	7.25	9.57	...

Figure 2 shows the collocation of the word acute in the COVID-19 corpus via the Sketch Engine software. The number of collocations for the word acute is 111,813 words with a span of -3 + 3, the words adjacent to the word acute are syndrome, respiratory, and severe, these are the top three words in the collocation of the word acute in the COVID-19 corpus and for more co-occurrences more indicated by the word respiratory around 30,752, then the word syndrome with 19,938 co-occurrence and severe words with co-occurrence of 18,089.

In the next discussion, the researcher wants to know the role and use of it in the collocation of the acute word. In this case, the author only takes a sample of 3 words that have the highest frequency in each colocation and focuses on the words that appear after the word acute and those that are side by side with the word acute in the COVID-19 corpus.

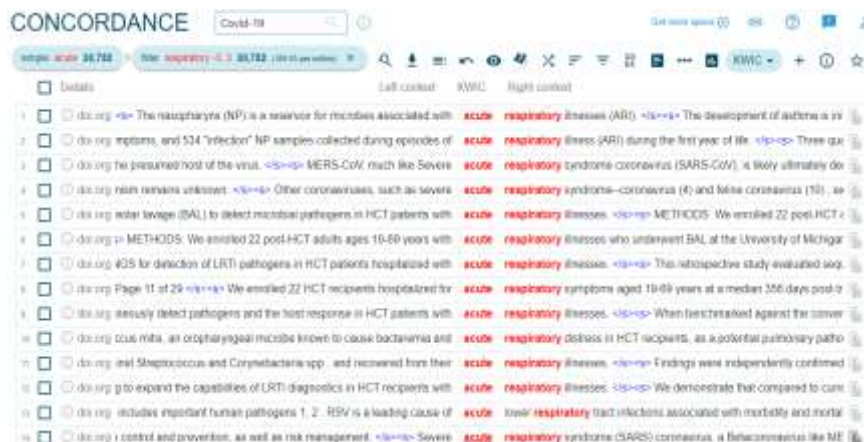**respiratory (a significant word comes after acute)**

Figure 3. Concordance features of the respiratory from acute word

The word collocation adjacent to the word acute and has the highest frequency is the word respiratory. This word tends to present or clarify the name of a virus, namely SARS and the word acute is an abbreviation of SARS, such as in the sentence "During the Severe Acute Respiratory (SARS) epidemic, broad quarantine interventions were applied in Taiwan and subsequently abandoned". In this case the word respiratory as a noun and its use to clarify a meaning of the word SARS so that the word respiratory tends to appear more closely together after the word acute.

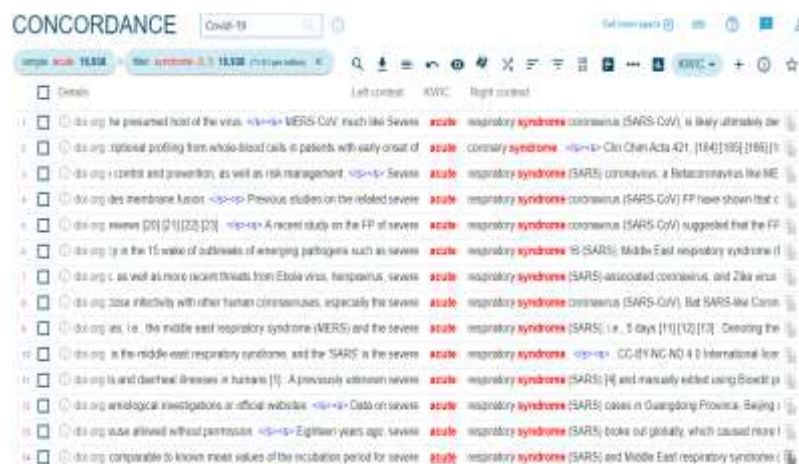**syndrome (the significant second word comes after acute)**

Figure 4. Concordance features of the syndrome from acute word

The word collocation adjacent to the word acute and has the second highest frequency is the word syndrome. This word is more likely to present or clarify the name of a virus, namely SARS and the word acute is an abbreviation of SARS, as in the sentence "MERS-CoV, much like Severe acute respiratory syndrome coronavirus (SARS-CoV), is likely ultimately derived from bats". This is the same as the word respiratory and mutually sustainable because the word syndrome stands for SARS virus. In this case, the word syndrome as a noun and its use to clarify a meaning of the word SARS so that the word syndrome tends to appear more closely after the acute word.

severe (a significant third word comes after acute)

Figure 5. Concordance features of the severe from acute word

In the word collocation adjacent to the word acute and which has the third highest frequency, it is severe. This word is more likely to present or clarify the name of a virus, namely SARS and the word acute is an abbreviation of SARS, as in the sentence "Other Coronaviruses, such as severe acute respiratory syndrome-coronavirus". This is the same as the words respiratory and syndrome and they are continuous because this severe word stands for the SARS virus, but the position for this word is before the word acute. In this case, severe word is a noun and its use is to clarify a meaning of the word SARS so that the word severe also appears alongside the word acute.

Conclusion

Based on the description above, the conclusions that can be drawn from this study are the first that there are three (3) collocations of acute words which tend to be more significant and the top one is side by side with the word acute, namely respiratory with 30,572 and the word syndrome adjacent to acute as many as 30,572. 19,938, and said severe as much as 18,089. Of the three collocations the word is a noun form and its use is to clarify a meaning and an extension of the name of the virus, namely SARS (severe acute respiratory syndrome).

References

Artha, A. F. (2018). REVOLUSI PEMERINTAHAN, SUDAHKAH BEREVOLUSI? KOLOKASI ADJEKTIVA KATA "INDONESIA" DALAM COCA DAN COHA PADA PERIODE PEMERINTAHAN ORDE LAMA, ORDE BARU

DAN ERA REFORMASI. *ETNOLINGUAL*. https://doi.org/10.20473/etno.v2i1.8441

Fan, M. (2009). An exploratory study of collocational use by ESL students - A task based approach. *System*. https://doi.org/10.1016/j.system.2008.06.004

Khoiriyah, A. R. (2018). KOLOKASI BERKONTRUKSI "NOMINA + VERBA" DALAM BAHASA JEPANG PADA MINNA NO NIHONGO SHOKYUU DAN NIHONGO CHUUKYUU. *Paramasastra*. https://doi.org/10.26740/parama.v5i2.3624

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.

McEnery, T. et al. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Routledge.

Saraswati, P. N. (2019). Analisis Berbasis Korpus: Preferensi Semantis Kata "Ayah" dalam Novel "Ayahku (Bukan) Pembohong" Karya Tere Liye. *ISOLEC Proceedings*, 216–219.

Suhardijanto, T., Dinakaramani, A. (2018). Korpus Beranotasi: Ke Arah Pengembangan Korpus Bahasa-Bahasa Di Indonesia. *Kongres Bahasa Indonesia*, 4–18.

Waskita, D. (2018). REPRESENTASI CALON PRESIDEN INDONESIA DALAM SURAT KABAR ELEKTRONIK ASING. *Jurnal Sosioteknologi*, *17*(1), 116–123. https://doi.org/10.5614/sostek.itbj.2018.17.1.11