

CORPUS DRIVEN ANALYSIS OF THE CORONAVIRUS PREVENTION HANDBOOK
Analisis Korpus pada Buku Pedoman Pencegahan Virus Corona

Ayu Humairoh Hakim^a dan Zamzani^b

^{ab}Universitas Negeri Yogyakarta

Jalan Colombo Nomor 1, Karangmalang, Yogyakarta, 55281, Indonesia.

Pos-el: ayu0058pasca.2019@student.uny.ac.id, zamzani@uny.ac.id

Naskah Diterima Tanggal 12 Maret 2021 — Direvisi Akhir Tanggal 13 Mei 2023 — Disetujui Tanggal 15 Juni 2023
doi: <https://doi.org/10.26499/rmh.v12i1.3483>

Abstrak

Informasi tentang pencegahan virus korona penting untuk diketahui, karena virus korona sebagai pandemi global setelah ditetapkan oleh Organisasi Kesehatan Dunia (WHO) setelah berjalan selama satu tahun. Tujuan dari artikel ini adalah untuk mencoba mencari tahu bukti linguistik dari aspek leksikal dan sintaksis. Kemudian dianalisis melalui analisis berbasis korpus menggunakan aplikasi Antconc, perangkat lunak ini digunakan untuk memproses data. Penelitian ini menggunakan metode campuran dan model campuran tidak berimbang, setelah itu mengunduh data, observasi, dan dokumentasi sebagai teknik pengumpulan data, adapun teknik data analisis menggunakan data konversi dan penyusunan data, data klasifikasi data pengurangan dan data tampilan. Hasil analisis kuantitatif dan kualitatif menunjukkan hasil AntConc dan menunjukkan aspek leksikal dan sintaksis yang paling umum seperti kelas kata, dan pola kalimat yang digunakan dalam buku pegangan pencegahan virus korona.

Kata-kata kunci: Analisis korpus, leksikal dan sintaksis, Perangkat Lunak AntConc.

Abstract

Information about coronavirus prevention is important to know because coronavirus as the global pandemic after being designated by the World Health Organization (WHO) for four years. The aim of this article is to find out linguistic evidences from the lexical aspects such as lexical items, lexical categories, lexical cohesion and lexical bundles. Then analyzed it through corpus-driven analysis using AntConc software, this software is employed to process the data. This research uses a mix method and concurrent embedded model, after that downloading the data, observation and documentation as technique of collecting data, while the technique of analysis data use converting and compiling data, reduction data classification data and display data. The results of the quantitative and qualitative analyses show the result of corpus driven analysis through AntConc show the most common of lexical and syntactical aspects such as word classes, and sentences patterns used in the coronavirus prevention handbook.

Keywords: Corpus driven analysis, lexical and syntactical, Antconc software

How to Cite: Ayu Humairoh Hakim dan Zamzani (2023). Corpus Driven Analysis of the Coronavirus Prevention Handbook. *Ranah: Jurnal Kajian Bahasa*. 12(1). 26—38. doi: <https://doi.org/10.26499/rmh.v12i1.3483>

INTRODUCTION

Corpus linguistics is a tool to facilitate language data processing in large quantities, such as counting the number of words, seeing trends in word usage, looking at word clusters, and so

on. In addition to language teaching, corpus features can be utilized in other areas of language, such as sociolinguistics and forensic linguistics, particularly authorship. The statistical significance of applying corpus linguistics can also help locate the degree of generality of research findings and conclusions (Chen, 2018). In addition, corpus linguistics can also increase "trust" empirical" research using language data on a broad scale and providing credibility and validity for analysis (Charles, 2011). In line with Tognini states two approaches to corpus linguistics, namely corpus-based and corpus-driven. Both have differences in viewing the corpus as evidence that can support a theory. The corpus-based approach is deductive, while the corpus-driven approach is inductive, which considers the corpus as evidence that must be a theoretical reference (Tognini, Bonelli, 2001).

Corpus driven analysis is inductive, with no predetermined words or collocations of words as search terms. Rather than driving the research, key words and collocations are derived from the corpus through a variety of quantitative analyses. Theory, then, does not exist before the corpus-driven analysis but is developed from it. In some ways, corpus-driven analysis performs the typically qualitative function of searching out what appears important in a collection of texts. In our corpus-driven analysis, the quantitative analysis identifies the salient features to be examined by a subsequent qualitative discoursee (Thompson, 2016).

Based on the explanation of the theory above, corpus driven analysis is believed to be able to analyze every word in information about coronavirus. Coronavirus pneumonia (Covid-19) was first reported from Wuhan China, then spread throughout China and even to other countries of the world. Confirmed Covid-19 cases have increased far exceeding SARS in 2003, and mortality cannot be ignored. Recognizing its human-to-human transmission capabilities, the world health organization identified it as a public health emergency for international concerns on January 31, 2020. These facts are sufficient to describe the severity and complexity of the outbreak. Then Indonesia faced a covid pandemic in early March 2020. Almost all sectors of life are paralyzed by the effects of the Covid-19 virus, both education, transportation, and the economy. After being designated as a global pandemic by the World Health Organization (WHO) for one year, the first detected coronavirus outbreak in the Chinese city of Wuhan at the end of 2019 is still spreading among the public (Dhama K, 2020). In between are tucked word and phrases that become a trend among language users. Nowadays many emerging and order the latest words about the coronavirus pandemic (Covid-19) that has spread throughout the world and is no exception in Indonesia. With this event, several studies examine the words or terms that occurred during the covid 19 pandemics.

Several studies examine the phenomenon of Covid-19 both in terms of health and language. The first study was conducted by Konul Hajiyeva which focuses on analyzing corpus based lexical analysis of subject-specific university textbooks which the purpose to explore lexical text coverage and frequency distribution of words from the Academic Word List and the British National Corpus. In this research using computational corpus analysis because this instrument used range softwere as a computer program that performs lexical text analysis. Then, the result is the Academic Word List is not as general an academic vocabularyas it was initially intended to be and, more important invalid source specified (Hajiyeva, 2015). The next research in line with this study was conducted by Euis Reliyanti Arum and Wiwin Winarti they analyzed that use of antconc in providing lexical and syntactic information of the textbook of radiographic positioning and related anatomy: a corpus linguistic study. This study aimed to acknowledge lexically and syntactically the linguistics features used in the textbook of Radiographic Positioning and Related Anatomy, then antconc show that has benefited the researchers since it has been able to assist the researchers in providing needed lexical and syntactical information of the textbook (Arum, 2020).

The other research from the editor of Oxford English Dictionary (2020) This study indicates that the frequency of use of vocabulary and linguistic studies in Coronavirus pneumonia has always increased significantly because many users use new vocabulary and terms in everyday life during this pandemic. The objective of this study was to examine of vocabulary used in each section of the coronavirus article. In line with, this research is relevant that conducted by NorFariza in his research using corpus driven analysis to analysed collocate on Covid-19, this research showed that the collocates reflect fear, anxiety from several newspaper reports from English newspapers in Malaysia. It also shows how the virus is taking a toll on Malaysia (NorFariza, 2020).

These studies are different from those conducted by researchers in this paper. All explanations about the previous research discussed above were focused on analyzing lexical and syntactic based on corpus study, but each research is not analyzed deeply. This research will complete the previous research missing points that need to be filled as an empirical gap in this study, by analyzing linguistic aspects that appear in Coronavirus Prevention Handbook. This study also is a varied use of the corpus where researchers use the English corpus on each handbook and there is also Chinese in the Coronavirus prevention handbook.

To support those cases, this research using Coronavirus Prevention handbook by Professor Wang Zhou from Wuhan Center for Disease Control and Prevention organized experts to compile a guidebook on an overview of the coronavirus and its transmission, detection, and treatment of its disease, precautions for individuals and public places, and the basics about infectious diseases. With graphic illustrations and plain language, this book is intended as a systematic introduction to scientific knowledge about Covid-19 (Zhou, 2020). From this book also, the researcher will examine linguistics aspects such as lexical and syntactical.

In line with, this research using corpus driven analysis, because this research combined with computational tools for analysis and the computational tools is antconc tools. This research to find out How corpus-driven analysis analyzes lexical and syntactical aspects in the coronavirus prevention handbook. Then analyzed it using antconc, this software is employed to process the data from the coronavirus prevention handbook.

THEORETICAL BASIS

Over the past ten years, corpora of language data have started to play an increasingly important role in determining how languages are taught. Corpora nowadays is also built in different language uses with different sizes based on different purposes of its studies. McCarthy argued that corpora now appear for many languages and compiled for different purposes (McCarthy, 2010). As Hunston writes, corpora have been applied in a wide range of areas, including translation studies, stylistics, and grammar and dictionary development (Hunston S. , 2012) It can be seen that the study of the corpus has given many contributions to the language fields, even in language teaching.

The ideas of a corpus study have been developed in many language research even in various aspects of practical language. According to Anna O'keff state that corpus is a collection of texts, written or spoken, usually stored in a computer database (Anne O'keff M. M., 2011). From the definitions it can be concluded that corpus is stored in a computer-based system, contains entire forms of text, even if they are written or spoken form. It can be said that computers have a significant role in the corpus. Computer-based programs can be used in collecting and storing corpus data. The use of computer-based programs creates the possibility of compiling a huge size of data quickly with a low budget. Moreover, computers can be used to manage the data of the corpus to conduct an analysis.

Talk about corpus linguistics has two primary research approaches are present under corpus linguistics, scholars presented similar definitions of corpus-based researches and corpus driven analysis. McEnery corpus-based researches often “use corpus data in order to explore a theory or hypothesis, typically one established in the current literature, in order to validate it, refute it or refine it” On the other hand, corpus-driven linguistics “rejects the characterisation of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language” (Ghani Lu, 2021).

On the other hand, corpus-driven analysis is virtually useless without a computer software tool to process it and display results in an easy to understand way. To study a corpus, there is a special program analytical tool is needed. To support the analysis using Antconc tools, AntConc contains seven tools that can be accessed either by clicking on their 'tabs' in the tool window or using the function keys F1 to F7. The seven tools are *Concordance Tool*, *Concordance Plot Tool*, *File View Tool*, *Clusters (N-Grams)*, *Collocates*, *Word List*, *Keyword List* (Anthony, 2011). The corpus-driven also as a methodology whereby the corpus serves as an empirical basis from which lexicographers extract their data and detect linguistic phenomena without prior assumptions and expectations (Tognini, Bonelli, 2001).

Furthermore, McEnery argue that corpus linguistics is indeed a methodology rather than an independent branch of linguistics in the same sense as phonetics, syntax, semantics or pragmatics and the other branch of linguistics (McEnery T. &, 2007). By using corpus-driven analysis this research will interest to analyze lexical analysis and syntactical analysis as one aspect of linguistic corpus analysis. Because when we are used corpora for lexical analysis, we want to find the lexical text coverage and frequency distribution of words in this pandemic from the coronavirus prevention handbook, such as lexical categories and lexical bundles.

However, lexicology studies various lexical units which is concerned with words, variable word-groups, phraseological units, and with morphemes which make up words. In line with Svensen in Borris Pritchard states that lexicology is a linguistic branch that examines the vocabulary of language, structure, and characteristics of the word, as well as the meaning of the word (Pritchard, 2015). Thus, Lexical forms include words or lexicon that have not undergone the process of deformity, concrete, and denotes. Simply lexicographically conceptualized over linguistic branches that can analyze the linguistic corpus around data collection, selection data, and word description units or word combinations in one or more languages.

The lexical categories are a topic that spans many of the traditional divisions of linguistics, including inflectional morphology, derivational morphology, syntax, and semantics. While this research just focus on lexical categories, a lexical category is a syntactic category for elements that are part of the lexicon of a language, according to (Hornby, 1972) also talk about lexical categories that lexical can be defined as the words of a language. Then the classification of lexical items into categories (also called ‘parts of speech’ or ‘word classes’ (Goethem, 2017). While The term ‘lexical category’ is generally used to describe the categories of noun, verb, adjective, and possibly certain others (Haspelmath, 2012). By using of lexical categories is to find out how choices of words involving various type of meaning. Meanwhile, Luwis also suggests the following taxonomy of lexical categories include base word, prefix, and suffix, prepositions, compound words, particles, numeral, abbreviation and acronyms, conjunction, word classes (Luwis, 1977).

In lexical bundles that are interrelated with lexical and syntactical aspects that the author analysed. Lexical bundles come from formulaic language, formulaic is one of the characteristic features of legal discourse, which manifests itself not only at the level of wording, “but also in the content, structure and layout” of legal texts (Ruusila, 2016). The study shows that lexical bundles are not an accidental by-product of corpus frequency

analysis. Rather, these word sequences turn out to be consistently functional, indicating that high frequency is a reflection of pre-fabricated or formulaic status. Although they go largely unnoticed by speakers, hearers, and analysts, lexical bundles are pervasive in spoken and written texts, where they serve basic discourse functions related to the expression of stance, discourse organization, and referential framing (Douglas Biber, 2006). Lexical bundles are simply a given length of repetitive strings of words, they are also known as N-Grams, in which “N” stands for sequences of contiguous number (“N”) of words. Below is the example of N-Grams (Granger, 2014).

The use of lexical in corpus analysis because can be simplified such as explanation above, its techniques are less complex than syntax analysis, secondly should be optimized or efficiency because is requires a significant portion of total compile time, and the last is Portability, it means the lexical analyzer is somewhat system dependent input processing. This study will examine the lexical aspect contained in the Coronavirus prevention handbook using a corpus driven analysis so that several lexical aspects can be found such as lexical items, lexical categories, lexical cohesion, and lexical bundles. Then, based on the book that will be examined there are several combinations of words during the Covid-19 pandemic. To complete the study, researchers also examined syntactical aspects contained in the coronavirus prevention handbook.

Meanwhile, Muliono in Amin said that syntax is a study of the rule of word combinations into larger units, namely phrases, clauses, and sentences. Thus, to understand syntax, it is necessary to understand these units (Mujid F Amin 1, 2021) In using corpora for syntactic research, we want to find instances of some pattern of linguistic relevance to explore, support, or refute a linguistic claim involving that pattern. In line with Syntax is the branch of linguistics that investigates word units and other units above words, their relationships to each other, and their arrangement became a speech unit (Chaer, 2015).

Talk about syntactic analysis, Ramlan state that syntax is a part or branch of linguistics that discusses the intricacies of discourse, sentence, clause and phrases (Ramlan, 2005). In line with, Tarigan state that *syntax in the studi of the patterns by which words are combined to make sentences* (Tarigan, 2008). This definition describes that discourse, sentences, clauses, and phrases are forms or units of language in which some intricacies need to be discussed or reviewed.

In syntactic research, at the fundamental empirical level, observes words, their form, order, and coocuranc in a sentence. The patterns of interest in syntactic research are, however, typically described in terms of generalizations and abstractions over the form and order of words (or groups thereof, for those syntactic paradigms that assume a notion of the constituency). This raises the question of how a syntactic pattern of interest can be characterized in terms of the properties of a particular corpus and its annotation. After that in the analysis through the concordance technique using the antconc application, the technique has several stages so that it can be known the linguistic features attached to a word will be known such as the word class, the type of phrase, clauses, and type of sentences.

RESEARCH METHOD

This research uses a mixed method and concurrent embedded model because this research methods combines qualitative and quantitative research methods by mixing the two methods unequally. According to Sugiyono Mixed methods are divided into two main models, namely the sequential model and the concurrent model. The sequential model is divided into two, namely *sequential explanatory* and *sequential exploratory*. Then the concurrent model also divided into two there are *the concurrent triangulation model* (a

mixture of quantitative and qualitative) and *the concurrent embedded model* (an unequal mix of quantitative and qualitative) (Sugiyono, 2011).

Meanwhile the data collection techniques in this study started from quantitative then qualitative. At the first stage, the quantitative method is used to obtain quantitative data which is carried out with the help of the AntConc concordance software. The Coronavirus prevention handbook in pdf format is converted into plain text, so that it can be processed by AntConc. Through the frequency feature on this device, a list of all words used in the book is generated based on their frequency of occurrence, from the most frequent to the one that only appears once. The list produced by the software is, then sorted during the reduction process to separate words that have full meaning and other words such as base word, acronym, and abbreviation, prefix, and suffix, are classified into word classes according to their function in the sentence. After obtaining the data that has been selected, the words are then analyzed qualitatively to determine the linguistics features attached to each word class to draw conclusions.

DISCUSSION

In this study, the researcher analyzed the Coronavirus prevention handbook which had 143 pages and was written by Professor Wang Zhou, from the Chinese Center for Disease Control and Prevention. As explained in the research question above that how corpus driven analyzed linguistic aspects such as lexical and syntactical. in the Coronavirus prevention handbook. Below are the result of word types and word tokens that found in the AntConc.

Rank	Freq	Word
1	1148	the
2	1028	and
3	761	of
4	648	to
5	459	in
6	458	a
7	371	or
8	353	with
9	352	x
10	336	is
11	329	for
12	320	covid

Picture 1. The AntConc Result

Based on the AntConc result, the features of *Word types* show the individual types of words used in the book. The number of words used is 3607 words which will then be reduced, classified and analyzed. Meanwhile the *words token* is the total number of words in the book including the words that are repeated or in other words. *Word tokens* are the total number of words resulting from the frequency of each word. Thus, from Coronavirus prevention handbook who has 143 pages contain a total of 28.590 words.

This study will examine the lexical aspect contained in the Coronavirus prevention handbook using a corpus driven analysis so that several lexical aspects can be found lexical categories and lexical bundles. Below are the classification of lexical categories.

a. Lexical categories

By using of lexical categories to find out how choices of words involving various type of meaning. Meanwhile, (Luwis, 1977) also suggests the following taxonomy of lexical categories include, *base word, prefix and suffix, abbreviation and acronyms, latin word and word classes*. The next stage the researcher focuses on the field data that has been collected.

Field data in the form of a list of base word on the frequency of their appearance is further reduced by separating meaningful words of the other words. At this stage of reduction than a meaningful word, it was also managed to collect a list of *base word, prefix and suffix, abbreviation and acronyms, latin word and word classes*. The reduction stage determines the degree of relevance to the purpose of research contained in the coronavirus prevention handbook.

Table 1.
The Result of Base Word

Total of Base Word 2582					
No	Freq	Base Word	No	Freq	Base Word
1.	132	Respiratory	29.	50	Healthcare
2.	125	Health	30.	49	Viruses
3.	118	People	31.	47	Care
4.	115	Disease	32.	46	Cases
5.	91	Hands	33.	46	Employees
6.	87	Contact	34.	45	Community
7.	80	Symptoms	35.	45	Fever
8.	79	Infection	36.	45	Guidance
9.	78	Medical	37.	45	Novel
10.	73	Public	38.	45	Prevent
11.	70	Spread	39.	45	Use
12.	68	China	40.	45	Water
13.	67	Risk	41.	44	Time
14.	65	Including	42.	43	Prevention
15.	65	Patient	43.	42	Illness
16.	65	Transmission	44.	42	Pneumonia
17.	64	Person	45.	41	Days
18.	63	Ion	46.	41	Infected
19.	62	Home	47.	41	Outbreak
20.	62	Patients	48.	40	Masks
21.	62	Sick	49.	40	Soap
22.	60	Animals	50.	39	Help
23.	58	Control	51.	38	Measures
24.	58	Diseases	52.	37	Confirmed
25.	57	Avoid	53.	36	Clean
26.	57	Do	54.	34	All
27.	57	Mask	55.	34	Travel
28.	52	Close	56.	33	Alcohol

The data above is the result of reducing the lexical categories of base word there is 2582 base words contained in the Coronavirus prevention handbook, however not all the data is embedded in the table above. Result of the data using antconc software and also as a result of quantitative research. Meanwhile below is a table of prefix and suffix contained in Coronavirus prevention handbook.

Table 2.
The Result of Prefix and Suffix

No	Prefix	Suffix
1.	Un-	-ed
2.	Work-	-able
3.	Diss-	-ing
4.	Pre-	-cation
5.	Post-	-tions
6.	Pro-	-ly
7.	Re-	-ness
8.	In-	-ers
9.	Dis-	-y
10.	Anti-	-er
11.	Healthy-	-nal
12.	Micro-	-ent
13.	Over-	
14.	Sub-	
15.	Miss-	

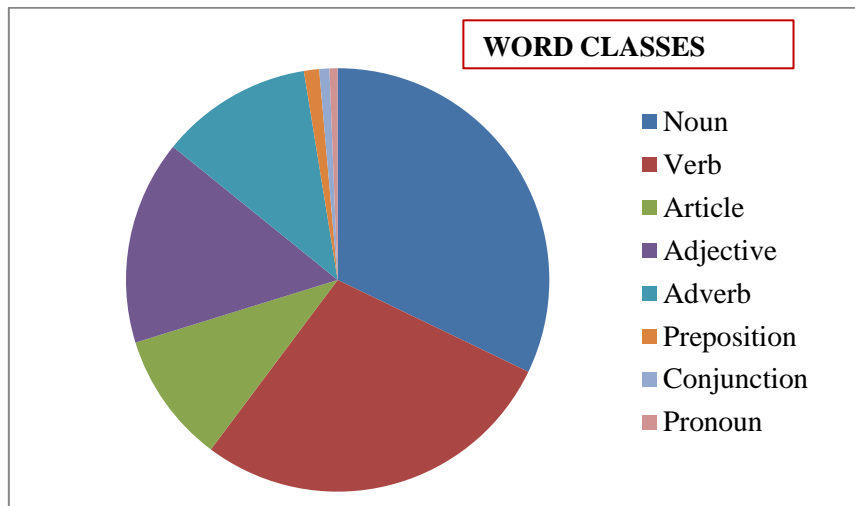
Table 3.
The Result of Abbreviation and Acronym

No	Abbreviation and Acronym	Description
1.	COVID-19	Coronavirus Disease 2019
2.	SARS	Severe Acute Respiratory Syndrome
3.	EVD	Ebola Virus Disease
4.	NPIS	Nonpharmaceutical Interventions
5.	CDC	Centers For Disease Control And Prevention
6.	WHO	World Health Organization
7.	SARS-COV	Severe Acute Respiratory Syndrome-Coronavirus
8.	COV	Coronavirus
9.	N-COV	Novel- Coronavirus
10.	MERS	Middle East Respiratory Syndrome
11.	MERS-COV	Middle East Respiratory Syndrome-Coronavirus
12.	ICTV	International Committee On Taxonomy Of Viruses
13.	ARDS	Acute Respiratory Distress Syndrome
14.	PHEIC	Public Health Emergency Of International Concern
15.	PHE	Public Health Emergency
16.	CAP	Community-Acquired Pneumonia
17.	VRE	Vancomycin-Resistant Enterococcus
18.	RRT	Real-Time Reverse Transcription
19.	PCR	Polymerase Chain Reaction
20.	PPE	Personal Protective Equipment
21.	ADA	The Americans With Disabilities Act
22.	TCM	The Most Commonly
23.	PUI	Person Under Investigation
24.	EOC	Emergency Operations Center
25.	USDA	The U.S.Department Of Agriculture
26.	FWS	Fish And Wildlife Service
27.	HCP	Healthcare Personnel
28.	HKU	Covid-19 Diagnosis

After the researcher reduced the collected data such as base words which amounted to 2582 words, after that the prefix was 15 words and suffix 12 words and the last one was abbreviation and acronyms amounted to 28 words contained in the Coronavirus prevention handbook, the researcher continued to the stage of technique data analysis namely classifying data.

Thirdly, the researcher classifies the selected data. Classification is done by selecting and separating the wordlist based on their respective word classes. These words are classified based on the theory from Gelderen which explain that the lexical categories (word classes) divided into 5 categories there are : Noun (N), Verba (V), Adjective (Adj), Adverb (Adv), and

Preposition (P) (Gelderen, 2010). The results show that the book is composed of 3607 words of words with the total words of 28.590 including the repeated words in the textbook. Of the 2582 words being classified by word classes, there are 830 nouns, 725 main verbs, 258 articles, 30 prepositions, 20 conjunctions, 403 adjectives, 300 adverbs, 16 pronouns. The percentage of each word class is presented in the following pie chart below.



Picture 2. Classification Result By Word Classes

Lexically, the use of noun has the biggest part of the pie showing the highest total of words occurring in the text. Then also pronoun including subjective, objective, and possessive pronouns. The linguistics features found in the noun list are common and proper noun, countable and uncountable nouns. In the opposite, there is also adjective and adverb having the lowest frequency. Articles has 258 words from word classes because almost used such as a, and the. In the Coronavirus prevention handbook, the articles used to describe noun which is countable or uncountable noun.

The most common verbs occur in the text are is, be, are, may, should, and can, which belong to the auxiliary verbs. The verbs to be, as the auxiliary verbs, are mostly used to form passive sentences. Samples of the lexical verbs are visit, living, required, balanced, identify, tested, taken, perform, plans, smoking, traveling, published, preventing. There are four form of verbs; base form, verb -ing, past form, and past participle. Syntactically, each verb form shows their grammatical features some of which related to the tenses and clause. Many verbs occur in base form display simple present tenses. However, not all -ing verbs form is used in present continuous tense. Some of -ing verbs are used to write reduced clauses and others are used as adjectives modifying nouns and forming noun phrases. Most of verbs occur in the past participle forming passive sentences, present perfect tense, and as adjectives modifying nouns. Only some of past verbs are used.

Based on word classes also adjective and adverb that smallest frequency the Coronavirus prevention handbook. In this book the function of the adjectives in addition to modifying nouns that form noun phrase, some of them also become the complement of verbs to be. This study also found some adjectives in the form of -ing verb and past participle. There is also the use of comparative degree with suffix and prefix, suffix -er and -est then prefix-un. After that, adverb

Samples of prepositions are of, to, in, for, with, on, by, from, at, and into. Meanwhile, kinds of conjunction found in the word list are coordinate conjunctions such as and, or, but, subordinate conjunctions such as which, when, where, whenever, whose, that, and correlative conjunctions such as also, only, either, and unless. In the stage of classifying data also obtain

of latin word as the explanation the kind of virus and as some ingredients of covid 19 medicine, below are the result of latin word.

Table 4.
The Result of Latin Word

No	Latin word
1.	Orthomyxoviridae
2.	Paramyxoviridae
3.	Paramyxoviruses
4.	Syncytial Virus
5.	Measles Virus
6.	Mumps Virus
7.	Hendra Virus
8.	Nipah Virus
9.	Togaviridae (Rubella Virus)
10.	Picornaviridae (Rhinovirus)
11.	Coronaviridae(SARS Coronavirus)
12.	Adenovirus
13.	Reovirus
14.	Coxsackie Virus
15.	ECHO Virus
16.	Herpes Virus
17.	Nidovirales
18.	Orthocoronavirinae
19.	Serotypic
20.	Genomic
21.	Nucleocapsid
22.	Polymorphic
23.	Alphacoronaviruses
24.	Betacoronaviruses
25.	Pulmonary Parenchymal
26.	Pathogens Include Mycoplasma
27.	Chlamydia
28.	Klebsiella Pneumoniae
29.	Escherichia Coli
30.	Staphylococcus Aureus
31.	Pseudomonas Aeruginosa
32.	Acinetobacter Baumannii
33.	Parainfluenza Virus
34.	Rhinovirus
36.	Adenovirus
37.	Human Metapneumovirus
38.	Haemophilus Influenza
39.	Mycoplasma
40.	Streptococcus
41.	Meningococcus (Neisseria)
42.	Benzoxazole/Methicillin-Resistant
43.	Vancomycin-Resistant Enterococcus
44.	Clostridium Difficile
45.	Sodium Hypochlorite
46.	Rhizoma Phragmitis (Lu Gen)
47.	Rhizoma Imperatae (Bai Mao Gen)
48.	Radix Angelicae Dahuricae (BaiZhi)
49.	Rhizoma Atractylodis Macrocephalae (Bai Zhu)
50.	Rhizoma Atractylodis (Cang Zhu)
51.	Honeysuckle (Jin Yin Hua)
52.	Herba Pogostemonis (Huo Xiang)
53.	Radix Et Rhizoma Rhodiolae Crenulatae (Hong Jing Tian)
54.	Rhizoma Dryopteridis Crassirhizomatis (Guan Zhong)
55.	Rhizoma Polygoni Cuspidati
56.	Fructus Tsaoko (Cao Guo)
57.	Pericarpium Citri Reticulatae
58.	Folium Mori (Sang Ye)
59.	Radix Astragali Praeparata
60.	Radix Ligustici Brachylobi (Fang Feng)
61.	Herba Eupatorii (Pei Lan).
62.	Chlorine-Containing Disinfectants
63.	Peracetic Acid
64.	chloroform
65.	lipid solvents
66.	Chlorhexidine
67.	chlorhexidine gluconate

b. Lexical bundles

Below are the result of lexical bundles In the Coronavirus prevention handbook examined are three-, four-, five, and six-word bundles. Lexical bundles are basically extended collocations based on the frequency of occurrence and the spread (or the range) of usage in the text.

Table 5.
The Result of Lexical Bundles

Structure Of Lexical Bundles	Sample Of Lexical Bundles	N-Grams
Type 1 : Verb/ adjective + to-clause fragment	Are likely to be	4-Grams
	Can be used to	4-Grams
	Can be related to	4-Grams
	To be able to	4-Grams
	has been shown to	4-Grams
Type 2 : Prepositional phrase with embedded <i>of</i> -phrase fragment	refer to the	3-Grams
	in the form of	4-Grams
	In the case of	4-Grams
Type 3 : Anticipatory <i>it</i> + verb phrase/adjective phrase	In this context of	4-Grams
	It is named	3-Grams
	It was found that	4-Grams
	It was related to	4-Grams
	It is found that	4-Grams
	It refers to the	4-Grams
	It showed from the	4-Grams
Type 4: Passive verb + prepositional phrase fragment	It is likely to	4-Grams
	is divided into	3-Grams
	Can be used in	4-Grams
	Can be seen in	4-Grams
	Can be shown in	4-Grams
Type 5 : Prepositional phrase with embedded <i>of</i> -phrase fragment	could be found in	4-Grams
	in terms of	3-Grams
	In the form of	4-Grams
	In the case of	4-Grams
Type 6 : Other expressions	In this context of	4-Grams
	as well as	3-Grams
	such as	2-Grams

As can be seen in Table 6 above, the most dominant structural form of lexical bundles in findings and discussion sections of the Coronavirus prevention handbook is structure type 3 Anticipatory *it* + verb phrase/adjective phrase. It is found that there are 7 lexical bundles appeared for type 3, such as *It is named. It was found that it was related to "It is found that it refers to the, It showed from the, it is likely too.* These bundles are used to report the writer's kind of stance related to studies, It is in line with the finding of the previous studies. They carried out a study on lexical bundles in research articles which found that the largest structural category of lexical bundles in research articles was other prepositional phrase.

The result of syntactic analysis in the Coronavirus prevention handbook is how words are set up and relation in shaping phrases, clauses, and sentences. In general, the syntactic structure consists of a composition of subject(S), predicate(P), object(O) and information relating to syntactic function. Syntactic function also called grammatical functions. In this handbook also found some of phrase and clause like in the result of lexical bundles. After seeing the words surrounding the verb above, it is concluded that there are types of active and passive sentences such as Simple present, Simple past, Present continuous, Simple present perfect, and the use of gerund also reducing relative clause.

CLOSING

Thus, it can conclude from the Coronavirus prevention handbook, this research found the linguistic aspects such as lexical and syntactical which is analyzed it through corpus-driven analysis using Antconc application, this software is employed to process the data from the coronavirus prevention handbook.

Using AntConc, has successfully conducted the quantitative analysis that generated the descriptive data of all words used in the textbook that were sorted by frequency. The data show that the book is composed of 3607 words with the total words of 28.590 including the repeated words in the textbook. 2582 words being classified by word classes, there are 830 nouns, 725 main verbs, 258 articles, 30 prepositions, 20 conjunctions, 403 adjectives, 300 adverbs, 16 pronouns. Besides, there were 15 prefixes and 12 suffixes, and the last one was an abbreviation and acronyms amounted to 28 words, not only that there was a Latin word of 67 words. For the lexical result, the researcher also found 27 lexical bundles that explain above.

Meanwhile, in the syntactic analysis, the AntConc tools to analyze the descriptive data qualitatively. This tools display the search term in a concordance line to see the keywords in context, group the words by their similarities, and show each search term in an individual text. All of them enable the researchers to observe each word and conclude the grammatical features attached to each word in each word classes, than also find out phrase, clause and sentence categories.

REFERENCES

- Anne O'keff, M. M. (2011). *From Classroom to Corpus*. New York: The United States of America by Cambridge University Press, New York.
- Anthony, L. (2011). AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom. *IEEE International Professional Communication Conference Proceedings* (pp. 729-737). Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Arum, E. R. (2020). Penggunaan Linguistik Korpus Dalam Mempersiapkan Bahan Ajar English For Specific Purpose Di Bidang Radiologi. *Jurnal Teras kesehatan*, 2, 58-69. <https://doi.org/10.38215/jutek.v2i2.39>
- Chaer, A. (2015). *Sintaksis Bahasa Indonesia: Pendekatan Proses*. Jakarta: Rineka Cipta.
- Chen, M. &. (2018). introducing data driven learning to PhD students for research writing purposes: A territory wide project in Hong Kong. *English for Specific purposes*. <https://doi.org/10.1016/j.esp.2017.11.004>
- Dhama K, K. S.-A.-M. (2020). Coronavirus Disease 2019-COVID-19. *Clin Microbiol Rev*, <https://doi.org/10.1128/CMR.00028-20>
- Douglas Biber, F. B. (2006). Lexical bundles in university spoken and written Register. *English for Specific Purposes Elsevier*, 263-286. <https://doi.org/10.1016/j.esp.2006.08.003>
- Gelderen, E. V. (2010). *An introduction to the grammar of English (Revised edition)*. John Benjamins Publishing Company.
- Ghani Lu, R. A. (2021). A comparative study of corpus-based and corpus-driven approaches. *Linguistics International Journal*, 15 (2), 119-132. <https://connect.academics.education/index.php/lij/article/view/157>
- Goethem, K. V. (2017). Lexical categories and processes of category change. Perspectives for a constructionist approach. *Zeitschrift für Wortbildung / Journal of Word Formation*, 2 (2), 31-61. <https://doi.org/10.3726/zwjw.2017.02.02>
- Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14, 58–72. <https://doi.org/10.1075/lic.14.1.04gra>

- Hajiyeva, K. (2015). A corpus-based lexical analysis of subject-specific university textbooks for English majors. *Ampersand*, 136-144. <https://doi.org/10.1016/j.amper.2015.10.001>
- Haspelmath, M. (2012). *How to compare major word-classes across the world's languages*. Los Angeles: Department of Linguistics, UCLA.
- Hornby. (1972). *Oxford Advanced Learner's Dictionary*. Oxford University Press.
- Hunston, S. (2012). *Corpora in applied linguistics*. Cambridge University Press.
- Luwis, M. (1977). *Implementing the Lexical Approach*. Hove. England: Language Teaching Publication.
- McCarthy. (2010). *The Routledge Handbook of Corpus Linguistics*. USA and Canada: Routledge.
- McEnery, T. &. (2007). *Corpus Linguistics: Method, theory and practice*. Cambridge University Press.
- Mujid F Amin 1, S. S. (2021). Syntactic Errors of Foreign Speakers in Learning Indonesian: A Case Study of Darmasiswa Learning Process in Semarang City. *EDP Sciences*, 317. <https://doi.org/10.1051/e3sconf/202131702028>
- NorFariza, M. (2020). Corpus Driven Analysis of News Reports about Covid-19 in a Malaysian Online Newspaper. *GEMA Online Journal of Language Studies*, 20(30), 199-220. <https://doi.org/10.17576/gema-2020-2003-12>
- Pritchard, B. (2015). On multiword lexical units and their role in maritime dictionaries. *Journal of academic purposes*, 1 (4), 40-64.
- Ramlan. (2005). *Ilmu Bahasa Indonesia Sintaksis*. Yogyakarta: C.V. Karyono.
- Ruusila, A. &. (2016). *Conditio sine qua non: on phraseology in legal language and its translation*. *Journal Language and Law*, 3 (1), 120-140.
- Sugiyono. (2011). *Metode Penelitian Kombinasi (Mixed Methods)*. Bandung: Alfabeta.CV.
- Tarigan, H. G. (2008). *Menulis Sebagai Suatu Keterampilan berbahasa*. Bandung: Angkasa.
- Tognini, Bonelli. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.6>
- Zhou, W. (2020). *The Coronavirus Prevention Handbook*. Wuhan: China: Skyhorse Publishing.