

THE COLLOCATION MEANING OF “VACCINE” ON SEMANTIC APPROACH: A CORPUS-BASED ANALYSIS**Wulansari**

Universitas Airlangga, Surabaya, Indonesia

Email: wulansari-2019@fib.unair.ac.id

Article history:

Submitted March 23, 2021

Revised April 13, 2021

Accepted May 17, 2021

Published June 29, 2021

ABSTRACT

Corpus semantics examines the words that are used in text and discourse using observations as evidence of meaning. This study aims to investigate the collocations of node ‘*vaccine*’ based on semantic preference and semantic prosody using the corpus-based approach since the source of data is the corpus data. The qualitative research method is used to find the hypotheses from the corpus data taken from sub-corpus in Covid-19 corpus using Sketch Engine software. This study demonstrates that the node word of ‘*vaccine*’ in the semantic preference was shown six words related to the Covid-19 corpus. Meanwhile, the category of semantic prosody was displayed four words negative in meaning, and there was two words displayed positive in meaning, the collocations with negative meaning which occur more frequently than positive meaning. Thus, the writer concluded that the author of a reputable journal explained the word ‘*vaccine*’ with a negative perspective in Covid-19.

Keywords: *collocation; covid-19; semantic preferences; semantic prosody*

ABSTRAK

Semantik Corpus mempelajari kata-kata yang digunakan dalam teks dan wacana menggunakan pengamatan sebagai makna bukti. Penelitian ini bertujuan untuk menyelidiki kolokasi node ‘*vaksin*’ berdasarkan preferensi semantik dan prosodi semantik menggunakan pendekatan berbasis corpus karena sumber data adalah data corpus. Metode penelitian kualitatif digunakan untuk menemukan hipotesis dari data corpus yang diambil dari sub-corpus dalam Covid-19 corpus menggunakan perangkat lunak Sketch Engine. Hasil penelitian ini menunjukkan bahwa simpul kata ‘*vaksin*’ dalam preferensi semantik ditampilkan enam kata yang terkait dengan Covid-19 corpus. Sementara itu, kategori prosodi semantik ditampilkan empat kata negatif dalam arti dan ada dua kata ditampilkan positif dalam arti, collocations dengan makna negatif yang terjadi lebih sering daripada makna positif. Dengan demikian, penulis menyimpulkan bahwa penulis

jurnal yang memiliki reputasi baik menjelaskan kata 'vaksin' dengan perspektif negatif dalam Covid-19.

Kata kunci : *kolokasi; covid-19; preferensi semantic; sematik prosodi*

INTRODUCTION

Recently, corpus linguistics has become a significant concern in last the decades in collecting objective, quantifiable and reproducible data using computers, that can facilitate the use of large corpora (see Barron (2019), Klimova (2014), Botella, et al. 2015), Grisot (2020). Corpus does not have a unique function, apart from being a sample of the language gathered for linguistic analysis (Murphy,2004). A corpus is a collection of texts that have been compiled to represent a particular use of a language, and it is made accessible utilizing corpus linguistic software that allows the user to search for a variety of language features (Cheng, 2011). Besides, the role of corpora means that corpus linguistics is evidence-based and computer-mediated. Corpus linguistics is the study of language stored in corpora, which computerized databases related to linguistic research (Klimova, 2014). Then, corpus linguistic analyses are usually carried out with specialized software programmer on a computer. Language features and testing hypotheses within a linguistic framework can be illustrated by a corpus-linguistic approach. Corpus is divided into corpus-based and corpus-driven. According to Tognini-Bonelli (2001), a corpus-based approach refers to the analysis which is under the certain frameworks which are limited in scope because of the restricted theoretical framework itself. McEnery et.al, (2011) clearly defines a corpus-based studies is a study that employs corpus data to explore theory or hypotheses to create the existing literature or ensure the literature. Meanwhile, corpus-driven approach uses corpus data as the only source of the hypotheses about such study of language. Corpus-driven approach is also closely related to collocation and discourse.

Corpus semantics examines the words that are used in text and discourse using observations as evidence of meaning. The terms 'text' and 'discourse' are conducted in text analysis is seen as the study of how language is organized in the sentence or clause (Stubbs, 2001). Semantics is the level of linguistic analysis, where meaning is analyzed (Bagha, 2011). Three main concepts in corpus linguistics that are often used in the process of analysis are collocation, semantic preferences, and semantic prosody (Yuliawati, 2016).

Collocation is a phenomenon around the fact or simplified that collocation is a word that is, among other words (McEnery et.al, 2011)). In the collocation exist node or lemmas analyzed as the focus of keywords, collocates are the basic words or lemmas that accompany the node in the corpus. Semantic preference is a relationship of words related to other words semantically (Stubbs, 2001). Semantic prosody is a word meaning whose relationship depends on the accompanying words, and some words have a tendency accompanied by words that describe a negative or positive situation (Sinclair, 2003). The frequency with which words appear and word selection patterns are important aspects in formulating the semantic profile of words (Yuliawati (2016).

For the first time, *Coronavirus Disease* (COVID-19) or (SARS-CoV-2) was discovered in Wuhan. This type of single positive RNA strain infects the human respiratory tract and is sensitive to heat and can effectively be activated by chlorine-containing disinfectants. The host sources come from animals, especially bats, and other vectors such as bamboo rats, camels, and ferrets. Common problems that arise are fever, cough, and difficulty breathing (Yuliana, 2020).

In recent years, a number of studies about collocational analysis using the corpus-based approach have been the interest of some researchers, such as Salama (2011) focused on the study of ideological collocation in opposing discourse on Wahhabi-Saudi and Yuliawati (2016) analyzed how women were interpreted in the collocation of words in the Sundanese language magazine,

Manglè, from Democracy to Reformation Era (1958-2013). However, there have been no attempts to examine the collocation node in Covid-19 corpus using Sketch Engine tools. The limited research about coronavirus focused on text and context-based and analyzed using corpus-based. Furthermore, the objective of this research is to investigate *vaccine* node interpreted from the analysis of collocation based on semantic preference and semantic prosody using the corpus-based approach. The result of the analysis would answer how the author of a reputable journal explained *vaccine* node to the readers in Covid-19.

THEORY AND METHOD

The methodology implemented in this research begins with Google databases accessed by Scimagojr.com and corpus compilation design to then applying Corpus Linguistics techniques (frequency lists, keywords, and concordance). The method used in this research was the combined method. Using mix-method, quantitative, and qualitative analysis methods would be combined to identify and interpret noun usage patterns. The aims of these methods are especially applicable to the lexical and constructional semantic relations.

The study is to investigate the collocate *vaccine* as a noun through Covid-19 corpora. Accessed by Scimagojr.com using “Medicine” as a subject and “Critical Care and Intensive Care Medicine” as a subject category, categorized by quartile 1 (Q1), was become the technic of filtering the Journal Rank in which ten reputable journals would be the references of the researcher.

It is worth mentioning again that the writer draws on a corpus-based approach Covid-19 corpora by using Sketch Engine tools in which it can be accessed at <https://www.sketchengine.eu/>. Sketch Engine is a corpus processing tool that can analyze text and collocation to find out the use of keywords in a context. The Sketch Engine is a leading corpus tool used in lexicography (Kilgarriff et al. 2004). The researcher considered the Covid-19 as a reference corpus and demonstrated how useful it is in analyzing the data.

The writer used some procedures to collect the data, first making sub-corpora, adding selected reputable journals, and naming the sub-corpora by *vaccine* COVID-19. Second, investigating the frequently collocates concordance with the query type of word of *vaccine* as a verb in Covid-19 corpora by using Sketch Engine. It was found 133,939 words collocate with *vaccine*. Then, the researcher selected some words which are most frequently collocate and having relevancy with Covid-19 and use -5 and +5 collocations. Finally, the chosen word "*vaccine*" would be analyzed using lexical and constructional semantic relations. The researcher investigated the word *vaccine* because she knew how academic researchers in the popular article viewed word of vaccine in Covid-19 as many people noticed that Covid-19 is one of the upcoming plagues in 2019 up to now 2020 in this world.

Finding and Discussion (Hasil dan Pembahasan)

The data of this study is the node word **vaccine** followed by some collocations using sub-corpora of Covid-19 corpus which were sorted based 10 reputable journals and formulated by the computer-based tool namely Sketch Engine tools. There are six frequently used collocations in the node word *vaccine* shown with a high number of MI in the collocation. As displayed in table 1 below.

Table 1. The collocation of *vaccine* in Covid-19

| Collocate | Freq | Coll. freq. | T-score | MI |
|------------|------|-------------|-----------|-----------|
| attenuated | 33 | 200.000 | 2.062.806 | 1.124.511 |
| induce | 41 | 173.205 | 1.989.987 | 1.063.077 |
| receive | 97 | 173.205 | 1.865.751 | 970.232 |
| influenza | 557 | 264.574 | 1.735.828 | 863.243 |
| early | 347 | 199.999 | 1.723.366 | 847.644 |
| effect | 266 | 173.204 | 1.720.214 | 842.006 |

Table 1 above points out that *vaccine* mainly collocates with quantity words such as *attenuated*, *induce*, *receive*, *influenza*, *early*, and *effect*. All collocates of *vaccine* that founded the word *attenuated* are in the highest value of collocates frequency; the total frequency is 200.000. The second place, the

word of *inducing* with total frequency is 173.205, followed by word *receive* with total frequency is 173.205, then the word *influenza* with total frequency is 264.574, the fifth-place word of *early* is 199.999, and the last word of *effect* with total frequency is 173.204.

| No | Left | Kwic | Right |
|----|--|----------------|--|
| 1 | all HCWs should undergo serologic testing to confirm <u>seropositivity</u> or be immunized with the live, <u>attenuated</u> | vaccine | 559 Susceptible HCWs should not care for patients with suspected or confirmed varicella or zoster infections, |
| 2 | <u>neutralizing</u> antibodies that may be clinically useful for treating other SARS patients. 43 An <u>adeno</u> virus based | vaccine | can <u>induce</u> strong SARS-CoV-specific immune responses in rhesus macaques and holds promise for the development of a |
| 3 | that obesity inhibits both virus-specific CD8+ T cell 137 responses and antibody responses to the seasonal <u>influenza</u> | vaccine | (11). The challenge for 138 future virus pandemics is therefore not only to protect those affected by under nutrition, |
| 4 | <u>influenza</u> , and <u>Staphylococcus aureus</u> . A total of 151 healthy children were randomized 1:1 to <u>receive</u> the | vaccine | starting either at recruitment (n = 74) or 28 days later (n = 77) in a stepped wedge fashion, allowing comparisons |
| 5 | <u>including</u> the same fixed and random effects as we used for the odds of carriage analysis. We modeled the <u>effect</u> of the | vaccine | <u>from</u> one visit after the vaccine was administered. Analyses were done on the modified intention-to-treat population |
| 6 | Definition of abbreviations: CI = confidence interval; OR = odds ratio. *OR of carriage at each visit in <u>early</u> | vaccine | <u>group</u> compared with control group with nursery (random effect, intercept) in the model. Definition of abbreviations |

Figure 1. Concordance of vaccine in Covid-19 corpora

The concordance in figure 1 describes the classification scheme of textual words. The co-textual information combined with the concordance, where the collocation span is *vaccine* + collate, and then these collocates in the terms of common semantic preference. The node word *vaccine* attracts different configurations and collocates that has different negatively shaded preference word: the combination word “*attenuated*” with “*vaccine*”, “*vaccine*” with “*induce*”, “*influenza*” with “*vaccine*”, “*receive*” with “*vaccine*”, “*effect*” with “*vaccine*”, “*early*” with “*vaccine*”.

The overall collocation, most of collocates of *vaccine* are positioned to left. However, only one word is positioned to right. The status of node VACCINE combines as an attributive adjective, noun, and verb. According to *Collins English Online Dictionary, Longman Online Dictionary* (2020) showed that *attenuated* as a verb, *induce* as a verb, *influenza* as a noun, *early* as an adjective, *effect* as a noun, and *receive* as a verb. The modified words of the vaccine in concordance made the writer easier for the writer to identify the purpose of the

journal author. The combination words also would be the identification of whether the author of a reputable journal explained *vaccine* with negative or positive way to the readers.

Another study on the topic by Lindquist (2009) asserts that semantic preference will see the relationship between one word to another that has relation meaning each other. See figure 1 in detail information, by investigating meanings based on semantic preferences, words of collocate on node *vaccine* determined the result of semantic prosody. Based on the significant collocate, it can be seen that the semantic preference for the word *vaccine* in the *Covid-19* corpus. The highest of MI is *attenuated*. It can be concluded that the author of a reputable journal described the node word *vaccine* inherently negatively.

The result from figure 1 that evidence from *Covid-19* corpus described in which the node word *vaccine* may have negative semantics prosody. According to *Longman Online Dictionary* (2020) showed the information of classifying word-class *attenuated, induce, receive, influenza, early, and effect*.

Table 2. The word class of collocate Word

| | |
|---|--|
| 1 | at·ten·u·ate /ə'tenjuet/ verb [transitive] <i>formal</i> to make something weaker or less |
| 2 | in·duce /ɪn'dju:s \$ ɪn'du:s/ verb [transitive] <i>formal</i> to persuade someone to do something. |
| 3 | in·flu·en·za /,ɪnflu'enzə/ noun [uncountable] <i>medical</i> an infectious disease that is like a very bad cold |
| 4 | re·ceive /rɪ'si:v/ verb [transitive] to be given something |
| 5 | ef·fect¹ /ɪ'fekt/ noun [countable, uncountable] a change that is caused by an event, action |
| 6 | ear·ly¹ /'ɜ:li \$ 'ɜ:rlɪ/ adjective (<i>comparative earlier, superlative earliest</i>) in the first part of a period of time, event, or process |

¹ , all HCWs should undergo serologic testing to confirm [seropositivity](#) or be immunized with the live, **attenuated** **vaccine** . 559 Susceptible HCWs should not care for patients with suspected or confirmed varicella or zoster infections.

Data 1 The combination of the word *attenuated* and *vaccine*

According to *Longman Online Dictionary Online*, the definition *attenuated* is to make something weaker. *Annuated* is a verb of past, it is combined with *vaccine* as a noun. The combination meaning could describe making *vaccine* to be weaker. Related with the data 1, these combinations between *annuated* and *vaccine* have negative meaning-the making *vaccine* to be weaker or less for patients with suspected.

| | | |
|---|----------------|---|
| <u>neutralizing</u> antibodies that may be clinically useful for treating other SARS patients. 43 An <u>adeno virus</u> based | vaccine | can induce strong SARS-CoV-specific immune responses in rhesus macaques and holds promise for the development of a |
|---|----------------|---|

Data 2 The combination of the word *induce* and *vaccine*

See table 2 shows that the place of *vaccine* collocation occurs on the right side. *Vaccine* is displayed as a subject (noun) and *induce* as a verb that integrates with *can* as modal auxiliaries; thus, this combination has a great influence on *vaccine* where it is seen as the meaning of *induce*, which is to persuade someone to do something. Therefore, the positive meaning of induce can affect the *vaccine* node in Covid-19. Authorized by looking at data 2, "*the vaccine can induce a strong SARS-CoV-specific immune response*". So, the combination in full sentence between *vaccine* and *induce* have strong meaning of *vaccine* in SARS-CoV.

| | | |
|---|----------------|--|
| that obesity inhibits both virus-specific CD8+ T cell 137 responses and antibody responses to the seasonal influenza | vaccine | The challenge for 138 future virus pandemics is therefore not only to protect those affected by under nutrition. |
|---|----------------|--|

Data 3 The combination of the word *influenza* and *vaccine*

By looking at table 2, the definition *influenza* is a very bad cold. This combination of *influenza* as a noun and *vaccine* as also noun includes negative meanings due to the meaning of *influenza* itself. Showed data 3, the sentence "*The challenge for 138 future virus pandemics is therefore not only to protect those affected by under nutrition*" appeared negative effect issue of Covid-19.

| | | |
|--|----------------|--|
| <u>influenza</u> , and <u>Staphylococcus aureus</u> . </s><s> A total of 151 healthy children were randomized 1:1 to receive the | vaccine | starting either at recruitment (n = 74) or 28 days later (n = 77) in a stepped wedge fashion, allowing comparisons |
|--|----------------|--|

Data 4 The combination of the word *receive* and *vaccine*

While, the meaning of *receive* is to be given something. The combination of *receive* (verb) and *vaccine* (noun) have positive meanings. Data 4 shows the sentence “*influenza, and Staphylococcus aureus. A total of 151 healthy children were randomized 1:1 to receive the vaccine*”, from this data means *vaccine* can reduce the influenza.

| | |
|--|--|
| <u>including</u> the same fixed and random effects as we used for the odds of carriage analysis. We modeled the effect of the | <u>from</u> one visit after the vaccine was administered. Analyses were done on the modified intention-to-treat population |
|--|--|

Data 5 The combination of the word *effect* and *vaccine*

The meaning of *effect* as noun is a change caused by an event, action (table 2). The combination of the word *effect* and *vaccine* has negative words. The sentence shows in data 5, “*from one visit after the vaccine was administered, analyses were done on the modified intention-to-treat population*” explained that there was no confirmation of treated.

| | |
|---|--|
| Definition of abbreviations: CI = confidence interval; OR = odds ratio. *OR of carriage at each visit in early | <u>group</u> compared with control group with nursery (random effect, intercept) in the model. Definition of abbreviations |
|---|--|

Data 6 The combination of the word *early* and *vaccine*

The definitive explanation is the combination of *early* (noun) and *vaccine* (noun), the fact that *noun* is described using *adjective*. So, it has big relation of them. On the other hand, the meaning EARLY is in the first part of a period of time, event, or process (see table 2). The sentence appeared in data 6 above, “*Definition of abbreviations: CI = confidence interval; OR = odds ratio OR of carriage at each visit in early*”. This combination of *early* and *vaccine* has negative words.

Semantic prosody is a word meaning whose relationship depends on the accompanying words, and some words have a tendency accompanied by words that describe a negative or positive situation (see Stubbs, 2001 and Sinclair, 2003. Part of speech of the high number of concordance has an impact combined with *vaccine*. Then, based on Collins English Online Dictionary, Longman Online Dictionary (2020) described the meaning of ‘vaccine’ is a substance that is used to protect people from bacteria or viruses causing

disease. It means that *vaccine* has a positive meaning than other collocations founded. The positive meaning would change automatically if combined with other collocations which have a negative meaning. Consequently, the node *vaccine* would be negative, meaning if it was associated with “*attenuated, influenza, early, and effect*”. Thus, the node *vaccine* would be positive meaning if it was combined with the word “*induce and receive*”.

Finally, the connotation of each collocation can vary because they depend on the context of the sentence. For instance, in data 1, 3, 5, and 6, ‘*be immunized with the live, attenuated vaccine*’, ‘*antibody responses to the seasonal influenza vaccine*’, ‘*the effect of the vaccine was given*’, and ‘*transporting in the early vaccine group compared with control group*’ which has a negative meaning in line with a sentence that weakens the vaccine in the virus, while the context of the positive meaning in data 2 and 3, ‘*vaccine can induce strong sars-cov*’ and ‘*receive the vaccine for healthy children*’ those examples of *vaccine* node can receive and strengthen the virus. The appropriate aspects with actual and potential contexts can be used to reflect the semantic properties of a lexical item. The negative and positive connotations can be found in the concordances of sub-corpus Covid-19. Furthermore, the most striking result is that collocations with negative meanings occur more often than positive meanings.

CONCLUSION

This present study concludes that the word *vaccine* in the semantic preference was shown six words related to Covid-19. The result of semantic prosody indicated that the *vaccine* node having four collocation words were associated with negative meaning and two collocations had a positive meaning. Thus, it was found that the collocation with negative meaning occur more frequently than it is with a positive meaning due to dependent the context in the interpretation of the definition stated in the dictionary. Based on the data, the author of a reputable journal explained the word *vaccine* with a negative perspective in Covid-19. However, this study has a limitation in certain aspect.

The writer suggests that the next researcher could investigate widely, in other words, using a pragmatic approach. It can be focused on how the reputable author guided the readers to the direct or indirect words based on the node vaccine using Covid-19 corpus.

REFERENCES

- Bagha, K. N. (2011). A short introduction to semantics. *Journal of Language Teaching and Research*, 2(6), 1411–1419. <https://doi.org/10.4304/jltr.2.6.1411-1419>
- Barron, A. (2019). Using corpus-linguistic methods to track longitudinal development: Routine apologies in the study abroad context. *Journal of Pragmatics*, 146, 87–105. <https://doi.org/10.1016/j.pragma.2018.08.015>
- Botella, A., Stuart, K., & Gadea, L. (2015). A Journalistic corpus: a methodology for the analysis of the financial crisis in Spain. *Procedia - Social and Behavioral Sciences*, 198(Cilc), 42–51. <https://doi.org/10.1016/j.sbspro.2015.07.417>
- Cheng, W. (2011). *Exploring corpus linguistics: Language in action*. new york, USA: Routledge.
- Dashboard | Sketch Engine. (n.d.). Retrieved July 15, 2020, from <https://app.sketchengine.eu/#dashboard?corpname=preloaded%2F covid19>
- Grisot, C. (2020). Cohesion, coherence and temporal reference from an experimental corpus pragmatics perspective. In *Pragmatics and Society* (Vol. 11). <https://doi.org/10.1075/ps.00034.zha>
- Klimova, B. F. (2014). Using corpus linguistics in the development of writing. *Procedia - Social and Behavioral Sciences*, 141, 124–128. <https://doi.org/10.1016/j.sbspro.2014.05.023>
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.
- McEnergy, T., & Hardie, A. (2011). *Corpus linguistics: method, theory and practice*. United Kingdom: Cambridge University Press.
- Murphy, M. L. (2004). Corpus linguistics at work (review). *Language*, 80(2), 359–359. <https://doi.org/10.1353/lan.2004.0092>
- Salama, A. H. Y. (2011). Synergy of corpus linguistics and critical discourse analysis. *JSTOR*, 22(3), 315–342.
- Sinclair, J. (2003). *Corpus, concordance, collocation*. Oxford: Oxford University Press

- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. United Kingdom: Blackwell publishers Oxford.
- Susi Yuliawati. (2016). Profil semantis nomina perempuan dalam korpus majalah berbahasa Sunda (Manglè, 1958-2013). *Kongres Internasional Masyarakat Linguistik Indonesia 2016*.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.