

## **A CORPUS-BASED LEXICAL ANALYSIS OF COVID-19 SPECIAL MESSAGES BY THE FORMER PRIME MINISTER OF MALAYSIA**

Shazila Abdullah, \*Norasyikin Abdul Malik, Mohamad Syafiq Ya Shak & Nuramira Anuar

Academy of Language Studies, Universiti Teknologi MARA Cawangan Perak,  
Seri Iskandar Campus, 32610 Seri Iskandar, Perak, Malaysia.

\*Corresponding author: [noras967@uitm.edu.my](mailto:noras967@uitm.edu.my)

Received: 17.03.2023

Accepted: 23.06.2023

### **ABSTRACT**

**Background and Purpose:** The global emergence and rise of the COVID-19 pandemic in 2020 seriously affect a nation's economic and social well-being, including Malaysia. During the lockdown period, the former Prime Minister delivered special messages on television about updates on COVID-19 issues in Malaysia. A specialised Malay corpus, i.e., Cov-19 Related Special Messages Corpus, consisting of 19 Special Messages texts by Muhyiddin Yasin, the former Prime Minister of Malaysia, was examined to reveal its lexical profiling.

**Methodology:** The specialised corpus was analysed using the Sketch Engine tool, fully utilising the Wordlist, Keyword, and Concordance functions for lexical analysis. The list of lexical items was then compared with the same list of items in the Malaysian Web (MalaysianWaC), a default Malay Reference Corpus set in Sketch Engine.

**Findings:** Through wordlist analysis, the most frequently used words in the analysed corpus are generally linked to political or government entities (e.g., *kerajaan*, *menteri*, and *perdana*) and the welfare of the Malaysians (e.g., *rakyat*, *kesihatan*, and *epidemik*). The keyword analysis indicates that words related to the containment of the pandemic and reference to the Malaysians (e.g., *saudara-saudari*, *pelitup*, and *penjarakan*) show high keyness values. Finally, through concordance analysis, some words with high keyness scores are used differently in specialised Malay and Reference Corpus. From the analyses, the most used words in the specialised corpus are related to the issue of COVID-19.

**Contributions:** This study's findings contribute to a deeper understanding of language use during a critical period in Malaysia's history. The analysis highlights the linguistic emphasis on government, societal welfare, and pandemic-related containment measures, offering valuable references for future studies on crisis communication and lexical profiling in the Malay language.

**Keywords:** Corpus analysis, specialised corpus, COVID-19, WordSmith Tool, lexical analysis.

**Cite as:** Abdullah, S., Malik, N. A., Shak, M. S., & Anuar, N. (2023). A corpus-based lexical analysis of Covid-19 special messages by the former prime minister of Malaysia. *Journal of Nusantara Studies*, 8(2), 44-72. <http://dx.doi.org/10.24200/jonus.vol8iss2pp44-72>

## 1.0 INTRODUCTION

COVID-19 was first detected in December 2019 in Wuhan City, Hubei Province in China (Zhu, Wei, & Niu, 2020). Then, the World Health Organisation has acknowledged this outbreak as a public health emergency of international concern on January 30, 2020, before it was declared as a world pandemic on March 11, 2020, as the outbreak continues to wreak havoc and claim lives around the world (WHO, 2020). As of the first week of May 2021, the number of COVID-19 cases shows no signs of stopping, with more than 150 million cases of COVID-19 and 3 million deaths being reported around the world, as some of the most affected regions such as India and Southeast Asian Countries are experiencing the fourth wave of the pandemic (OECD, 2020a).

To curb the spread of the pandemic, many countries including Malaysia have advocated lockdowns and home quarantine. The first phase of lockdown in Malaysia was announced by the country's former Prime Minister from March 18th, 2020, to April 28th, 2020 (Tang, 2022). Meanwhile, climbing cases of the pandemic has forced the country to impose another lockdown from January 2021 to February 2021 (Yusof, 2021). In May 2021, another set of targeted lockdowns were being imposed in designated areas around the country as the country embraces itself for the fourth wave of COVID-19 cases during the period of Eid, the most widely celebrated festivity in the Muslim-majority Malaysia (Yusof, 2021). These measures included restrictions on movement, gatherings, and international travel, along with mandatory 14-day quarantines for Malaysians returning from abroad. Educational institutions and non-essential services were also temporarily closed. As a result, the population experienced increased fear and anxiety.

Political leaders then have scrambled to reduce fear and calm anxiety among the population of their countries. Almost all of them have attempted to do so by delivering special messages on issues related to COVID-19, which are broadcast on national television. The former Prime Minister of Malaysia, Muhyiddin Yassin, also followed the similar measure, in which he delivered almost 30 Special Messages during these various lockdown period, with the first one being delivered on March 16th, 2020, live on all television channels and social media platform.

It is assumed that such a measure is needed during these uncertain times, as Yu (2020) suggested that it could encourage and boost public sentiment. Salayo (2020) seemed to echo this, as she mentioned that it could act as an instrument to elaborate a government's policy and agenda, while advancing its desire to bring people together in harmony and solidarity. Sultan and Muhammad Rapi (2020) in addition stated that this measure is crucial as it reveals how a "government portrays a pandemic event and influences citizens' behaviour through the use of language" (p. 252).

Due to its scale, magnitude, and implications that expand all aspects of life, the COVID-19 pandemic has been extensively studied in recent times, including in the field of linguistics. Joharry and Turiman (2020) for example, who reviewed the discourse in letters to editors in Malaysian newspapers, concluded that Malaysians are viewing this pandemic as a challenge for them. Meanwhile Nor Fariza and Adlyn Syahirah (2020) disclosed that news reports in Malaysia on the pandemic used collocates that signify fear, anxiety, and uncertainty. In the study, some of the collocates include 'spread', 'curb', 'fight', 'outbreak', and 'death', are associated with fear due to the rapid and wide-reaching impact of the virus on Malaysians' health, economy, and social well-being. The term 'outbreak' suggests an unpleasant and difficult-to-control event, inducing anxiety among the masses regarding the spread of Covid-19.

Several studies also attempted to examine how messages on the COVID-19 pandemic are being framed by governments and political leaders around the world. Among others, the US former president, Donald Trump seemed to employ expressions that incited abuse, humiliation, displeasure, and discrimination in his COVID-19 speeches and election rallies (Olimat, 2020). This is in a stark contrast with the New Zealand Prime Minister, Jacinda Ahern. Attaminy, Junining, and Khasanah (2020) illustrated that her use of language in relaying communication on COVID-19 tends to be more straightforward and fact-based. She also chose "good words so that people's worries do not get bigger" (p. 209).

In terms of the speech acts used in communications about COVID-19, African presidents utilised performative and constative speech acts in their COVID-19 communications to command, request, urge, appeal, direct, and inform their citizens about the disease and necessary measures (Anyanwu & Abana, 2020). For instance, Buhari commended governors, gave instructions to security agencies, and mentioned strategies. Ramaphosa expressed determination, gratitude, unity, and congratulations, while Akufo-Addo focused on implementation and the government's assistance with water bills. These speeches exemplify the presidents' active engagement and commitment to addressing the challenges posed by COVID-19.

In Southeast Asia, Ala'yun (2020), who analysed the COVID-9 related speeches by the Singapore Prime Minister, Lee Hsien Loong, discovered that he mostly used the speech acts of informing, stating, claiming, convincing, hypothesising, commanding, greeting, and hoping. For instance, the Prime Minister subtly conveys the message that Singaporeans should not excessively fear the new virus, as it can have negative consequences for them. He also emphasises the importance of Singaporeans coming together, supporting one another, and remaining calm as they face the challenges posed by the new virus in the concluding remarks.

Apart from that, recent research of linguistic analysis of the COVID-19 pandemic usually involves the use of metaphors and phrasal expressions in language to serve specific purposes and achieve goals. Amairah (2022) examined the metaphorical language used in Jordanian English newspapers to depict the COVID-19 disease. The study explores conceptual metaphors that liken COVID-19 to war, water, and a person. These metaphors are widely employed in the corpus to shape the understanding of the pandemic and reveal political and ideological implications. Zhang, Lin, and Jin (2022) also focused on metaphors in framing COVID-19 in Chinese newspapers editorials. This study shows that metaphors related to seasons, disease and medicine, and homework serve pragmatic purposes in the pandemic. They evoke emotions, simplify social issues, and emphasise the need for democratic governance in disease control based on shared socio-cultural knowledge.

However, to the best of the researchers' knowledge, studies in analysing special messages (*perutusan khas*) by Malaysian political leaders in general and its Prime Minister specifically are still limited. At least two studies were conducted to analyse the speeches by the country's premiere. Earlier, Kamil and Bahrn (2020) attempted to analyse the direct and indirect speech acts in the text of the Malaysian Prime Minister's special message on the first day the Movement Control Order (MCO), where they reported that "that expressives and directives were used the most in the direct speech acts, while the indirect speech acts carried

the purpose to direct or persuade the audience” (p. 1). Then, Anuar and Ngadiran (2021) examined Muhyiddin Yassin’s first maiden speech on March 2, 2020, a day after his official appointment as the prime minister for his use of politeness strategies in the speech. They discovered that the speech was overwhelmingly populated by positive politeness strategies as opposed the negative ones.

Nevertheless, these two studies were focusing on speeches that took place at the early stage of the COVID-19 pandemic before the full scale of its severity was widely known. The speech observed by Anuar and Ngadiran (2021) in addition did not focus on the pandemic at all as it outlined the visions of the Prime Minister for the country in general after he was appointed; only one sentence on the pandemic was merely offered in the final part of the speech where the Premiere asked Malaysians to take a good care of their health amidst the threat of COVID-19. Apart from that, these studies also centered on a single speech by the Prime Minister, rather than a series of special speeches that focused solely on COVID-19.

Hence, this study intends to analyse the lexical items related to COVID-19 outbreak in the series of special messages by the former Prime Minister of Malaysia. The findings of the study could disclose how language is being employed by the local government in managing an unprecedented outbreak like COVID-19. This is crucial as Kim and Kreps (2020) warned that communication mismanagement during this period could potentially heighten distress among the people who are already experiencing pandemic fatigues, while leading to more uncertainties, misinformation, and an extended period of the pandemic.

In addition, up until now, there is no corpus which has been built based on the series of special messages by the former Prime Minister of Malaysia. Hence, no analysis has been done before in terms of the lexical items that appear in such series of speeches. Therefore, it is essential for such a corpus to be compiled to see if the lexical items inside the speeches are deemed as fit for further analyses.

The purpose of this study is to identify the lexical items in special messages texts by Muhyiddin Yasin, the former Prime Minister of Malaysia on the updates of COVID-19 issues in Malaysia. This study is based on a specialised Malay corpus, compiled from a series of special messages delivered by the former Prime Minister from March 2020 to January 2021. It aims at identifying the most frequently used lexical items in special messages texts by Muhyiddin Yasin on the updates of COVID-19 issues, while determining the relation of these lexical items to the issues related to COVID-19. It will attempt to answer these questions:

1. What are the most frequently used lexical items in special messages texts by Muhyiddin Yasin, the former Prime Minister of Malaysia on the updates of COVID-19 issues?
2. How are the lexical items related to COVID-19 issue in Malaysia?

## **2.0 LITERATURE REVIEW**

### **2.1 Corpus-Based Studies on Specialised Corpora**

Corpora are noted as important sources for academic purposes- teaching, learning, and research. From manually transcribed corpora in the early 1700s, the field has advanced to digitalised mediums and innovative methodologies that we could witness today. In the context of Malaysia, we have benefited from the corpora research in Malay and English languages, respectively. However, largely, the existing corpora are in the English language due to the emergence of studies in pedagogy (Lee, 2010). In terms of Malay corpora, researchers believe that they are still underdeveloped (Bakar, 2020; Juhary et al., 2021, Omar et al., 2017; Rahim, 2014). Juhary et al. (2021) identified there are at least 15 Malay corpora, but four are inaccessible and two are not annotated. They propose that creating a new corpus would enhance the language and benefit speakers, users, and researchers by incorporating diverse Malay corpora. Rahim (2014) argued that despite corpus linguistics being utilised as a research methodology for Malay and English studies in Malaysia for the past thirty years, there has been a lack of substantial discussion regarding the development and advancement of corpus research in the country. It is suggested that corpora of various national contexts should be developed as this is vital to encourage ‘natural language processing’ (Alfred et al., 2012).

In relation to corpus linguistics, there are generally two main research approaches: corpus-based approach and corpus-driven approach (Tognini-Benelli, 2001). Despite both approaches are both practical and efficient in corpus linguistic research, nevertheless, they are very much differing between one another. Exclusively, corpus-based linguistics is a deductive top-down approach, that “involves the use of data in corpus to refute, exemplify, or validate a theory” (Lu, Ali, & Ghani, 2021, p. 122). Conversely, corpus-driven linguistics is an inductive bottom-up approach which permits researchers to identify patterns in a text from a neutral and unbiased perspective (Goźdz-Roszkowski, 2018). Since the present study involves a newly built specialised corpus which is only a small-sized and focused corpus, corpus-based approach was opted as the research design of this research. In corpus-based approach, the size of the corpus is not highly considered to be significant, if the sampled texts are sufficiently representative and balanced (Lu et al., 2021). According to Lewis (1997), through this analysis, we could categorise vocabulary (words with fixed meanings) and lexis (single words and word

combinations). Furthermore, lexical analysis would benefit as a basis for the growth of word list which is required to the public's specific needs and consequently improves their understanding of the situation. Based on past literature, a framework was designed for the purpose of the present study. Figure 1 illustrates the framework of the corpus-based approach employed in the present research.

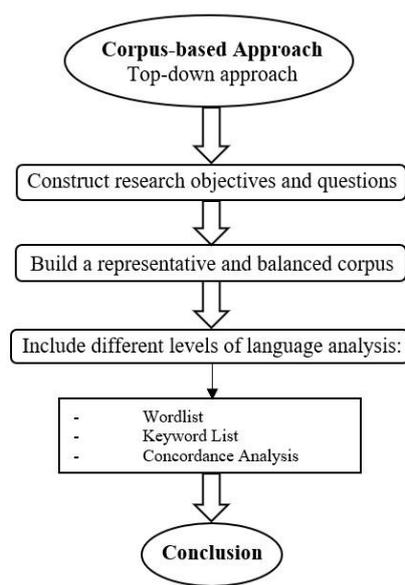


Figure 1: Framework of corpus-based approach

The specialised corpora examined in the study would be the special messages (*perutusan khas*) delivered by the former Prime Minister of Malaysia, Muhyiddin Yassin. Political speeches are considered as specialised texts devised to the speaker's specific communications goals and political agenda (Schaffner, 1997). The speeches are persuasive in nature. A few important elements to be considered regarding political speeches would be issues, locales, and purposes of delivering the speeches (Schaffner, 1997; Reisigl, 2008).

In a study by Attaminy et al. (2020) on speeches delivered by Jacinda Ardern, the Prime Minister of New Zealand, it was found that the speeches were direct, and she avoided figurative language that might undermine the pandemic situation. In addition, she also stressed on the situations that might occur in the future in relation to the pandemic in the speeches. In another study which examined the speeches made by the Prime Minister of Singapore, Lee Hsien Loong, it was reported that based on illocutionary speech analysis, the speeches contain directives, assertives, and expressives (Ala'yun, 2020). The assertives category includes speech acts such as informing, stating, claiming, convincing, and hypothesising. The directive

category consists solely of commanding, while the expressives category comprises greeting and hoping. It was also observed that Lee emphasised truths in his speeches and avoided promises and offerings to the citizens.

Meanwhile, in the context of Malaysia, Rajandran (2021) compared Malaysian and Singaporean Prime Ministers' speeches and found the use of war metaphors in media broadcasts to the public. COVID-19 is metaphorically depicted as an 'enemy' in Malaysia and Singapore. It was found that the terms 'silent enemy' and 'common enemy' are used by the respective Prime Ministers. These metaphors highlight the danger of the virus, although its characteristics are not clearly described. COVID-19 is intangible and can only be sensed through symptoms, making it a unique enemy. The battle against COVID-19 is described as 'unprecedented' as it requires simultaneously harming the virus while healing the human body.

In another study, Maisarah and Fatin Fatinah (2020), conducted a speech act analysis of the Prime Minister's speech during the Movement Control Order (MCO) period in Malaysia. From the qualitative study, it was found that in the direct speech acts, expressives and directives were most used. For instance, the choice of the word '*merayu*' ('beg') in the speech is a highly expressive term, particularly given his position as the country's leader. Meanwhile, the absence of any mitigation devices in conjunction with the words 'stay' and 'protect' indicates that the Prime Minister was issuing a direct instruction to the public to remain at home, reflecting a directive speech act. In contrast, to convince and direct the public, indirect speech acts classes were commonly used by Malaysia's former Prime Minister during MCO. Although this study analysed Malay language; however, it focused on the pragmatics aspects of the language- and much is less known on the Malay corpora- which would increase understanding on the topic. Other studies in the field that focused on Malay language reported that there are many words or terms added in Dewan Bahasa dan Pustaka (DBP) at the start of the spread of COVID-19 (Kasdan, Baharudin, & Shamsuri, 2020) which further suggests the need for lexical analysis. In another study, Anuar and Ngadiran (2021) investigated the positive and negative politeness strategies in the Malaysian former Prime Minister's maiden speech. In this study, it was found that the former Prime Minister applied positive politeness strategy in his speech such as 'notice or attend to receiver's interest, wants, or needs' so that he would be viewed as agreeable and having sincere intent towards the public. Better understanding of the phenomena could be achieved by a discussion surrounding the Malay language corpora which was limited at the time. Studies such as the present one would benefit in terms of enlightening researchers on the factors relating to politeness strategies employed by political speakers. Therefore, with the

above insights in mind, although both studies are important, as highlighted, they did not specifically analyse the lexical items of Malay language.

While there is a plethora of corpus-based studies on speeches delivered by political figures, corpus-based studies specifically looking at the use of lexical items in relation to the pandemic issues are limited, especially in Malaysian context. Hence, in relation to the present research, this study would contribute to the bilingual Malay-English corpora with the focus on Malay language based on the former Prime Minister's special messages broadcasted during the nationwide MCO in Malaysia. As the pandemic is discussed not only in English language but also in the native language, i.e., Malay, at this juncture, it is believed that the present study could help address the gaps outlined.

## **2.2 The Significance of Corpus-Based Analysis**

Texts, in both written and spoken form, could be difficult to comprehend when they are lexically dense. The advancement in terms of technology and methodology has contributed to the way texts in digital form are analysed today. Now, corpus-based research studies related to COVID-19 are limited. In relation to the present study, it is apparent that linguistic investigations in relation to the COVID-19 pandemic are blanketed with metaphors or phrasal expressions for specific purposes and goals using language. Building a wordlist surrounding the pandemic is crucial to help the public understand, address, and react to it, especially in a national context as exemplified by previous studies. Sinclair (2005) further specifies that the corpus must be selected according to the communicative usage in the community identified. As a starting point, in corpus-based research, identifying the keywords and the frequency of the words are vital. It is essential to identify words that were used frequently by the speechmakers in addressing the pandemic, concerning the present study, the Malaysian former Prime Minister. According to Nurmukhamedov and Webb (2019), word frequency lists and a representative corpus of a domain to examine the number of words is necessary to cover figures associated with the speeches with adequate comprehension.

## **2.3 The Need for a Keyword Analysis**

Keywords can be identified by making a comparison of word frequencies within a text with word frequencies of another larger reference corpus. According to Ensslin and Johnson (2006), in conducting keyword analysis, corpus linguists must be careful in building the corpus, and to decide on the selection of lexical items based on the research objectives. The selection of the significant lexical items is based on the keyness scores after the Cov-19 Special Messages

Corpus is compared with the Reference Corpus. Apart from the keyness scores, the lexical items are also selected based on the high relevancy to the COVID-19 pandemic. This is because, since the current research focuses on the lexical items used by the former Prime Minister of Malaysia in his special messages related to COVID-19 pandemic, the salient lexical items that are related to the pandemic were prioritised to be further examined in the later stage.

#### **2.4 Concordancing the Keyword**

A linguistic analysis of the speeches from the concordance list was examined to shed light on the speeches made by the former Prime Minister of Malaysia on the COVID-19 pandemic. According to Baldwin and Awab (2006) regarding the present study, since Malay imitates English in terms of its character encoding, pre-existing concordance tools for English could be used. Concordancing, according to O’Keeffe and McCarthy (2012), is a useful analytical technique to researchers as many items can be brought together in one place while maintaining their original contexts. Thus, applying these corpus linguistics methods to analyse a specialised corpus will contribute to the significance of the corpus itself. In particular, the corpus-based study of speeches made by national leaders, such as the former Prime Minister of Malaysia, on the COVID-19 pandemic will further add to the global discussion on the crisis from corpus linguists’ perspectives.

### **3.0 RESEARCH DESIGN**

In corpus linguistics, researchers generally employ either two major research approaches: “corpus-based” and “corpus-driven” (Biber, 2015). According to Tognini-Bonelli (2001), corpus-based studies generally utilise corpus data to investigate a theory or hypothesis with the aim to prove it, disprove it, or improve it. Conversely, corpus-driven studies deny the depiction of corpus linguistics as a method. Instead, the corpus itself should be the sole foundation of hypotheses about language. In other words, the research is more inductive (Biber, 2015), so that the corpus itself embodies a theory of language (Tognini-Bonelli, 2001). Since the aim of this study is not to construct new theory or hypotheses, a corpus-based approach was selected in investigating a specialised Malay corpus of the Special Messages related to COVID-19 made by Muhyiddin Mohd Yasin, the former Prime Minister of Malaysia. A corpus analysis of the former Prime Minister’s special messages was executed to attain insight and understanding about his speeches with regards to COVID-19 issues. The present study involves the analysis of wordlists, keywords, and concordance lines in identifying the salient features of the lexical items used in the former Prime Minister’s COVID-19 related Special Messages corpus.

### **3.1 Cov-19 Special Messages Corpus**

The specialised Malay corpus used in this study, which is referred to as the Cov-19 Special Messages Corpus in the analysis, was built from the COVID-19 related special messages delivered by the former Prime Minister of Malaysia, Muhyiddin Mohd Yasin. The special messages, that were delivered in Malay language, were retrieved from the Prime Minister's Office of Malaysia official website (<https://www.pmo.gov.my/speech/>). The researchers decided to collect the data from March 13th, 2020. This is because during that time the cases of COVID-19 started to spike and prompted the government to impose the first Movement Control Order (MCO) in the country. By using the keyword 'COVID-19', a total of 27 speeches of the special messages, dated from March 13th, 2020, until January 31st, 2021, were extracted. An average of six (6) special messages were delivered every month. Nevertheless, not all speeches were related to COVID-19, hence, due to this reason, only a small number of texts could be compiled to build this specialised Malay corpus. Even though it is understood that a corpus should be maximally representative of a variety and avoid being skewed (Agnostou & Weir, 2006), the limited number of available and suitable speeches should be enough to represent the discourse made by Muhyiddin Yassin in his special messages. This is because, if the constructed small corpus can address the issue by collecting samples that are characteristic of the specific speech domain being represented, the representative of the small corpus can be ensured (Vaughan & Clancy, 2013). The Cov-19 Special Messages Corpus contains a total of 33,805 words. While it may be considered small for a corpus-based research paper, it is sufficient for the lexical profiling objective. Even small corpora can provide meaningful insights, as Sinclair (2005) noted that frequent linguistic features are significant even in a corpus of 2000 words. The Cov-19 Special Messages Corpus represents complete speech events, making it appropriate for analysis and generalisation. Xiao (2010) emphasised the importance of a corpus being balanced and able to address research questions. In this case, the specialised corpus focuses on the speeches of the former Prime Minister of Malaysia regarding COVID-19, fulfilling both situational and linguistic criteria according to Biber's (1993) perspectives. By situational here it means that the specialised corpus covers a specific account of event, i.e., the speeches delivered by the former Prime Minister of Malaysia; and linguistically speaking, the corpus is based on COVID-19 related lexical items.

### 3.2 The Corpus Tool: Sketch Engine

The corpus-based analysis used the Sketch Engine (Kilgarriff et al., 2004) tool, a leading online corpus management tool that has been used by many corpus analysts for lexicography and language studies since 2004. The software enables corpus researchers to do two tasks: investigating corpora and building and managing corpora. In addition, the corpus software has high-level resources not just for English, but also for many languages such as Chinese, Spanish, Arabic, Japanese, Russian, Portuguese, and many more, including Malay (Kilgarriff et al., 2014), hence, makes it a suitable corpus tool to be used in the study. In terms of its features, Sketch Engine supported several languages in terms of POS tagging, lemmatisation, word sketches, and terms, but Malay language is not one of them (Kilgarriff et al., 2014). Nevertheless, the corpus software acknowledges roman alphabetical languages like the Malay language, thus, wordlist, keyword, and concordance lines of the keywords are possible to be generated (Sketch Engine, 2021). Due to this, the corpus analysis involved in the present study is limited to the analyses of wordlist, keyword list, and concordance lines. Finally, Sketch Engine possesses a default Malay Reference Corpus, i.e., Malaysian Web (MalaysianWaC) with 182,578,743-word count (Sketch Engine, 2021), making it easier for researchers of Malay corpus to analyse their data since there is a scarcity of publicly available pre-existing Malay reference corpus (Baldwin & Awab, 2006).

### 3.3 The Methodological Procedure of Corpus-Based Analysis

To explicitly explain the methodological procedure involved in this study, a step-by-step flowchart of the corpus-based analysis was devised as follows:

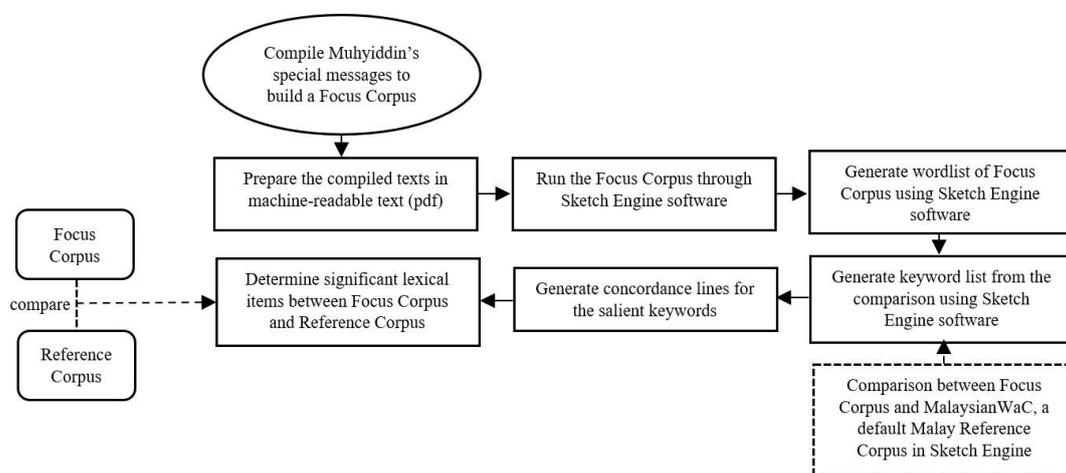


Figure 2: Flowchart of corpus-based analysis used in the study

Once the identified special messages which are related to COVID-19 were retrieved, the speeches were converted into PDF format before they were uploaded into the corpus software since Sketch Engine allows data to be processed in the said format. Cov-19 Special Messages Corpus examined in the present study has a total of 33,805 words. By using the Sketch Engine, a wordlist was generated based on the compiled specialised corpus. The wordlist displays the frequency of words that are used in the Cov-19 Special Messages Corpus. From the wordlist, as the purpose of this study is to investigate the use of lexical items, only the lexical items were then extracted. The list of lexical items was then compared with the same list of items in the Malaysian Web (MalaysianWaC), a default Malay Reference Corpus set in Sketch Engine with 182,578,743-word count (Sketch Engine, 2021). The reason to use a Reference Corpus in this study is to contrast the frequency and percentage of use of the lexical items in both corpora, hence salient keywords can be identified in the Cov-19 Special Messages Corpus. A keyword analysis is necessary to identify significant lexical items in the COVID-19 Special Messages Corpus. The analysis involves comparing word frequencies with a larger reference corpus and selecting key items based on keyness scores and relevance to the pandemic. According to Geluso and Hirsch (2019), even though common keywords representing a target corpus can be generated without the use of a reference corpus, utilising a reference corpus in an analysis can give a substantial difference. They further explained that the use of a reference corpus of the same sub-register as the target corpus can better accentuate the content unique to each target corpus. Meanwhile, when a reference corpus of a different register is used, words that reflect the register that the target corpora represent can be better revealed (Geluso & Hirsch, 2019). This would indicate whether the lexical items used in the Cov-19 Special Messages Corpus are only commonly used in the corpus or otherwise. Next, the keyword analysis was performed on both corpora to extract a list of lexical items with high keyness values that would be significant for further analysis. These lexical items were then run through the concordance function in the Sketch Engine to display concordance lines for a textual linguistic analysis of the same lexical items used in both corpora. The concordance list provides linguistic insights into the former Prime Minister's speeches on the pandemic. This linguistic analysis would further unveil the salient of the significant lexical items used in Muhyiddin Yasin's special messages relating to COVID-19.

## **4.0 ANALYSIS AND DISCUSSION**

This section presents the results based on the corpus-based analysis mentioned previously. The salient findings of the results will be further discussed in relation to COVID-19 related speeches. The research questions of this study will then be answered based on the discussion of the results. However, since this corpus-based analysis only involves the analysis of lexical items with the purpose of determining the possibility of using this Cov-19 Special Messages Corpus for further research in the future, the discussion of the findings shall revolve around the lexical items and not beyond this.

### **4.1 The Wordlist Analysis**

The data on the Wordlist analysis show the frequency of each word used in the Cov-19 Special Messages Corpus. Since lexical analysis involves the analysis of vocabulary items which can reveal the nature of a text, only the lexical items were extracted from the Wordlist. 49% of the words were found to be lexical items out of the first 200 frequently used words in the corpus. And from these 97 lexical items, 13 words have more than 100 occurrences in the Cov-19 Special Messages Corpus. Table 1 shows the list of the first 200 frequently used words in the examined corpus, whereas Table 2 shows the list of 13 words with more than 100 occurrences.

Table 1: Wordlist of 200 frequently used words in the COV-19 special messages corpus

No	Word	Freque	No	Word	Freque	No	Word	Freque	No	Word	Freque
1	-	2,463	51	<i>dari</i>	95	10	<i>sentias</i>	56	151	<i>sabah</i>	41
2	-	1,485	52	<i>terus</i>	94	10	<i>ramai</i>	56	152	<i>mene</i>	41
3	<i>yang</i>	1,141	53	<i>boleh</i>	93	10	<i>sektor</i>	56	153	<i>kadar</i>	40
4	<i>dan</i>	1,084	54	<i>itu</i>	93	10	<i>vaksin</i>	56	154	<i>atas</i>	40
5	<i>1</i>	755	55	<i>saudara</i>	92	10	<i>bebera</i>	56	155	<i>kehid</i>	40
6	<i>ini</i>	678	56	<i>saudara</i>	90	10	<i>menga</i>	56	156	<i>beker</i>	39
7	<i>di</i>	581	57	<i>negeri</i>	90	10	<i>tahun</i>	55	157	<i>progr</i>	39
8	<i>kita</i>	507	58	<i>rm</i>	89	10	<i>serta</i>	55	158	<i>kasihi</i>	39
9	<i>untuk</i>	502	59	<i>penular</i>	89	10	<i>bahawa</i>	55	159	<i>memb</i>	39
1	<i>saya</i>	474	60	<i>dapat</i>	89	11	<i>jumlah</i>	54	160	<i>diben</i>	39
1	<i>akan</i>	408	61	<i>Kes</i>	88	11	<i>antara</i>	54	161	<i>situas</i>	39
1	<i>2</i>	407	62	<i>2020</i>	87	11	<i>pekerja</i>	53	162	<i>selain</i>	38
1	<i>9</i>	361	63	<i>orang</i>	87	11	<i>bulan</i>	53	163	<i>penin</i>	37
1	<i>/</i>	332	64	<i>rumah</i>	87	11	<i>pkp</i>	53	164	<i>kluste</i>	37
1	<i>denga</i>	331	65	<i>hari</i>	85	11	<i>mesti</i>	52	165	<i>8</i>	37
1	<i>dala</i>	318	66	<i>sop</i>	85	11	<i>semula</i>	52	166	<i>meng</i>	37
1	<i>keraj</i>	310	67	<i>secara</i>	83	11	<i>iaitu</i>	51	167	<i>setiap</i>	37
1	<i>covid</i>	306	68	<i>kement</i>	81	11	<i>sebagai</i>	51	168	<i>pand</i>	37
1	<i>telah</i>	276	69	<i>covid</i>	80	11	<i>kes-kes</i>	51	169	<i>sahaj</i>	37
2	<i>negar</i>	262	70	<i>perlu</i>	79	12	<i>ialah</i>	50	170	<i>sebel</i>	37
2	<i>3</i>	252	71	<i>jangkita</i>	78	12	<i>6</i>	49	171	<i>diberi</i>	37
2	<i>5</i>	237	72	<i>pergera</i>	77	12	<i>wa</i>	49	172	<i>allah</i>	36
2	<i>kepad</i>	232	73	<i>19</i>	77	12	<i>awam</i>	49	173	<i>priha</i>	36
2	<i>mente</i>	214	74	<i>kawalan</i>	75	12	<i>perniag</i>	49	174	<i>mema</i>	36
2	<i>0</i>	203	75	<i>anda</i>	74	12	<i>7</i>	48	175	<i>melib</i>	36
2	<i>juga</i>	202	76	<i>ada</i>	73	12	<i>sosial</i>	48	176	<i>jadi</i>	36
2	<i>bagi</i>	187	77	<i>juta</i>	72	12	<i>;</i>	48	177	<i>mung</i>	35
2	<i>atau</i>	184	78	<i>masih</i>	71	12	<i>jika</i>	47	178	<i>wala</i>	35
2	<i>perda</i>	174	79	<i>seperti</i>	71	12	<i>the</i>	47	179	<i>bany</i>	35
3	<i>pejab</i>	174	80	<i>lagi</i>	70	13	<i>keadaa</i>	46	180	<i>amat</i>	35
3	<i>tidak</i>	172	81	<i>satu</i>	69	13	<i>tetapi</i>	46	181	<i>memb</i>	35
3	<i>malay</i>	171	82	<i>kerana</i>	65	13	<i>ditetap</i>	45	182	<i>hingg</i>	34
3	<i>semu</i>	171	83	<i>perintah</i>	62	13	<i>pelaksa</i>	45	183	<i>komu</i>	34
3	<i>pada</i>	164	84	<i>ingin</i>	62	13	<i>virus</i>	45	184	<i>baik</i>	34
3	<i>ekono</i>	150	85	<i>pihak</i>	61	13	<i>penting</i>	45	185	<i>petug</i>	34
3	<i>oleh</i>	148	86	<i>sebanya</i>	61	13	<i>melalui</i>	45	186	<i>syari</i>	33
3	<i>ke</i>	139	87	<i>positif</i>	60	13	<i>kesela</i>	45	187	<i>ming</i>	33
3	<i>rakya</i>	138	88	<i>sehingg</i>	60	13	<i>keluarg</i>	45	188	<i>luar</i>	33
3	<i>kesih</i>	134	89	<i>usaha</i>	60	13	<i>diri</i>	45	189	<i>perin</i>	33
4	<i>darip</i>	123	90	<i>2021</i>	60	14	<i>melaks</i>	44	190	<i>apa</i>	33
4	<i>adala</i>	122	91	<i>tempat</i>	59	14	<i>:</i>	43	191	<i>pusat</i>	32
4	<i>lebih</i>	119	92	<i>supaya</i>	59	14	<i>bukan</i>	42	192	<i>kuara</i>	32
4	<i>waba</i>	115	93	<i>sekalian</i>	59	14	<i>tindaka</i>	42	193	<i>masy</i>	32
4	<i>bahar</i>	110	94	<i>4</i>	58	14	<i>dilaksa</i>	42	194	<i>beria</i>	32
4	<i>tempo</i>	102	95	<i>memasti</i>	58	14	<i>bersam</i>	42	195	<i>and</i>	32
4	<i>)</i>	101	96	<i>sedang</i>	58	14	<i>seluruh</i>	42	196	<i>mena</i>	31
4	<i>masa</i>	100	97	<i>berada</i>	58	14	<i>tak</i>	42	197	<i>berla</i>	31
4	<i>sauda</i>	98	98	<i>termasu</i>	57	14	<i>aktiviti</i>	42	198	<i>pesak</i>	31
4	<i>(</i>	96	99	<i>peratus</i>	57	14	<i>hospita</i>	42	199	<i>sepan</i>	31
5	<i>merek</i>	95	10	<i>lain</i>	57	15	<i>bawah</i>	41	200	<i>10</i>	31

Table 2: List of 13 lexical items with a frequency of > 100

No	Item (English translation)	Frequency
1.	<i>kerajaan</i> (government)	310
2.	covid- (covid-)	306
3.	<i>negara</i> (country)	262
4.	<i>menteri</i> (minister)	214
5.	<i>perdana</i> (prime)	174
6.	<i>pejabat</i> (office)	174
7.	malaysia (Malaysia)	171
8.	<i>ekonomi</i> (economy)	150
9.	<i>rakyat</i> (the people)	138
10.	<i>kesihatan</i> (health)	134
11.	<i>wabak</i> (epidemic)	115
12.	<i>baharu</i> (new)	110
13.	<i>tempoh</i> (duration)	102

To consider that the frequently used lexical items are exclusive to COVID-19 related speeches, the frequency of these lexical items is then compared to the frequency of the same lexical items in the Malaysian Web (MalaysianWaC), a default Malay Reference Corpus in Sketch Engine. Table 3 exemplifies the list of frequently used lexical items in the Cov-19 Special Messages Corpus as compared to the Reference Corpus.

Table 3: List of frequently used lexical items in the Cov-19 special messages corpus as compared to Reference Corpus

No	Item (English Translation)	Cov-19 Special Messages Corpus		Reference Corpus	
		Frequency	Percentage	Frequency	Percentage
1.	<i>kerajaan</i> (government)	310	0.91	165,204	0.090
2.	<i>covid-</i> (covid-)	306	0.90	0	0
3.	<i>negara</i> (country)	262	0.77	290,147	0.158
4.	<i>menteri</i> (minister)	214	0.63	85,333	0.046
5.	<i>perdana</i> (prime)	174	0.51	38,080	0.020
6.	<i>pejabat</i> (office)	174	0.51	30,608	0.016
7.	<i>malaysia</i> (Malaysia)	171	0.50	224,320	0.122
8.	<i>ekonomi</i> (economy)	150	0.44	88,759	0.048
9.	<i>rakyat</i> (the people)	138	0.40	187,256	0.102
10.	<i>kesihatan</i> (health)	134	0.39	26,541	0.014
11.	<i>wabak</i> (epidemic)	115	0.34	2,189	0.0011
12.	<i>baharu</i> (new)	110	0.32	3,385	0.0018
13.	<i>tempoh</i> (duration)	102	0.30	27,176	0.014

As exemplified in Table 3, the word ‘COVID-’ is not even included in the Reference Corpus. This is most probably due to the period of data compilation which was before the pandemic broke out globally in late 2019. Therefore, these 13 words have a high tendency to be exclusive to COVID-19 speeches.

Analysing these 13 lexical items further reveals more interesting findings. The lexical items such as *kerajaan* (government), *negara* (country), *menteri* (minister), and *perdana* (prime) that can be clustered under government or political entity indicate that the former Prime Minister highlighted the role of the government during the pandemic.

Apart from that, the lexical items *Malaysia* and *rakyat* (the people) related to Malaysia and its people. The inclusion of a subject in a discourse usually aims to engage the public (Hermans & Gyldensted, 2019). Therefore, it is not surprising for a Prime Minister to use the words frequently when addressing the people (the *rakyat*) in his or her discourse to engage the people

to accept the delivered message. As for the word *wabak* (epidemic), it is related to the subject matter itself that is COVID-19, the virus that has caused a global pandemic, hence the reason for the delivery of these speeches by the former Prime Minister. Whereas the words *kesihatan* (health) and *ekonomi* (economy) indicate the former Prime Minister's concern with the people's welfare during the pandemic. This is also expected as the COVID-19 pandemic has been portrayed as 'the worst global health crisis of our time' affecting colossal loss and harm to people's health, jobs, and wellbeing (OECD, 2020b). Meanwhile, *baharu* (new) and *tempoh* (duration) may be related to either one's country or COVID-19. To further justify these findings, a further qualitative analysis of the whole discourse is required.

In summary, the lexical analysis of the Cov-19 Special Messages Corpus can be seen from the wordlist analysis revealing the most frequently used lexical items that may be pertinent to a speech delivered by a Prime Minister, and about COVID-19 which is the main subject matter of the speech. There have been other studies on political speeches (Rajindran, 2021; Anuar & Ngadiran, 2021), on COVID-19 (Joharry & Turiman, 2020; Nor Fariza & Adlyn Syahirah, 2020), and on COVID-19 based political speeches (Ala'yun, 2020; Anyanwu & Abana, 2020; Attaminy et al., 2020; Kamil & Bahrn, 2020; Maisarah & Fatin Fatinah, 2020; Ojimat, 2020), as mentioned in the Introduction and Literature Review sections. These studies, even though they have the same genre, that is political speeches, and/or a similar subject matter, which is COVID-19, they are not directly comparable to the present study as the objectives of these studies are not the same as they look beyond the corpus itself, analysis the discourse functions used in the corpus. The specialised corpora in those studies are only assumed to be situation- and time- sensitive and not linguistically examined to see whether it corresponds to the materials analysed as suggested by Biber (1993). Thus, the findings of the above comparison between the wordlist analysis of Cov-19 Special Messages Corpus and Reference corpus are evidence of situation- and time- sensitive corpus. Table 4 presents the top 10 lexical items with the highest keyness scores in both specialised corpus and Reference Corpus.

Table 4: Top 10 lexical items with the highest keyness scores in both the Cov-19 Special Messages Corpus and Reference Corpus

No	Item	Frequency (Cov-19 Special Messages Corpus)	Frequency (Reference Corpus)	Relative frequency (Cov-19 Special Messages Corpus)	Relative frequency (Reference Corpus)	Score
1	covid-	306	0	7293.2	0	7294.2
2	covid	80	0	1906.7	0	1907.7
3	<i>saudara-saudari</i>	90	525	2145.1	2.3	654.5
4	<i>pkp</i>	63	270	1263.2	1.2	582.1
5	<i>pelitup</i>	23	2	548.2	0	544.5
6	<i>sop</i>	85	696	2025.9	3	504.1
7	<i>rahmatullahi</i>	26	88	619.7	0.4	449.1
8	<i>pandemik</i>	37	241	881.9	1	431.5
9	<i>penjarakan</i>	25	114	595.8	0.5	399.3
10	<i>sosial</i>	32	232	762.7	1	380.5

Table 4 is extracted from the data of the keyword analysis between the Cov-19 Special Messages Corpus and Reference corpus. Although the lexical items listed in Table 4 are not entirely the same as the list of most frequently used lexical items shown in Table 3 earlier, there is one lexical item that appears in both lists, that is COVID-. There is also a very high keyness value when comparing the use of this word in both corpora. In keyword analysis, keyness value is used to determine the significance or importance of a lexical item within a corpus compared to a reference corpus. It is calculated by comparing the frequency of a word in the target corpus with its frequency in the reference corpus. The higher the keyness value, the more ‘key’ or distinctive the lexical item is in the target corpus compared to the reference corpus. This clearly indicates that the lexical items COVID-19 are indeed a characteristic of the Cov-19 Special Messages Corpus. Whereas the other eight lexical items in Table 4 also show high keyness values compared to other lexical items. The words are *saudara-saudari*, *pkp*, *pelitup*, *sop*, *rahmatullahi*, *pandemik*, *penjarakan*, and *kuarantin*. Even though it was mentioned earlier that the most frequently used lexical items could have a high tendency of being exclusive to COVID-19 speeches, the keyword analysis has proven otherwise. Therefore, it can be said that

the top 10 lexical words with high keyness values are the distinct features of the examined corpus.

#### **4.2 Concordance Analysis**

A further concordance analysis was conducted on the 10 lexical items with high keyness values to compare their usage in both corpora. Since the words are the distinct features of the Cov-19 Special Messages Corpus, analysing how they are used in sentences may reveal a different pattern of use which could also be a distinct feature of the examined corpus.

Table 5 shows the extraction of concordance lines from eight of the lexical items in Table 4. The words COVID- and COVID were not included as these words do not appear in the Reference Corpus.



7 *penjarakan*

sihatan. </s></s> Sentiasa amalkan **penjarakan** sosial, jaga kebersihan diri dan apabila  
> Selagi virus ini masih ada, amalan **penjarakan** sosial mesti diteruskan. </s></s> Kita n  
ementingkan keselamatan pekerja, **penjarakan** sosial dan kebersihan tempat bekerja.  
</s></s> </s></s> Amalkan **penjarakan** sosial. </s></s> Pakai penutup mulut di  
ITERI Jadi, berhati-hatilah. </s></s> **Penjarakan** sosial, memakai topeng mulut dan hid  
ih mematuhi SOP kesihatan seperti **penjarakan** sosial, pemakaian penutup mulut dan l  
mpat yang sesak. </s></s> Amalkan **penjarakan** sosial. </s></s> Pakai penutup mulut da  
> Saya fikir antara sebabnya ialah **penjarakan** sosial yang sentiasa diamalkan dalam  
tapa pentingnya kita mengamalkan **penjarakan** sosial untuk mengelakkan diri kita dari  
> PEJABAT PERDANA MENTERI **penjarakan** sosial walau di mana sahaja saudara-e

arolok-olokkannya, lalu dia berkata : </s></s> **Penjarakan** dia l </s></s> </s></s> Dengan itu, saudara akan me  
rik penyediaan, Asmah telah siap aku **penjarakan** </s></s> Usianya baru 17 tahun. </s></s> Ore  
ranan seperti binatang, diabikan dan di **penjarakan** selama enam tahun dimana makanan pun dib  
itu bahayanya Anwar, itu sebab beliau **penjarakan**, untuk memudahkan penggantian. </s></s> t  
im-macam strategi saya gunakan untuk **penjarakan** nafsu ini, sampai membuat perintah berkurun  
orang yang kononnya menyatakan nak **penjarakan** kamu. </s></s> Buat lah laporan Polis. Insaf le  
rbincangan itu, kata Abul Buhturi, kita **penjarakan** Muhammad. </s></s> Kata Abu Jahal dan orar  
slam. </s></s> " Yang tidak mau, saya **penjarakan**, " ungkapnya di sela-sela pelun-uran Majalah  
ih dunia dan masalah. </s></s> Jangan **penjarakan** dia di dalam keseksan nafsu yang memang li  
tg seperti Sheikh Umar Mukhtar atau di **penjarakan** tanpa perbicaraan seperti Ustaz Abu Bakar Al I

8 *kuarantin*

ara ini untuk menjalani perintah **kuarantin** selama 14 hari di stesenstesen kuar  
selama 14 hari di stesenstesen **kuarantin** yang ditetapkan oleh kerajaan. </s></s>  
kita. </s></s> Oleh itu peraturan **kuarantin** yang ketat di tempat-tempat tertentu  
pendapat rawatan atau menjalani **kuarantin** di tempat tinggal. </s></s> Pengas  
pekerja yang sedang menjalani **kuarantin** untuk datang bekerja kerana ini aka  
t saringan COVID-19 dan pusat **kuarantin**. </s></s> 33. </s></s> Bagi aktiviti p  
an sebagai Pusat Tahanan dan **Kuarantin** Sementara Pendatang Asing Tanpa  
ijadikan pusat pengasingan dan **kuarantin** sementara. </s></s> Usaha-usaha a  
yang lalu, saya telah menjalani **kuarantin** di rumah kerana dikategorikan seba  
> Peraturan-peraturan berkaitan **kuarantin**, pemakaian penutup hidung dan mu

in, Pusat Kastam, Imigresen dan **Kuarantin**, CIQ, bukan jalan tikus. </s></s> Lebih mi  
baidi berkata, kemudahan ladang **kuarantin** dan eksport yang dimiliki GLC membolehkan  
ri Perkhidmatan Pemeriksaan dan **Kuarantin** Australia (Aquis), kemudahan GLC juga li  
</s></s> Di samping itu, Khidmat **Kuarantin** dan Pemeriksaan Malaysia (MAQIS) akar  
bat. </s></s> Dan mereka perlu di **kuarantin** kan di rumah untuk tempoh tertentu jika ad  
diteruskan ialah penyinaran untuk **kuarantin** hama dan thrips (MARDI) dan kajian impa  
<kompleks Kastam, Imigresen dan **kuarantin** (CIQ Complex) di Bukit Cagar akan menji  
in 1984, Akta Pestsid 1984, Akta **Kuarantin** Tumbuhan 1976, Akta Perlindungan Kepel  
<kompleks Kastam, Imigresen dan **Kuarantin** (KIK) beroperasi, kerajaan memutuskan t  
Form yang dikeluarkan oleh pihak **kuarantin** Republik Rakyat China. </s></s> Pemeriks

The comparison of the concordance lines of the eight lexical items with high keyness scores discloses an intriguing finding as not all the words are used in the same way in both corpora. The words which have the same usage in both corpora are *saudara-saudari*, *pandemik*, *rahmatullahi*, and *kuarantin*. Whereas *pkp*, *pelitup*, *sop*, and *penjarakan* have different usages in the Cov-19 Special Messages Corpus. PKP in both corpora are abbreviations. As shown in the concordance lines, the abbreviations have different and several meanings in both corpora. However, in the Cov-19 Special Messages Corpus, the abbreviation has only one meaning which is *Perintah Kawalan Pergerakan* (Movement Control Order). The abbreviation is one of the words that may be most prevalent in the COVID-19 pandemic. The word *pelitup* has only two occurrences in the Reference Corpus. One of them is to indicate the cover of a roof and the other one to represent a mask. As for the use of *pelitup* in the Cov-19 Special Messages Corpus, the concordance lines clearly show that *pelitup* is used to refer to a face mask where the right collocate of *pelitup* is *muka* (face). Thus, it can be said that *pelitup muka* may be a collocation that is distinct to the Cov-19 Special Messages Corpus.

The word *sop* extracted using the Sketch Engine tool has another meaning in the Reference Corpus. This is because the tool does not differentiate the case of the word. Both lower and upper cases are extracted and therefore, the concordance lines in the Reference Corpus also show the use of *sop* as 'soup' or 'broth'. SOP is in fact an abbreviation of Standard Operating Procedure; the Cov-19 Special Messages Corpus has all occurrences of SOP to refer to this. This is not surprising as the breakout of the pandemic has resulted in several Standard Operating Procedures which are meant to curb the spread of the contagious virus.

And finally, the word *penjarakan*; it is the only word in the list that shows an entirely different meaning as compared to the Reference Corpus. *Penjarakan* in the Reference Corpus

is used to refer to ‘imprisonment’; whereas, in the Cov-19 Special Messages Corpus it shows ‘distance’ and all instances of *penjarakan* have a right collocates of ‘*sosial*’. *Penjarakan sosial* seems to be the translation of ‘social distance’, an expression that may be ubiquitously found in the Standard Operating Procedure of COVID-19 to control the spread of the virus. The difference in meaning indicates a need for a further morphological analysis of the word *penjarakan*. Although the spelling is similar, the root word may not be the same, causing the semantic meaning of the word to be different from how it has been normally used.

It is evident from both the keyword and concordance analyses that the most used words in the Cov-19 Special Messages Corpus are related to the issue of COVID-19. As a matter of fact, certain words like *pelitup*, *sop*, and *penjarakan* have meanings that are different when used in other contexts. This clearly would have implications towards the overall lexis of the language which could be the reason for the addition of new words or terms in Dewan Bahasa and Pustaka (DBP).

In terms of how these lexical items are linked to COVID-19 pandemic, it is evident that the former Prime Minister was found to be direct in his speeches. Based on the list of the most frequent keywords, no figurative language was used to deliver his messages. This is similar with the finding in Attaminy et al.’s (2020) study on speeches delivered by Jacinda Ardern, the Prime Minister of New Zealand in which her speeches were also found to be direct because figurative language might undermine the pandemic situation. The use of figurative language such as violent metaphors (e.g., war metaphor) in describing COVID-19 pandemic should be avoided as it may cause extreme anxiety, legitimise authoritarian governmental measures, and imply those who succumb to the pandemic did not fight hard enough (Semino, 2021).

Apart from that, most of the keywords indicate the government’s efforts in combating the pandemic. For example, the keywords *pkp* (referring to movement control order), *pelitup* (face mask), *sop* (standard operating procedure), *kuarantin* (quarantine), and *penjarakan* (social distancing) employed by the former Prime Minister in his speeches highlights the efforts made by the Malaysian government in combating the outbreak as well as call for action among the Malaysians to take their own initiative as preventive measures in stopping the spread of the virus. According to Malik, Shak, and Hasni (2022), when a newspaper highlights the role of a country’s government in combating a pandemic, it can be constructive. This is because the emphasis of any response from the governments such as providing awareness or actions to solve an issue can positively influence the readers or listeners (Fatima, 2020) in accepting the delivered message.

Finally, the keywords *saudara-saudari* that refers to the Malaysians and *rahmatullahi*, a part of Arabic utterance, are common salutations used in speeches in Malay language. These phrases of salutations portray an assertive politeness strategy. The evident is in line with Anuar and Ngadiran's (2021) study who found that the former Prime Minister applied positive politeness strategy in his speech to be viewed as agreeable and having sincere intent towards the public in his maiden speech. Effective use of speech acts by leaders in government can increase good governance via correct speech act choices and their perlocutionary acts (Akinkurolere & Ariyo, 2015).

## **5.0 CONCLUSION**

The corpus-based analysis of the Cov-19 Special Messages Corpus, in comparison with the Reference Corpus, has not only provided valuable insights into speeches about COVID-19, but it also carries significant implications for various stakeholders.

Firstly, the identification of distinctive features and salient lexical items related to COVID-19 in the corpus can greatly benefit policymakers and public health authorities. By understanding the language used in these speeches, they can effectively communicate important information, shape public perception, and devise targeted strategies to combat the pandemic.

Secondly, the evidence of language evolution within the corpus highlights the dynamic nature of communication during times of crisis. This insight can assist language researchers, linguists, and scholars in studying the linguistic changes and adaptations that occur in response to unprecedented situations, contributing to the broader field of sociolinguistics and language studies.

Moreover, the creation and analysis of the Cov-19 Special Messages Corpus addresses the existing gap in COVID-19 Malay corpora for research purposes. The availability of this specialised corpus enables future researchers to delve deeper into the pragmatic and linguistic aspects of the speeches, paving the way for comprehensive investigations into discourse patterns, rhetorical strategies, and sociocultural implications surrounding the pandemic.

In conclusion, this research not only enhances our understanding of COVID-19 discourse but also has practical implications for policymakers, contributes to the study of language evolution, and expands the research possibilities in the field.

## REFERENCES

- Akinkurolere, S. O., & Ariyo, K. S. (2015). Speech act features of select extracts in Chinua Achebe's *Arrow of God: Leadership perspective*. *Journal of Education, Society and Behavioural Science*, 9(2), 139-145.
- Ala'yun, A.V. (2020). *Critical discourse analysis of speech by Lee Hsien Loong about Coronaviruses in Singapore*. (Unpublished project paper). Universiti Muhammadiyah, Surakarta, Indonesia.
- Alfred, R., Jie, C. C., Wei, N. Z., Tahir, A., & Obit, J. H. (2012). Optimizing clusters alignment for bilingual Malay-English corpora. *Journal of Computer Science*, 8(12), 1970-1978.
- Amairah, H. (2022). Covid-19 is war, water & a person: Metaphorical language of the Coronavirus disease in "the Jordan Times" newspaper. *Theory and Practice in Language Studies*, 12(7), 1286-1293.
- Anagnostou, N. K., & Weir, G. R. S. (2006). From Corpus-based collocation frequencies to readability measure. <https://www.semanticscholar.org/paper/From-corpus-basedcollocation-frequencies-to-AnagnostouWeir/be513abe8f3aef31c372af18505c227db71984e0>
- Anuar, N., & Ngadiran, N. M. (2021). An appeal to the people: Politeness strategies in the Malaysian prime minister's maiden speech. *International Journal of Modern Languages and Applied Linguistics*, 5(2), 1-19.
- Anyanwu, E. C., & Abana, I. (2020). Speech act analysis of public speeches of three African Presidents amid COVID-19. *British Journal of English Linguistics*, 8(3), 10-18.
- Attaminy, H. F., Junining, H., & Khasanah, I. (2020). Speech act analysis of Jacinda Ardern about COVID-19. *Elite Journal*, 7(2), 209-224.
- Bakar, N. S. A. A. (2020). The development of an integrated Corpus for Malay language. In Alfred R., Lim Y., Havaluddin H., & On C. (Eds.), *Computational Science and Technology. Lecture notes in Electrical Engineering* (pp. 425–433). Springer.
- Baldwin, T., & Awab, S. (2006). Open source corpus analysis tools for Malay. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy*. European Language Resources Association.
- Biber, D. (1993). Representativeness in Corpus design. *Literary and Linguistic Computing*, 8(4), 243-57.
- Biber, D. (2015). Corpus-based and Corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistics analysis* (pp. 159–192). Oxford University Press.

- Ensslin, A., & Johnson, S. (2006). Language in the news: Investigating representations of 'Englishness' using WordSmith tools. *Corpora*, 1(2), 153-185.
- Fatima, S. S. (2020). *Understanding the construction of journalistic frames during crisis communication: Editorial coverage of COVID-19 in New York Times*. (Unpublished master's dissertation). Södertörn University, Stockholm.
- Geluso, J., & Hirsch, R. (2019). The reference corpus matters: Comparing the effect of different reference corpora on keyword analysis. *Register Studies*, 1(1), 209-242.
- Goźdz-Roszkowski, S. (2018). Between corpus-based and corpus-driven approaches to textual recurrence: Exploring semantic sequences in judicial discourse. In J. Kopaczyk & J. Tyrkkö (Eds.), *Applications of pattern-driven methods in corpus linguistics* (pp. 102–104). John Benjamins Publishing Company.
- Hermans, L., & Gyldensted, C. (2019). Elements of constructive journalism: Characteristics, practical application and audience valuation. *Journalism*, 20(4), 535-551.
- Joharry, S. A., & Turiman, S. (2020). Examining Malaysian public letters to editor on COVID-19 pandemic: A corpus-assisted discourse analysis. *GEMA Online Journal of Language Studies*, 20(3), 242–260.
- Juhary, J., Bakar, E. W., Shamsudin, M., & Alias, A. (2021). Understanding Malay corpora: A content analysis of 15 Malay corpora. *Journal of Advances in Linguistics*, 12(1), 18-26.
- Kamil, M. A., & Bahrn, F. F. S. (2020). A speech act analysis of the Prime Minister's speech pertaining to the Movement Control Order (MCO) to address the COVID-19 pandemic in Malaysia. *International Journal of Modern Languages and Applied Linguistics*, 4(4), 1-19.
- Kasdan, J., Baharudin, R., & Shamsuri, A. (2020). COVID-19 dalam Korpus peristilahan Bahasa Melayu: Analisis sosioterminologi [COVID-19 in the Corpus of Malay terminology: A socio-terminological analysis]. *GEMA Online Journal of Language Studies*, 20(1), 221-241.
- Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). ITRI-04-08 The Sketch Engine. Information Technology. [https://www.researchgate.net/publication/260387608\\_ITRI-04-08\\_the\\_sketch\\_engine](https://www.researchgate.net/publication/260387608_ITRI-04-08_the_sketch_engine)
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M. Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36.
- Kim, D. K. D., & Kreps, G. L. (2020). An analysis of government communication in the United States during the COVID-19 pandemic: Recommendations for effective government health risk communication. *World Medical & Health Policy*, 12(4), 398-412.

- Lee, D. Y. W. (2010). What corpora are available? In O'Keeffe, A. & McCarthy, M. (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107–121). Routledge.
- Lewis, M. (1997). *Implementing the lexical approach: Putting theory-into practice*. Language Teaching Publications.
- Lu, R., Ali, A. M., & Ghani, C. A. A. (2021). A comparative study of corpus based and corpus driven approaches. *Linguistics International Journal*, 15(2), 119-132.
- Maisarah, A. K., & Fatin Fatinah, S. B. (2020). A speech act analysis of the Prime Minister's speech pertaining to the Movement Control Order (MCO) to address the COVID-19 pandemic in Malaysia. *International Journal of Modern Languages And Applied Linguistics*, 4(4), 1.
- Malik, N. A., Shak, M. S., & Hasni, N. A. (2022). A Corpus-based approach to frame 'COVID-19 vaccination' in Malaysian English newspapers. *Jurnal Komunikasi: Malaysian Journal of Communication*, 38(4), 23–42.
- Nor Fariza, M. N., & Adlyn Syahirah, Z. (2020). Corpus driven analysis of news reports about Covid-19 in a Malaysian online newspaper. *GEMA Online Journal of Language Studies*, 20(3), 199–220.
- Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, 52(2), 188-200.
- OECD. (2020a). *OECD policy responses to Coronavirus (COVID-19): The territorial impact of COVID-19: Managing the crisis across levels of government*.  
<https://www.oecd.org/coronavirus/policy-responses/the-territorial-impact-of-covid-19-managing-the-crisis-across-levels-of-government-d3e314e1/>
- OECD. (2020b). *The world economy on a tightrope*.  
<https://www.oecd.org/economicoutlook/june-2020/>
- O'Keeffe, A., & McCarthy, M. (2012). *The Routledge handbook of corpus linguistics*. Routledge.
- Olimat, S. N. (2020). Words as powerful weapons: Dysphemism in Trump's COVID-19 speeches. *3L: The Southeast Asian Journal of English Language Studies*, 26(3), 17–29.
- Omar, N., Hamsani, A. F., Abdullah, N. A. S., & Abidin, A. Z. Z. (2017). Construction of Malay abbreviation corpus based on social media data. *Journal of Engineering and Applied Sciences*, 12(3), 468-474.
- Rahim, H. A. (2014). Corpora in language research in Malaysia. *Kajian Malaysia: Journal of Malaysian Studies*, 32(1), 1-16.

- Rajandran, K. (2021). 'A long battle ahead': Malaysian and Singaporean Prime Ministers employ war metaphors for COVID-19. *GEMA Online Journal of Language Studies*, 20(3), 261–267.
- Reisigl, M. (2008). Rhetoric of political speeches. In R. Wodak, & V. Koller (Eds.), *Handbook of communication in the public sphere* (pp. 243-269). Mouton de Gruyter.
- Salayo, J. D. (2020). Social distancing, community quarantine and bullets: A critical discourse analysis of Pres. Rodrigo Duterte's speeches on the war against COVID-19 pandemic. *Middle Eastern Journal of Research in Education and Social Sciences*, 1(2), 233-256.
- Schaffner, C. (1997). *Analysing political speeches*. Multilingual Matters Ltd.
- Semino, E. (2021). "Not Soldiers but Fire-fighters" – Metaphors and Covid-19. *Health Communication*, 36(1), 50-58.
- Sinclair, J. (2005). Corpus and text-basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1-16). Oxbow Books.
- Sketch Engine. (2021). Malay text corpora. <https://www.sketchengine.eu/corpora-and-languages/malay-text-corpora/>
- Sultan, & Muhammad Rapi. (2020). Positive discourse analysis of the Indonesian government spokesperson's discursive strategies during the COVID-19 pandemic. *GEMA Online Journal of Language Studies*, 20(4), 251-272.
- Tang, K. H. D. (2022). Movement control as an effective measure against Covid-19 spread in Malaysia: An overview. *Zeitschrift fur Gesundheitswissenschaften*, 30(3), 583-586.
- Tognini-Bonelli, T. (2001). *Corpus linguistics at work*. John Benjamins.
- Vaughan, E., & Clancy, B. (2013). Small Corpora and pragmatics. In J. Romero-Trillo (Ed.), *Yearbook of Corpus linguistics and pragmatics* (pp. 53-73). Springer.
- WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Xiao, R. (2010). Corpus creation. In N. Indurkha, & F. Damerau (Eds.), *Handbook of natural language processing, second edition* (pp. 147-165). CRC Press.
- Yu, L. (2020). Analysis of the queen's speech on COVID-19 from the perspective of transitivity. *Advances in Language and Literacy Studies*, 11(6), 7-11.
- Yusof, A. (2021, August 25). Timeline: How the COVID-19 pandemic has unfolded in Malaysia since January 2020. *Channel News Asia*.

<https://www.channelnewsasia.com/asia/timeline-how-covid-19-pandemic-has-unfolded-malaysia-january-2020-2082081>

Zhang, C., Lin, Z., & Jin, S. (2022). What else besides war: Deliberate metaphors framing COVID-19 in Chinese online newspaper editorials. *Metaphor and Symbol*, 37(1), 114-126.

Zhu, H., Wei, L., & Niu, P. (2020). The novel coronavirus outbreak in Wuhan, China. *Global Health Research and Policy*, 5(6), 1-3.