Analysis Of The Corona Word In German Newspaper In 2020 Based On Corpus Linguistics

Sri Kurniati Almucharomah*  
Pratomo Widodo  
Wening Sahayu

1*Universitas Negeri Yogyakarta, Indonesia, Email: uniealmucharomah@outlook.com  
2Universitas Negeri Yogyakarta, Indonesia, Email: prat_wid@yahoo.com  
3Universitas Negeri Yogyakarta, Indonesia, Email: wenweni88@yahoo.com

*Corresponding author: Sri Kurniati Almucharomah

ABSTRACT
Massive coverage of the Corona virus around the world has caused the word "Corona" to be more familiar to the ear. Likewise in Germany, this virus has caused Germany to implement a lockdown to protect its citizens from the Corona virus attack. This is the background of this research, namely to find out how many times the word "Corona" appears in German newspaper reports. The data in this study were taken from the data corpus of the University of Leipzig, Germany and analyzed using Antconc software. The results showed that the word "Corona" appeared 2,993 times in German newspapers throughout 2020. This study uses quantitative and qualitative methods. The data in this study were taken from German newspapers which took place throughout 2020. The data source was obtained from the website of the University of Leipzig, Germany. There is a lot of corpus data on the page, but this research is limited to 100,000 sentences. Of the many corpus software, in this study, a corpus analysis software called AntConc was used to process corpus data. The steps taken in this study were to download the corpus data on the website of the University of Leipzig. The corpus data in the form of sentences in newspapers in Germany published throughout 2020 amounted to 100,000 sentences. Then the corpus data is processed using AntConc software. After obtaining the data, then the data was analyzed qualitatively to determine the linguistic features. The results of the analysis show that the use of corpus software is very helpful in making it easier to analyze corpus data. Based on the existing corpus data, it can be seen that the word "Corona" throughout 2020 appeared 2,993 times in German newspapers. This shows that the word "Corona" has become the main topic in newspaper coverage in Germany.

Keywords: Corona, Corpus Linguistics, Antconc, German Newspaper.

INTRODUCTION
Since the outbreak of the Coronavirus Disease 2019 (COVID-19) pandemic, countries around the world have adopted various measures to combat the pandemic, among which the sealing of cities from the outside is a common practice, but the corresponding effects vary. Taking Germany and the UK as an example, since the outbreak of the COVID-19 pandemic, Germany has implemented a mandatory lockdown on 8 April 2020 and on 23 May 2020 each country has a significant difference in the effectiveness of epidemic prevention, as evidenced by the fact that Germany quickly contain the spread of the epidemic within months, while the number of confirmed coronavirus cases in the UK continues to rise. Such differences are related to many reasons, such as initial policies and administrative orders, socio-economic disparities and vulnerabilities, mobility and social distancing, availability of health facilities, climate, economic issues, balancing open data and privacy protection policies. (Yu et al., 2021), differences in rates of promulgated social distancing measures, as well as differences at the time of promulgation between countries (Muhyiddin, 2020).
After the corona virus began to attack, the government set rules for working, studying, and worshipping from home (Ihsanuddin, 2020). Since then, the stay-at-home campaign has spread through social media, with the hashtag #stayhome going viral on Instagram and Twitter. The #stayhome campaign was carried out by various groups, including local governments and community organizations who participated in campaigning for the recommendation using the local language.

Because the differences between countries are visible (Susanto, 2020; Yu et al., 2021), which calls for multifaceted analysis. However, there are still insufficient studies in analysing how news is presented as an important way for citizens to obtain information and policies about the pandemic affect public behavior from a linguistic perspective. Therefore, this study tries to explore how the discourse of German-language newspaper media in Germany affects public behavior. This study tries to reveal how often the word Corona appears in German-language newspapers throughout 2020.

Departing from the objectives of the research above, this study selected German-language newspapers circulating in Germany as a source for carrying out related to the word Corona from the report from the beginning of the year to the end of 2020. To analyse this research, Antconc software was used, quantitative and qualitative methods were applied to analyse the number of words used in newspapers in Germany.

The corpus is the most recent development in the field of linguistics. Along with the growth of technology, corpus-based research continues to grow and continues to be a conversation. According to Renouf (1988), the corpus is used to describe a collection of documents, both written and oral, which are stored and processed on a computer for the purpose of linguistic investigation. The corpus does not only talk about a few texts, but texts that number hundreds of thousands to millions.

A corpus is a collection of information, either ordinary information or digital information, in written form that contains various linguistic data, ranging from word level, structure, meaning, and discourse, which can be used for research. Cheng (2012), argues that corpus linguistics works by analysing data through a compilation process based on categories and units of analysis. In the next stage, the most important categories and units of analysis in corpus linguistics are word lists, keywords, and concordance. The word list contains a large number of words contained in the linguistic data and the number of times they occur in the source text. In addition to the word list, there are also keywords which are a collection of words with the highest frequency of occurrence and are assumed to represent the core topic in the text, while concordance is a list of words that collocate with other words and form a new meaning based on the context and context (Adolphs, 2006).

Before computer technology was discovered and developed, all linguistic data in written form was still in the form of manuscripts or handwriting on paper or other objects that could be used as a means of writing. Since the discovery of computer technology and the start of the digital era, now linguistic data in the form of writing or text and spoken language that is converted into digital writing can be stored and processed more easily. This is where corpus linguistics plays an important role by processing digital linguistic data.

There are several corpus processing software, including Antconc, Antgramm, WordCruncher, Nooj, TextStat, MonoconcEsy, Aconcord, and so on. The average corpus software provides free services. However, there are also paid corpus software, such as WordSmith. Through this software can make it easier to analyze corpus data.

In December 2019 there was an outbreak of Corona or Covid-19 which originated from the city of Wuhan, China and spread to various countries in the world including Germany, so WHO declared it a global pandemic. This global pandemic is forcing to change the world order. Starting from the social, economic, and educational aspects. Health protocols are implemented everywhere. The name Corona is taken from the Latin meaning crown, because the shape of the corona virus has spikes that protrude like a crown and the corona of the sun.

The spread of the Corona virus to various parts of the world and causing daily news about the Corona virus is the background of this research to analyze what words follow the word Corona,
how many Corona words appear in German newspapers throughout 2020. In analyzing the research This is using Antconc software.

In addition, the corpus linguistic method is used to analyze linguistic phenomena using the corpus (Ali et al., 2011). This method has been used as a means to analyze actual patterns of language use and also as a tool to develop language teaching materials in the classroom (Prasetya et al., 2020). In relation to the development of linguistic theory, (Astia & Yunianti, 2020) explains that the main contribution of the linguistic corpus is to document the existence of linguistic constructs that are not recognized by current linguistic theory. Even farther (Wang, 2021) stated that the corpus of linguistics contributes to the field of translation and language teaching. In addition, corpus linguistics can also be used in the fields of lexicography and terminology, ideological and cultural studies, registers and genre characterization, as well as in the field of forensic linguistics. There are four main characteristics in this method, namely the empirical approach, focusing on authentic data, the meaning approach, and the use of digital tools to process language data that provides objective results. (Ali et al., 2011).

Based on these problems, this study aims to analyze how many Corona words appeared throughout 2020 in German-language newspapers using a corpus linguistic approach.

RESEARCH METHODS
This research uses mixed methods, namely quantitative and qualitative. Quantitative methods are used to obtain data to be studied and the data obtained are analyzed using qualitative methods. The data in this study were taken from German newspapers which took place throughout 2020. The data source was obtained from the website of the University of Leipzig, Germany. There is a lot of corpus data on the page, but this research is limited to 100,000 sentences.

In this study, a corpus analysis software called AntConc was used to process corpus data. This program is a tool for analysing text to find certain patterns in a language. This software is very easy to download and its use is also not too difficult. Antconc was created by Lawrence Anthony in 2002, a Professor and Chair of the Centre for English Education at Waseda University, Japan. The steps taken in this study were to download the corpus data on the website of the University of Leipzig. The corpus data is already in plain text, so there is no need to convert the data again. The corpus data in the form of sentences in newspapers in Germany published throughout 2020 amounted to 100,000 sentences. Then the corpus data is processed using AntConc software to obtain quantitative data. After obtaining the data, then the data was analysed qualitatively to find out the linguistic features attached to each word class and then conclusions were drawn.

RESULTS AND DISCUSSION
Since the onset of COVID-19, news reports about the epidemic have increased exponentially. Linguists from various countries have collected discourses related to COVID-19 and have built epidemic corpora to interpret the relationship between epidemic progress and related discourses from a linguistic perspective (Debnath & Bardhan, 2020). In general, discourse studies related to COVID-19 can be divided into two strands. One of them is the analysis of the discourse on the COVID-19 topic on various social platforms. Studies have found that there are significant differences in the discourse about epidemics of different groups on social platforms. For example, a study found that among COVID-19-related content posted by Arabic twitter, topics related to religion and health were more dominant and generally passive. (Samsi et al., 2021), which also happened in Malaysia where the majority of online letters associated COVID-19 with negative expressions (Muslimah, 2020). In contrast, another similar study conducted on Germans found that they had a positive attitude in fighting the epidemic (Nabila & Abdulrahman, 2021). Another strand is the analysis of mainstream media news discourse in various countries. Consensus was reached that news discourse about epidemics in different countries could reveal their distinctive ideological and cultural backgrounds. For example, a study of headlines in eight newspapers from four countries found that differences in the naming of epidemics were related to ideological differences (Leo, 2020). It is evident that people's understanding of the concept of “influencer” during an epidemic is strongly influenced by regional and socio-cultural backgrounds (Muslimah, 2020). Meanwhile, analogous conclusions can be drawn by taking the German mainstream media as the object of research. For example, through the discourse-historical approach, it was found that the German media closely linked the arguments of globalism and nationalism (Yu et al.,
In addition, research on the use of fences in news about the epidemic in Germany reveals that the way German media reports on the epidemic reflects the determination and courage of the citizens, as well as the great efforts made by the government for epidemic control. (Muslimah, 2020). From the above review, it can be concluded that there are significant differences in discourse regarding the epidemic in the news media of different countries. As a pioneer country fighting epidemics, Germany has achieved remarkable success in epidemic prevention and control. Today, when the number of infected people is still soaring, taking lessons from the discourse related to the Chinese epidemic and making cross-country comparisons is very important for all countries currently affected by the epidemic. However, through literature review, there is a dearth of studies comparing epidemic discourse between Germany and other countries (Cahyati et al., 2021). Based on this, this study seeks to fill this research gap by comparing the corpora of news discourse from different stages of epidemic development in Germany and the UK. By exploring the characteristics of news discourse related to COVID-19 in the two countries, this study aims to explore the differences in media focus and attitudes in Germany, as well as the ideological differences behind them.

The corpus is defined by its form and purpose (Susanto, 2020). With regard to form, linguists describe a corpus as a collection of naturally occurring examples of language, consisting of everything related to a text, from a few sentences to a set of written texts or tape recordings, collected for linguistic study. Meanwhile, regarding the purpose, the corpus is planned and designed for several linguistic purposes. The specific purpose of the corpus design determines the selection of the text by considering the principles of representation and balance (Uyun, 2016).

In compiling a general corpus that can be used for all kinds of purposes, the linguistic data collected is divided into seven categories, namely (1) imaginative works, (2) journalistic works, (3) books, (4) academic works, (5) ephemeral, (6) documents, and (7) references (Astia & Yunianti, 2020). Meanwhile, in terms of the size of the corpus, according to (Uyun, 2016) there is no minimum size to define a collection of text called a corpus. However, when referring to the initial standard size used in Brown Corpus, the minimum number of words (tokens) that is used as the standard is one million words. This is supported by research conducted by (Prasetya et al., 2020) by using a special corpus taken from the publication of scientific articles in the field of engineering with a token number of less than two million words in the analysis of language phrases for academic purposes.

Antconc software has succeeded in creating a vocabulary list sorted by occurrence. The results showed that this book consisted of 12395 types of words with a word count of 428117 including words that were repeated in the textbook. Of the 3250 words classified by word class, there are 1717 nouns, 619 main verbs, 636 adjectives, 100 adverbs, 29 prepositions, 18 conjunctions and 9 pronouns.

Patient, position, exposure, joint, contrast, projection, imaging, body, bone, inch, anatomy, collimated joint, spine, section, rotation, interest, bone, system are examples of nouns taken from the first 20 nouns that appear based on frequency. The linguistic features contained in the noun list are common nouns and nouns, countable and countable nouns, use of articles. On the other hand, there are also pronouns including subjective, objective, and possessive pronouns which have the lowest frequency.

The keyword in this study is corona which is included in a pandemic disease at this time. To use the Antconc software, first the data to be analyzed is uploaded. Because this program only reads data in *.txt form, the corpus data must be converted first. However, the Leipzig University website already provides data in the form of *.txt, so the data does not need to be converted again. After the corrupted data is in the form of *.txt, then the data is uploaded by clicking on file in the menu bar and then selecting open file. After the data is uploaded, the data processing process begins immediately.
1. Concordance
This menu is used to display certain words that are being searched for. In this study the word searched was the word "Corona". Below is shown the concordance menu.

From the results above, it can be seen that the word "Corona" appears 2993 times followed by other words, such as Corona ab Montag, Corona abgesagt, Corona Abstandregeln, and so on. Then it can be seen that every word that appears following the word Corona is a verb, preposition, conjunction, and noun. Very few Corona words are followed by adjectives. The search results using concordance, it was found that the word Corona appears in two different orthographic forms. Corona was the first to use the symbol for the glottal sound. From the text, apart from the very few typos, it can be seen that Corona is written in Indonesian spelling. For this reason, in this study, the word Corona, based on the frequency of occurrence, appears 2993 times in the corpus concordance data.

2. Concordance Plot
This menu is used to display the word being searched for. Because this research is devoted to searching for the word "Corona", then the word to be searched is the word "Corona". The word "Corona" appears 2,993 times, so the Concordance Plot appears black, because the word Corona is located at the beginning to the end of the sentence. The image display is as follows.
3. File View
File View is used to view all uploaded text and to display the searched word. The word "Corona" is entered in the File View column, then all the words "Corona" will be highlighted to make it easier to see. And for Hit Location, you can enter the number of numbers you want to find. For example, if you want to find the word "Corona" with a total of 10, then in Hit Location, enter the number 10, so that in File View, 10 words "Corona" will be displayed. The following shows an image of the File View.

![Figure 3. File View](image)

4. Cluster/N Grams
This menu displays groups of words based on search. Groups of words are displayed based on the frequency that occurs frequently. N Grams serves to display 2 phrases according to the desired number of phrases. If you want 3 phrases, then 3 phrases will be displayed. In this study there is the word "Corona-Krise" which ranks first with 529 times and the second place is occupied by the word "Corona-Pandemie" as much as 354. In third place is the word "Corona-Fälle". For more details, it is shown below.

![Figure 4. Cluster/N Grams](image)
5. Collocates
This menu serves to show the collocation of the searched word. There are about six main types of collocations: adjective + noun, noun + noun (such as collective noun), verb + noun, adverb + adjective, verb + prepositional phrase (phrasal verb), and verb work + adverb. Below is a view of Collocates.

![Figure 5. Collocates](image)

6. Word List
Word List is used to count all the words in the corpus data sequentially, making it easier to find which words appear most often in a corpus data. Here is how it looks.

![Figure 1. Word List](image)

The data shows that the word "die" ranks first frequently with 53,329 occurrences, followed by the word "der" in second place with 50,271 occurrences. While the word "und" ranks third with 33,215 occurrences.
After seeing the frequency of occurrence of the word, then by using the collocates feature in the AntConc application, the collocation of the word will be analyzed. However, it is important to
note that collocation analysis with tools such as AntConc uses a statistical approach so the range used in the settings will affect the validity of the analysis results. For example, in language collocation analysis (Yu et al., 2021) did not use the 5L and 5R default settings (5 words on the left and 5 words on the right), but used the settings proposed by (Uyun, 2016) so that the setting of the range of each word class is different. However, because there is no corpus research in German that can be used as a reference, this study still uses the initial settings in the AntConc collocation feature to get the results of data analysis first. After that, the data will be filtered by paying attention to psycholinguistic validity. In addition, the collocation with the highest frequency displayed in the collocates feature will also be manually confirmed by looking at the context in which the word is used in concordance.

Corpus software is very helpful in analyzing corpus data. The corpus itself has the advantage of being easily accessible, making it easier to use in the field of research as a means of seeking evidence and confirmation in the field of linguistics. By using the linguistic corpus, it proves that linguistic position is very important for language development.

Corpus linguists are constantly trying to build larger and larger corpora. This trend is most clearly seen in Figure 3, which shows a measure of the word count of the seven most influential corporations of the last 50 years. One of the most important reasons for this trend is that certain linguistic features are rare and will have a near-zero occurrence frequency in small corpora. Obviously, in this case, a large corpora is unavoidable.

However, another reason stems from the view of corpus linguistics as corpus-driven. In this view, the corpus does not serve to test linguistic models but to create linguistic models. Consequently, if that small corpus can provide only a small window on the investigated language phenomenon and hence, the results will only provide a partial picture of its 'true' complexity. On the other hand, the large corpus will provide a more complete view of the phenomenon and thus will always be superior to the smaller corpus. The argument was briefly put forward by (Cahyati et al., 2021), when he wrote:

“There is no virtue in being small. Small is not beautiful; it's just a limitation. If in a small corpus dimension, using the corpus technique, you can get the results you want to get, then your methodology is impeccable - but the results will be very limited...”

However, others argue that small corpora can also be useful. (Arum & Winarti, 2019), for example, providing an interesting example of a corpus study of critical discourse using small corpora, and (Mcenery & Hardie, 2017) analyses the first story in Samuel Beckett's “Text for nothing” series to reveal patterns about the author's writing. Explain the importance of small corporate studies using astronomical analogies. In astronomy, some researchers may be interested in studying galaxies and from this analysis model the universe and how it was born. In contrast, other researchers may be interested in studying single stars, such as our sun, and understanding their life cycles, solar flares, and radiation emission patterns. Few doubt the importance of studying our own sun, and similarly, there is value in studying a small corpus, such as the story of Samuel Beckett, the work of J. K. Rowling, or a number of research articles in biochemistry.

The value of a corpus obviously doesn't depend on its size, but on what kind of information we can extract from it. Therein lies the importance of the corpus tool; we need to have tools that can give us the information we want. For example, if we are interested in observing which characters interact with Harry Potter through the J. K. Rowling Harry Potter series, we need a tool that can record and visualize these interactions in several ways. Unfortunately, the mainstream corpus toolkit does not provide such a feature. It is conceivable that more progress could be made in this regard if researchers focused less on debating the benefits and drawbacks of corporate measures and more on the validity, interest, and value of the analyses they undertake within the scope of the tools available to them.

The results of the quantitative analysis carried out by the AntConc word list tool that generates word lists based on frequency allow researchers to sort and classify words by word class easily. Classification of words based on frequency and then based on word class becomes important
because the results of the classification will help radiography students to determine nouns, verbs, adjectives or adverbs that must be understood and memorized first, especially words that are closely related to radiography process. In addition, it will also inspire researchers to decide how to present these words in a pocket dictionary for compilation. Through AntConc’s concordance, cluster/N-gram and file view tools, researchers were able to qualitatively analyse search terms. By knowing the grammatical rules attached to each verb, besides being able to understand the texts in books, students are expected to be able to understand other English radiology sources such as magazines or journal articles. Furthermore, based on the syntactic information given in the pocket dictionary students can generate their own texts and use them in spoken and written communication.

CONCLUSION
Based on the analysis above, it can be concluded that the use of corpus software is very helpful in facilitating the analysis of corpus data. Of the many corpus software, in this research, AntConc software is used. One of the advantages of AntConc is that it looks user friendly, making it easy to use. This tool also has many menus that help you find the desired word.

Based on the existing corpus data, it can be seen that the word "Corona" throughout 2020 appeared 2,993 times in German newspapers. This shows that the word "Corona" has become the main topic in newspaper coverage in Germany. Considering that this virus first appeared at the end of 2019 and spread massively throughout the world, making the word "Corona" more familiar to the world's citizens.

For further studies in Indonesian, similar research can still be developed by adding Indonesian language corpus data sourced from oral texts with a larger size to enrich studies related to corpus linguistics in German.

REFERENCES

