A CORPUS-BASED SENTIMENT ANALYSIS OF COVID-19 VACCINATION NEWS REPORTS

SITI NUR AINA BINTI MOHD HASHIM

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITI MALAYA
KUALA LUMPUR

2024

A CORPUS-BASED SENTIMENT ANALYSIS OF COVID-19 VACCINATION NEWS REPORTS

SITI NUR AINA BINTI MOHD HASHIM

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS (LINGUISTICS)

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITI MALAYA
KUALA LUMPUR

2024

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Siti Nur Aina Binti Mohd Hashim

Matric No: 17138682

Name of Degree: Master of Arts (Linguistics)

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

A Corpus-Based Sentiment Analysis Of COVID-19 Vaccination News Reports

Field of Study: Discourse Analysis

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 14 Mac 2024

Subscribed and solemnly declared before,

Witness's Signature Date: 14 Mac 2024

Name:

Designation:

A CORPUS-BASED SENTIMENT ANALYSIS OF COVID-19 VACCINATION NEWS REPORTS

ABSTRACT

COVID-19 is the world's most critical global health emergency at present and administering an effective vaccination program is crucial in keeping the pandemic under control. However, the mainstream views on COVID-19 vaccinations are rather divided. By using a corpus-based approach, this study intends to investigate how sentiments regarding COVID-19 vaccination are reflected in linguistic elements and how such sentiments change over time in a local online newspaper in Malaysia. Adopting a mixed method approach, this study employs NVivo and Wmatrix and the selected news articles will be carried out by descriptive analysis to gain insights into elements that constitute sentiments of COVID-19 vaccines. Furthermore, the linguistic elements are examined using discursive news value analysis (DNVA) in pursuance of the transition in sentiments between 2020 and 2022. Based on NVivo and Wmatrix results, the sentiment is negative as the words pertaining to vaccinations consist of more words with negative connotations compared to positive ones. The findings identified three themes in 2020 and eight themes in each of 2021 and 2022. The transition in vaccination sentiments was portrayed as positive from 2020 to negative in 2021 and neutral in 2022 as indicated by the amounts of themes. The limitation of this study is that the researcher only focuses on a limited time frame (the month of March in 2020, 2021 and 2022) and only indicates the sentiment at that certain period. This study is significant in providing insights into the public's attitudes, underlying concerns and acceptance of the vaccines, which can be utilized to inform and improve vaccination policies.

Keywords: COVID-19 vaccination, sentiment analysis, corpus-based approach, discursive news values analysis

ANALISIS SENTIMEN BERASASKAN KORPUS TERHADAP LAPORAN BERITA VAKSINASI COVID-19

ABSTRAK

COVID-19 merupakan sebuah kecemasan kesihatan yang dianggap kritikal oleh seluruh dunia dan pelaksanaan program vaksinasi amatlah penting untuk mengawal wabak ini. Meskipun begitu, pandangan umum tentang vaksinasi COVID-19 agak bercelaru. Dengan menggunakan pendekatan berdasarkan korpus, kajian ini bertujuan untuk menyelidik bagaimana sentimen mengenai vaksinasi COVID-19 dipaparkan dalam elemen linguistik dan bagaimana sentimen tersebut berubah dari semasa ke semasa dalam sebuah akhbar dalam talian tempatan di Malaysia. Dengan menggunakan pendekatan kaedah campuran, kajian ini menggunakan NVivo dan Wmatrix. Analisis deskriptif akan dilakukan terhadap artikel berita yang dipilih untuk mendapatkan pandangan tentang elemen-elemen yang membentuk sentimen vaksin COVID-19. Unsur linguistik yang diperolehi akan diperiksa menggunakan *Discursive News Values Analysis (DNVA)* antara tahun 2020 hingga 2022. Hasil daripada NVivo dan Wmatrix menunjukkan bahawa keseluruhan sentimen yang dilihat dalam korpus adalah negatif memandangkan kata kunci berkaitan dengan vaksinasi mengandungi lebih banyak kata-kata berkonotasi negatif. Kajian ini juga mengenal pasti tiga tema pada tahun 2020 dan lapan tema pada kedua-dua tahun 2021 dan 2022. Sentimen terhadap vaksinasi ini juga dilihat mengalami peralihan ketara memandangkan sentimen umum kelihatan positif pada tahun 2020 dan berubah kepada negatif pada tahun 2021, seterusnya menjadi neutral pada tahun 2022. Kajian ini dibataskan pada jangka masa yang terhad iaitu pada bulan Mac 2020, 2021 hingga 2022 serta memperlihatkan sentimen umum pada tempoh masa tersebut sahaja. Kajian ini dilihat penting untuk memberikan gambaran penerimaan umum, kebimbangan dan penerimaan vaksin yang seterusnya boleh digunakan oleh pihak berkuasa untuk memaklumkan dan menambah baik dasar vaksinasi sedia ada.

Kata kunci: vaksinasi COVID-19, analisis sentimen, pendekatan berasaskan korpus, *discursive news values analysis*

ACKNOWLEDGEMENTS

Finally, the day when my “InsyaAllah” turns into “Alhamdulillah” has arrived. All praise and thanks to Allah SWT for giving me His blessings to complete this study. My challenging and stressful Masters's journey has strengthened my faith in His perfect timing in every circumstance.

I would like to express my deepest gratitude to the most supportive, thoughtful and reliable supervisor, Dr. Yeo Siang Lee for providing guidance, crucial feedback, constantly checking up on me and assuring me whenever my self-doubt kicks in. I consider myself really fortunate to have you as a supervisor who is concerned not just with my work but also with my mental wellness.

To my parents, saying “Thank you” to someone who deserve so much more is hard. There’re no words to describe how grateful I am to have you in my life and I am forever thankful for your unconditional love and support and for always being there for me whenever I need you. Without your prayers, I wouldn’t be able to achieve anything. This dissertation is dedicated to you two.

To my siblings and sisters-in-law, your existence alone gives me so much strength and happiness. Nothing makes me feel stronger than knowing I have a family who has my back. Let’s bicker and fight each other as hard as we love each other until we are old.

To my friends, my miserable moments are not completely devastating just because you are there to support me and listen to me even when I don’t act like myself. You know who you are, thank you!

Last but not least, to my idol, your presence acts as my shield and makes me smile even when I’m in pain. I appreciate how your music illuminates my world.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
CBA	:	Corpus-Based Approach
CDA	:	Corpus-Driven Approach
CL	:	Corpus Linguistics
DNVA	:	Discursive News Value Analysis
LDA	:	Latent Dirichlet Allocation
ML	:	Machine Learning
NLP	:	Natural Language Processing
NRC	:	National Research Council
OM	:	Opinion Mining
PICK	:	Nationwide COVID-19 Immunization Campaign
PICKids	:	National COVID-19 Immunization Program for Children
SA	:	Sentiment Analysis
SNA	:	Social Network Analysis
USAS	:	UCREL Semantic Analysis System
WHO	:	World Health Organization

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Introduction

This chapter offers an overview and the background of the study which focuses on sentiment analysis news reports of COVID-19 vaccination in Malaysia. The analysis of how language is used to represent the sentiment towards vaccination in the news provides valuable insights into public opinion on vaccination. This chapter also includes the problem statement and the literature gap in previous studies, namely the lack of linguistic perspective in sentiment analysis. The research objectives and research questions, together with the significance of the study which underlines the contribution of the current study will also be included.

1.2 Background

Pandemics have happened periodically throughout human history, resulting in millions of fatalities, detrimental economic and societal impact on civilization. Without exception, the recent COVID-19 pandemic has caused far-reaching repercussions throughout the world due to its high infectivity and the lack of effective treatments at the beginning stages. The highly transmissible coronavirus illness 2019 (COVID-19) began spreading in late December 2019, affecting every element of life in all communities throughout the globe. Within the first year of the global pandemic, almost all countries were affected with varying degrees of infection and mortality rates. The World Health Organization (WHO) classified COVID-19 a pandemic in March 2020 taking into account more than 110,000 cases had been reported in over 114 countries (WHO, 2020).

The evolution of an effective vaccine and the implementation of an immunization program was vital to contain the transmission of COVID-19. COVID-19 vaccines were developed at a rapid rate than expected by a number of pharmaceutical firms and

institutions (Yang & Sornlertlamvanich, 2021). A year after pandemic had begun, the first vaccinations were finally ready and vaccination campaigns were implemented throughout the world. A number of vaccinations have been authorized across the world and among them are Pfizer/BioNTech, Moderna and AstraZeneca/Oxford (Marcec & Likic, 2022). The Malaysian government's flagship vaccination program, The Nationwide COVID-19 Immunization Campaign (PICK), began on February 24, 2021, in response to an increase in COVID-19 infections in Malaysia (Radu, 2021). While many nations were attempting to expedite the vaccination program so their citizens could be vaccinated, many individuals were perceptibly hesitant to participate (Alabid & Katheeth, 2021; Allington et al., 2021; Amjad et al., 2021; Yang & Sornlertlamvanich, 2021).

Public opinion and sentiments regarding the spread of infectious illnesses and how vaccination could prevent them have been a popular area of research as the public awareness of COVID-19 vaccination concerns grew during the pandemic. To study a society's acceptance of COVID-19 vaccines, many approaches have been adapted and adopted with Sentiment Analysis (SA) being one of them. SA is a burgeoning field at the crossroads of computational linguistics and computer science that aims to naturally detect sentiment in texts. Defined as computer-based automated approaches to extract people's opinion about a particular topic of interest (Esuli & Sebastiani, 2006), SA was originally invented to categorize texts that deal with subjective remarks, with the goal of extracting sentiments from natural language text using computer approaches. Due to the tremendous popularity of social media, there is an enormous amount of such materials which are publicly available. Each social media site, including lengthy blogs and forum posts, often contains a large amount of opinion texts which are challenging to process. It will be a struggle to find relevant sites, extract and summarize all the thoughts contained within them. As a result, automated SA tools are created and utilized. As SA has been found to be useful in extracting customer or public feedback on goods and services, its use has

expanded to many sectors in recent years, from consumer goods to social activities and political elections (Liu, 2015; Pang & Lee, 2008).

The techniques used by the machine to automatically annotate sentiment may be split into two categories: machine learning and lexicon-based methods (Nasukawa & Yi, 2003; Mountassir et al., 2013). Machine learning divides sentiment categories based on specified features where it gains knowledge from models that have been trained using an algorithm. To simplify, it starts with a collection of labelled training data that is used to construct a classification model that will be used to categorize new data that does not have labels. The lexicon-based method employs a word bank dictionary containing opinion terms and matches a specified collection of words in a text for detecting polarity. Unlike machine learning approaches, this approach does not need data preprocessing or the training of a classifier (Taboada et al., 2011).

With the growing demand of the Internet, studies of SA have exploded as the Internet has offered a simple and effective way for individuals to express their opinions (Qin, 2017). In addition, people can share their thoughts and experiences through numerous social media platforms (e.g., Twitter, Facebook, Instagram) and blogging platforms (e.g., WordPress, Blogspot, Wix). It is a particularly appealing source for SA considering the quantity and relevance of the accessible content where the users' thoughts and opinions are expressed in their platforms. For these reasons, many researchers (Karami et al., 2021; Luo et al., 2021; Nezhad & Deihimi, 2022) have tapped into the Internet to analyze people’s perceptions towards COVID-19 vaccine as the public has openly expressed how they feel about the issue via various online channels.

This enormous data set may be utilized to obtain valuable observation into the general public's thoughts and feelings about COVID-19-related topics. Nonetheless, the

language characteristics employed to determine the sentiments of vaccines are also important because they tend to reveal the true meaning and explanations of widely held vaccination beliefs. While the past studies provide useful information about COVID-19-linked issues, certain gaps have been found. First, most studies did not examine sentiments towards vaccination in a newspaper domain and from a linguistic standpoint. Second, there is no research observing and comparing the sentiment changes between 2020 and 2022, namely from the time before the vaccination program began to after the vaccination program had been implemented. As a result, this study is relevant and significant in contributing to the body of knowledge with regard to sentiments about COVID-19 vaccination, notably in Malaysia. By adopting a corpus-based approach, this research plans to study how language is utilized to express sentiments regarding COVID-19 vaccination in a local online newspaper in Malaysia. The current study also aims to explore the changes in the sentiments surrounding COVID-19 vaccination from 2020 to 2022. This research will be using news articles on vaccination in Malaysia from The Star, an online English Malaysian newspaper, as samples for analysis. NVivo and Wmatrix are the chosen software to process the data. Furthermore, this corpus-based research will reveal how Malaysians' opinions regarding vaccination have changed over time and thus, contributing to linguistics studies on COVID-19 vaccination.

1.3 Problem statement

COVID-19 vaccines have been created to help people develop immunity against potentially deadly infections caused by the coronavirus. However, there are many influencing factors that affect people's willingness to participate in the immunization program, thus making it difficult to gauge their acceptance of the vaccine. According to Liu et al. (2022) age, gender, education level, cognitive and behavior traits were the most important factors influencing an individual's decision on whether to receive or reject the vaccination. Some of the key concerns include the vaccine's safety, efficacy and probable

adverse effects. Not only that, the most significant contributing factors to people's hesitation in getting vaccinated are their skepticism about certain vaccine components and fear of severe side effects (Patterson et al., 2022). As it is normal to feel anxious about something new, some people are fearful of receiving COVID-19 vaccination, despite the fact that it has been shown to reduce infection rates, serious complications and hospitalization.

Until now, there is no systematic technique to assess public opinion on vaccinations in the mainstream media. Moreover, the sentiments of COVID-19 vaccine may vary from person to person or change over time based on the media platform they receive information from, as the information channel has a tremendous impact on how the readers' perspectives and perceptions are formed. With these in mind, SA has been chosen as the approach to systematically assess the opinions expressed towards vaccination in news sources in this study. This is also based on the ground of SA being an effective tool in helping researchers gauge how the general population feels about the vaccines and analyze in concrete terms the breakdown of linguistic elements that form the sentiments.

While previous studies focused on examining public opinions of COVID-19 (Barkur et al., 2020; Liu & Liu, 2021), recent studies began to examine the public perception on COVID-19 vaccines (Alabid & Katheeth, 2021; Karami et al., 2021; Luo et al., 2021; Nasyaya et al., 2023; Nezhad & Deihimi, 2022; Park & Suh, 2023; Saleh et al., 2023; Sattar & Arifuzzaman, 2021; Sutrave et al., 2021; Yang & Sornlertlamvanich, 2021; Yousefinaghani et al., 2021). Although SA has been used in various research to examine people's opinion of vaccination, the papers above have only analyzed the polarity of the sentiments and none incorporated the linguistic viewpoint. As of now, there is clearly a lack of research work in Malaysia that focuses on sentiments of vaccination from

a linguistic or discourse viewpoint. Not only that, while most previous SA research on COVID-19 vaccination relies on information gathered from social media platforms like Twitter and Facebook (Klimiuk et al., 2021; Liu & Liu, 2021; Nezhad & Deihimi, 2022; Park & Suh, 2023; Saleh et al., 2023; Sattar & Arifuzzaman, 2021; Sutrave et al., 2021; Yousefinaghani et al., 2021), the SA community has paid considerably less attention to the news genre (Bučar et al., 2018).

1.4 Research aim & questions

The concerns of SA in vaccination have a wide range of possibilities, not only because of the crisis, but also due to the controversies surrounding vaccination, which have caused doubts among the public about getting vaccinated. Along these lines, this study aims to investigate how sentiments with regard to COVID-19 vaccination are expressed in news reporting and to identify the emerging themes from the COVID-19 vaccination news reporting. Based on these research aims, this research attempts to answer the following research questions:

1. What are the linguistic elements used in news reporting to represent sentiments towards COVID-19 vaccination in Malaysia?
2. What are the emerging themes from the COVID-19 vaccination news reporting between 2020 and 2022?

1.5 Significance of study

In general, the mainstream view on COVID-19 vaccinations is ambiguous. This study will look at how language is used in a local Malaysian online newspaper to depict sentiments surrounding the COVID-19 vaccine, as well as how sentiments vary over time thus enabling us to obtain a better grasp of public perceptions of the vaccination. In other words, the findings of this research are to provide insights into linguistic elements that

are relevant to signify sentiments on news reports. Understanding the language aspects used to explain how people feel about vaccinations can help relevant parties, such as decision-makers, to plan and implement appropriate and adaptive vaccination policies in order to engender confidence and cooperation among the people. Not only will it help in detecting important issues and taking preemptive actions to give people more confidence in vaccinations, it will also be a critical component in combatting the COVID-19 pandemic. Plus, the current work is at the right time in light of the contribution to linguistics in relation to COVID-19 vaccination. In addition, the research is also expected to shed light on the changes in sentiments toward vaccination and themes from the COVID-19 vaccination news reporting throughout the times. In other respects, from a linguistics standpoint, the current research will contribute to the body of literature and research on sentiments of vaccination in the Malaysian newspaper context as none of the previous researchers has focused on both of this topic and domain. Furthermore, this work will boost to the small number of studies of sentiment on vaccination that use a corpus-based approach as past research has primarily focused on machine learning. On top of everything, the findings of the study may be used to suggest methods to exploit the linguistic components to enhance SA research and the construction of more sentiment systems for real-world applications as SA is particularly important since it allows us to get a deep sense of public viewpoint on certain issues. As a result, this work may open up opportunities for future research combining the areas of linguistics and the field of SA.

1.6 Overview of dissertation

This dissertation is separated into five chapters including this as the introductory chapter. The next chapter is the literature review which provides an overview of the related literature for the study. The methodological approaches used to conduct this research will be described in Chapter 3. Chapter 4 explains the data analyzed by using both Wmatrix and NVivo as well as the interpretations. The findings from the study are

organized in this chapter to best address the research questions. Last but not least, Chapter 5 brings the dissertation to a close by summarizing the research findings. It also highlights some of the implications and suggests some suggestions for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The background information on SA presented in this chapter is derived from two fields of study, namely computer science and linguistics. A massive number of studies have been carried out over the years with the aim to define and understand the concept of SA better particularly in the field of computer science. However, there has been little focus on the systematic study of how sentiment is detected in text manually for the development of Natural Language Processing (NLP) systems. Each study field offers useful information to help the researcher to understand more about the methods and applications of SA in text.

The first section of this literature review looks at several viewpoints about SA in text by distinguishing it from other similarly related areas to discover how these ideas are used to describe concepts linked to SA. The second section focuses on several development phases of SA in order to provide a fuller picture of its background and methodologies. The following section describes the background of COVID-19 vaccination, investigates the past literature on SA and last but not least examines some significant issues that has not been addressed by any previous study in this field.

2.2 Sentiment Analysis (SA)

In recent years, there is a growing body of research in the computational field in what is known as "sentiment analysis" (Pang et al., 2002; Pang & Lee, 2008; Taboada & Grieve, 2004, inter alia) and "subjectivity analysis" (Abdul-Mageed & Diab, 2012; Xuan et al., 2012; Wiebe et al., 2004).

Liu (2015) stated that Nasukawa and Yi (2003) may have coined the term ‘sentiment’ but according to Pang and Lee (2008) the word ‘sentiment’ is first used in the context of a task involving computational text analysis by Das and Chen (2001) and Tong (2001) owing to the researchers' interest in market sentiment. Das and Chen (2001) defined the term “sentiment” as having a precise meaning which refers to the sum of positive and negative sentiment expressed on a stock's message board. It then appeared in Turney (2002) and Pang et al. (2002) where both of the authors’ papers were published in the proceedings of the Association for Computational Linguistics' (ACL) annual conference.

In the recent years, SA has been defined as a field of study that have computers automatically identify and comprehend the opinions expressed by internet users on a certain topic. These opinions and remarks are then gathered for the system to evaluate and conduct a wide range of tasks on them (Mountassir et al., 2013; Pang & Lee, 2008; Taboada, 2016). SA may be defined as the process of identifying, extracting, and categorizing opinions, emotions, and sentiments represented in textual input, in anattempt to include the whole body of work that has been done in the area (Liu, 2015).

2.2.1 Related terms

Many terminologies associated with SA indicate that the meanings of these terms overlap significantly, and that various scholars looked at the subject from various angles and alternative terminology. Thus, this section will be providing a brief summary for each of the terminology that is related to SA.

The huge development and collection of online writings in the blogosphere has boosted subjectivity analysis research substantially in recent years (Abdul-Mageed & Diab, 2012; Panicheva et al., 2013). Subjectivity Analysis can be defined as the

computational job at hand to distinguish between sentences that provide opinions and phrases that present factual information objectively (Abdul-Mageed & Diab, 2012; Hatzivassiloglou & Wiebe, 2000; Wiebe et al., 2004). Wiebe et al. (2004) claimed that subjective language is a type of language that is used in a speech or text to represent a private state and described that subjectivity analysis is concerned with detecting "private states". Opinions, sentiments and emotions all fall under the umbrella of the phrase "private states" (Quirk et al., 1985). Subjectivity was defined by Wiebe et al. (2004) as an umbrella term that comprised opinions, emotions, sentiments and other private states. The author was motivated by linguist Banfield's (1982) work, which defines subjective sentences as sentences that assume a character's point of view. In short, Subjectivity Analysis aims to separate subjective and objective information in each text.

Montoyo et al. (2012) stated that due to the vague nature of the terms "subjectivity" and "sentiment," research in these fields has covered a wide range of techniques and applications, making it impossible to offer a single definition. The author also argued that a substantial body of research has tackled SA separately from Subjectivity Analysis, and it has, in many cases, been tied to the task of emotion detection. However, Pang and Lee (2008) observed that many researchers currently use the word ‘Sentiment Analysis’ to refer to the computational analysis of sentiment, opinion, and subjectivity in text. Wiebe and Mihalcea (2006) noted that in the literature, these two words ‘sentiment’ and ‘subjectivity’ are commonly used interchangeably. In this regard, both ‘sentiment’ and ‘subjectivity’ will be utilized interchangeably in this dissertation. Besides, Taboada (2016) expressed that from the linguistic point of view, the basic objective of SA is to figure out whether a text, or a portion of it, is subjective and, if so, whether it conveys a positive or negative sentiment.

Liu (2015) stated that as SA was born out of computer science rather than linguistics, there has been some debate among practitioners and even researchers over the distinction between sentiment and opinion, as well as whether the area should be named SA or opinion mining (OM). In terms of the usage of these two terms, SA is predominantly used in industry, but both OM and SA are widely used in academia. According to Pang and Lee (2008) and Liu (2015) the term "Opinion Mining" was first seen as a computational task in a work by Dave et al. (2003). He defined OM as a collection of written documents containing opinions about an object and the goal of OM is to extortaspects and components of the object that have been remarked on in each document as well as to identify whether the comments are positive, negative, or neutral. Esuli and Sebastiani (2006) defined OM as a new field at the intersection of information retrieval and computational linguistics that is concerned with the expression of an opinion rather than the content of a document. Pang and Lee (2008) stated that in some ways, the development of the term "Sentiment Analysis" is similar to that of "Opinion Mining." OM and SA are two terms that are frequently used interchangeably as sentiment and opinion are linked concepts, and when one is addressed, the other is mentioned as well (Hemmatian & Sohrabi, 2019; Ligthartet al., 2021l; Mountassir et al., 2013; Yousefpour et al., 2014). In such an instance, the researcher has decided to use ‘Sentiment Analysis’ throughout this dissertation due to its widespread adoption and prevalence in linguistic research.

2.2.2 Applications in computer science

Researchers from a variety of fields including natural language processing (NLP) and machine learning (ML) are interested in the area of SA in view of the way they are connected. When it comes to SA, a section that demonstrates the architectures in the domain of computer science will be pointed out too. This section of this literature review

will look at NLP and ML which will be briefly described here, along with an explanation of how each of these subjects relates to SA.

Pustejovsky and Stubbs (2012) explained that NLP is a branch of computer science and engineering that arose from the study of language and computational linguistics in the area of AI. The purpose of NLP is to create and construct applications that use natural language to assist human interaction with machines and other devices. Indurkhya and Damerau (2010) explained that NLP can be referred to as computational linguistics as it focuses on the comprehension and creation of human language in both writing and spoken that combines computer science with linguistic concepts and methodologies. Moving on to ML, Pustejovsky and Stubbs (2012) described it as a branch of computer science and AI that involves the construction and implementation of systems that improve over time as they collect more data. Ligthart et al. (2021) described that ML is a bottom-up approach that uses subsymbolic AI. They stated that some of the most essential topics to acquire in the domains of language technology and computational linguistics are a text's sentiment categorization, syntactic structure (sentence parsing) and Named Entity (NE) identification. According to him, it is the traditional "bag of words" method where in order for the algorithm to learn, each word is considered as a feature. Other types of common features are Term Frequency (TF), Part of Speech (POS), Negation and Syntactic dependency (Ligthart et al.,2021; Pang & Lee, 2008; Salah et al., 2019). Following the extraction of a set of features, an algorithm is used to learn those features and build a classifier based on the algorithm that learned from the features.

2.3 COVID-19 vaccination

On March 11, 2020, the World Health Organization (WHO) declared COVID-19 as a pandemic. Since then, experts all around the world have been working hard to develop a vaccine to end the crisis and let people resume their normal lives (Awijen et

al., 2022). The first round of vaccines was released at the end of 2020, with some advanced nations launching early immunization campaigns and others were still acquiring vaccines and waiting until early 2021 to get immunized. Several vaccinations had been produced and licensed in a relatively shorter period to combat the pandemic which is less than a year after the pandemic had been declared (Marcec & Likic 2022). Over twenty vaccines had been licensed in various regions of the globe by November 2021 (Mohamed et al., 2022). Over eight billion shots of eight front-runner vaccines had been provided to the world population at that time, with the overwhelming majority scheduled to be administered in 2021. Subsequently, over 4.4 billion individuals or roughly 56% of the world's population, had received one or more doses by December 2021 (Mallapaty et al., 2021).

A global vaccination drive was underway to put a stop to the outbreak, however the success of such a campaign was hugely reliant on people's desire to get vaccinated (Marcec & Likic, 2022; Syed Alwi et al., 2021). While vaccines proved to be crucial, there were many people who appeared to be hesitant to get vaccinated (Alabid & Katheeth, 2021; Allington et al., 2021; Amjad et al., 2021; Yang & Sornlertlamvanich, 2021). Patterson et al. 2022 stated that this is due to their skepticism about certain vaccine components and fear of severe side effects. Klimiuk et al. (2021) found out that there was a huge amount of information in the conspiracy theory and misleading sections regarding vaccination may imply that the writer of such remarks lack faith in the scientific achievements of medicines. Yang and Sornlertlamvanich (2021) stated that people in many countries were becoming more anxious about the vaccine's adverse effects and reliability as a result of local mortality caused by varied conditions and unknown causes. According to a study on the debate over the efficacy of COVID-19 vaccines, the majority of the public did not deny them as the study indicated that vaccination reluctance could be caused by factors outside of the vaccine. The public's understanding and attitude

concerning vaccines, as well as their primary information needs, are reflected in online media platforms (Gao et al., 2022).

In short, the next concern will be dealing with public opinion once safe and reliable vaccines become accessible. The public's approval for vaccination is critical to combat the pandemic. Respectively, it is critical to learn what the public thinks about vaccination. As of early 2021, various SA studies on the COVID-19 vaccination have been conducted. In addition, there are also supplementary studies that looked into public opinion of vaccination in Australia (Kwok et al., 2021), Indonesia (Wibowo & Musdholifah, 2021), Canada (Yan et al., 2021), Philippines (Villavicencio et al., 2021), Poland (Klimiuk et al., 2021), Iran (Nezhad & Deihimi, 2022), India (Barkur et al., 2020) and many more.

On the home front, Tan Sri Muhiyiddin Yaasin, the then-prime minister, was the first Malaysian to be immunized under the government's flagship vaccination program, the nationwide COVID-19 immunization campaign (PICK), with his first shot being televised live on different media channels on February 24, 2021.

PICK, which was executed in four phases, sought to vaccinate 80% of Malaysia's adult population in order to create herd immunity and prevent COVID-19 from spreading throughout the nation. With the first stage of the process began at the end of February 2021, the country's frontline workers, who had been feeling the strain of the pandemic's onslaught on a frequent basis for over a year, were chosen for vaccination. To guarantee that the immunization campaign would achieve its goal in a comprehensive way, walk-in and outreach programs were established, particularly in remote communities in the country's inland areas. Thanks to such initiatives, almost 97% of the adult population in Malaysia had received their vaccines by December 2021 (Radu, 2021).

On September 8, 2021, PICK for teenagers was initiated involving nearly three million teenagers between the ages of 12 and 17 in Malaysia. By October 2021, 80% of teenagers had been administered at least one dose of the Pfizer vaccine thus making it one of the world's quickest teenage COVID-19 immunization rollouts ("Malaysia's vaccination rollout," 2021). However, the government efforts to contain COVID-19 transmission in Malaysia did not stop here. A few months after the vaccination program for teenagers had been launched, National COVID-19 Immunization Program for Children (PICKids) commenced on February 3, 2022 for children aged five to twelve, according to Health Ministry Malaysia. As of March 25, 2022, over 36% of the Malaysian child population had been given the first shot of COVID-19 vaccine under the program. ("Malaysia to roll out," 2022; "Over 36% of children," 2022)

2.4 DNVA and past studies

SA is a burgeoning field at the crossroads of computational linguistics and computer science that aims to automatically detect sentiment in text. For the past decade, the studies of SA had been growing immensely since the rapid rise of social media. Additionally with the sudden appearance of the infections illness and the horror of being infected with COVID-19, an enormous amount of research has gone into developing and deploying COVID-19 vaccines. These SA research studies, to be presented in the next three paragraphs, look into a number of public opinions towards COVID-19 vaccine in Malaysia by utilizing automated computer-based approaches like Latent Dirichlet Allocation (LDA) (Feizollah et al., 2022; Ong et al., 2022**)** and corpus-based approaches, in particular NVivo, together with alternative analytical strategies such as thematic analysis (Wong et al., 2023).

In order to comprehend the debate surrounding halal vaccinations in social media Facebook and Twitter in Malaysia, Feizollah et al. (2022) utilized Latent Dirichlet

Allocation (LDA) to determine the theme and continued with SA by applying National Research Council (NRC) of Canada Emotion Lexicon to further explore the emotions in the data. The study testified that the sentiment about halal vaccination is largely neutral in Twitter platform but positive in Facebook platform while trust is the most prevalent emotion in both datasets, followed by anticipation and fear.

By employing the same method as Feizollah et al. (2022), Ong et al. (2022) employed LDA as well as SA to identify the common topics discussed in tweets related to the COVID-19 vaccine booster to explore Malaysians' perceptions of the booster. The study shows that Malaysians have a range of concern about the COVID-19 booster shot and the sentiments demonstrated that the public was more likely to have positive sentiments about the national immunization program than to have negative sentiments about the mixing of vaccine boosters.

Wong et al. (2023) aimed to uncover views that opposed COVID-19 vaccinations among Malaysians on Facebook by using NVivo and LDA. The study indicated that the skepticism and difficulties surrounding the COVID-19 vaccine have raised a number of significant concerns that might impede the vaccine acceptance. The lifelong side effects, safety, efficacy, duration of protection, the halal status of the vaccine, the preference for treatment over vaccination and more were established as the reasons for Malaysian to oppose vaccine in this research by using thematic analysis.

Moving forward, SA studies on the subject of vaccination have been carried out in nations including the United States, South Korea, Indonesia and China. The next seven paragraphs will present key studies in SA which focus on using computer-based automated methods and corpus-based approach for obtaining public perceptions on sentiment of vaccination. The methods include NLP (Sattar & Arifuzzaman, 2021), LDA (Melton et

al., 2021; Park & Suh, 2023; Saleh et al., 2023; Sutrave et al., 2021), SNA (Luo et al., 2021) and Corpus-Assisted Discourse Studies (CADS) (Nasyaya et al., 2023). Most of the data from these studies were gathered from social media platforms like Twitter (Park & Suh, 2023; Saleh et al., 2023; Sattar & Arifuzzaman, 2021; Sutrave et al., 2021), Reddit (Melton et al.,2021), YouTube (Nasyaya et al., 2023) and Weibo (Luo et al., 2021).

Sutrave et al. (2021) aimed to detect the emotions and themes of discussion on Twitter around the COVID-19 vaccination by using LDA, a topic modelling approach. By using ML, the researcher utilized NRC Sentiment Lexicon in order to categorize the tweets into eight different emotion groups and also categorize them into positive and negative. The data indicate that the public has a favorable opinion of the vaccine in general. Those who have positive opinions about the COVID-19 vaccination applaud the efforts of the government, medical professionals, and pharmaceutical corporations in creating it. Nevertheless, there are others who have negative opinions concerning vaccines due to skepticism of the authorities and fears about vaccine efficacy and severe effects.

Park and Suh (2023) conducted a comprehensive analysis of Korean Twitter discourse surrounding COVID-19 vaccines, specifically public sentiments and topics related to the brand of vaccines following their rollout. They analysed their research in a manner similar to Sutrave et al. (2021), who utilized LDA alongside the programming language Python. The research not only sought to explore the changes of sentiments over the time but also delved into the specific themes and sentiments associated with adverse events (AEs) related to vaccination. The data resulted in the analysis of public sentiment, especially regarding the Pfizer vaccine, which showed a prevailing negative sentiment

from the early stages of vaccination. This sheds light on the diverse public perceptions and concerns related to specific vaccine brands.

Melton et al. (2021) gathered vaccination-related data from the Reddit social media network to evaluate public reaction against the COVID-19 vaccine by using LDA. This study used the same approaches of Sutrave et al. (2021) and Park and Suh (2023); however, there is only a slight difference in how the tools were employed in collecting and analyzing data. Whileboth studies preprocessed the data using Python, Sutrave et al. (2021) used NRC for categorizing the tweets in Twitter and Melton et al. (2021) used TextBlob for the purposeof classifying the debate in Reddit platform. The results of this research showed that the sentiment in these social media forums were generally more positive than negative and did not change much since December 2020. In spite of that, during the LDA topic modelling, terms suggesting vaccine hesitancy were identified.

Saleh et al. (2023) took a different approach compared to Sutrave et al. (2021), Park and Suh (2023), and Melton et al. (2021) by using a unique method for topic modeling. They employed the Correlation Explanation (CorEx) ML algorithm to identify clusters of topics in tweets related to COVID-19 vaccines. Their findings shed light on the changes in public attitudes over time, indicating that several topics exhibited a monthly trend towards more positive sentiment. Interestingly, tweets that initially had strongly negative sentiments when comparing COVID-19 to the influenza vaccine showed improvement over time. Despite an overall positive shift in public perception, Saleh et al. (2023) pointed out concerning trends, especially within specific topic and demographic clusters, suggesting potential areas of COVID-19 vaccine hesitancy.

Sattar and Arifuzzaman (2021) used NLP approaches to extract information on public awareness of COVID-19 vaccine and the consequences of vaccination in terms of

health precautions. The studies also look at the public's reactions to COVID-19 safety precautions after they had received the shots. In terms of maintaining safety precautions against COVID-19 among the vaccinated population, positive opinions outweigh negative emotions as people were cooperative in keeping social distance while simultaneously being optimistic about maintaining hygiene.

To investigate public opinions surrounding the COVID-19 vaccination, Luo et al. (2021) focused on the cultural sensitivity aspect and used the SNA software Gephi to explore the semantic networks. Results show that domestic vaccination policy, priority groups, challenges from COVID-19 variations, and the worldwide pandemic scenario were all topics that the social media users discussed. Luo et al. (2021) stated that the research only caught a glimpse of public opinions on the COVID-19 vaccination during its early stages. As a result, additional efforts should be made in the future to analyze how perceptions change over time, such as probable shifts in public opinion following major events. In addition, although vaccination is a frequently debated topic on social media, it is unclear if the insights gleaned from social media users' sentiments can be applied to the entire community. The researcher suggested that future studies should retrieve corpus from many sources to increase external validity.

Nasyaya et al. (2023) delves into the perspectives of the Indonesian population concerning COVID-19 vaccination by using YouTube comments as the source of the data. Focusing on the initial phase of vaccination in Indonesia, the research employed a mixed-methods approach, combining Corpus Linguistic (CL) methods with Critical Discourse Analysis (CDA). The authors used AntConc software to analyze the pattern and grammar and the results showed a range of public opinions that are divided into fifteen categories based on the comments left on YouTube videos. The study discovered that positive opinions are influenced by trusting accurate information, the impact of

community leaders as influencers, transparent data, and effective communication from health authorities. On the flip side, ignoring misinformation can result in negative opinions, indifference, and even public rejection of the COVID-19 vaccine.

In addition to SA, researchers have used additional analytical techniques to investigate the topic of vaccination in Malaysia, namely SPSS (Syed Alwi et al., 2021), Mann-Whitney and Kruskal-Wallis Test (Mohamed et al., 2021) and Framing Analysis (Ghazali et al., 2020). Not only that, thematic analysis and AntConc (Nor & Zulcafli, 2020) were utilized to look into the COVID-19 issue in Malaysia as shown in the following four paragraphs.

The study by Syed Alwi et al. (2021) was about determining Malaysians' fears and acceptance level of the COVID-19 vaccine. By using a snowball sampling strategy, the researchers conducted an online questionnaire to 1,411 Malaysians aged 18 years and above recruited on social media platforms. The poll included 14 questions about sociodemographic traits, medical illnesses, sources of COVID-19 info, approval of the COVID-19 vaccine, and doubts about the COVID-19 vaccine. With the help of SPSS, a descriptive analysis was carried out in this study and the findings show that Malaysians have a high level of acceptance for the COVID-19 vaccine. As a result, the researcher indicated that the Malaysian government and other relevant organization should ramp up their campaign and get ready to roll out the COVID-19 mass immunization program to the public. Despite the high acceptance rate, it is still necessary to address concerns among the vaccine sceptics by establishing faith in the vaccine's safety and effectiveness through proper vaccine information.

Mohamed et al. (2021) chose to investigate Malaysian adults' understanding, approval, and perceptions of the COVID-19 vaccine and an online survey was carried out

over two weeks in December 2020. The questionnaire was divided into four components including demographic and COVID-19 status, COVID-19 vaccination awareness, COVID-19 vaccine approval and perspective based on the Health Belief Model (HBM). This research approach was the same as Syed Alwi et al. (2021) as both gathered data from social media by using online questionnaires. The only variation was in how the data was analyzed. In this study, the Mann-Whitney test, Kruskal-Wallis test and SPSS were used to analyze the data while in Syed Alwi et al. (2021) a descriptive analysis was used in the study with the help of SPSS. This research resulted in more than half of those surveyed having little understanding about the COVID-19 vaccine and a majority of them were willing to receive one. The study also found out that statistically and significantly higher scores were linked to a higher educational qualification, a higher income group, and living with someone who is more likely to have severe COVID-19. The research found out that the public were more inclined to get vaccinated if they were in a lower age group, have higher education levels and were female.

Ghazali et al. (2020) examined vaccine coverage in two Malay newspapers in an attempt to determine whether the news media properly addressed the vaccination issues. By using framing analysis, all news was examined to determine the quantity of coverage, frame kinds, and sources utilized in reporting on vaccination concerns. The data revealed some disparities in the numbers and methods in which the two newspapers covered about immunization and discovered that the media place a low priority on the topic which is often regarded as news to be reported, rather than an issue to be highlighted and stressed.

Nor and Zulcafli (2020) studied themes from headlines about COVID-19 in the news and terms that collocate with the node COVID-19 by using AntConc. This study described how a corpus-driven method discovered new information about COVID-19 where the most common verb and noun collocations indicated what Malaysian

government was doing, how Malaysians reacted to the virus and the latest COVID-19 news. Based on the collocations, this study found out that there are feelings of anxiety and menace from the disease.

The following two paragraphs describe studies on the issue of vaccinations by adopting Discursive News Value Analysis (DNVA), despite the fact that thereare very few SA-related articles in this area.

Fatah et al. (2022) conducted a qualitative approach to investigate the connection between media coverage and public concern by addressing the religion as a noteworthy component for media reporting on COVID-19 vaccinations in Malaysia. By using DNVA on 10 news headlines from Malaysian newspaper and conducting a focus group discussion, the findings showed that religion is a newsworthy component of COVID-19 vaccine reporting in Malaysia and it showed that religion is a powerful factor that can alter public attitude on vaccines.

Chen and Liu (2023) looked into the disparities in how the COVID-19 vaccine is portrayed in Chinese and US foreign media sources. This study uses DNVA to look at the importance of keywords, nomination techniques, and images in news stories. It reported that the COVID-19 vaccine was presented as a foreign aid in order to improve diplomatic relations and promote international ties in Chinese media sources, yet it was emphasized as a local commercial product in the US media sources. The different methods that each country presents the COVID-19 vaccination have aided in the global acceptability of vaccines from both countries.

2.5 Literature gap

Although SA has been used in various research to examine people's opinion of vaccination, the past papers have only analyzed the polarity of the sentiments and none

from a linguistic viewpoint. In terms of research about COVID-19 vaccination in the Malaysian setting, these works looked into vaccination acceptance in Malaysia's overall population by using online questionnaires (Syed Alwi et al., 2021 & Mohamed et al., 2021) and relationship between the news media and coverage with vaccination topics (Fatah et al., 2022 & Ghazali et al., 2020). Furthermore, although looking extensively online, the researcher was unable to locate published research that studied COVID-19 vaccination from a linguistic point of view in the context of Malaysia but instead found one in the context of Indonesia (Nasyaya et al., 2023). The researcher did find one (Nor & Zulcafli, 2020), which is the closest and related to the study; however, the study focused on the discourse around COVID-19 news rather than the sentiment of COVID-19 vaccines. On top of everything, NLP and ML were mostly used to analyze the results of the public's sentiments. As of now, there is clearly a lack of research work in Malaysia that focuses on sentiments of vaccination from a linguistic or discursive viewpoint. Given the importance of identifying significant vaccine concerns, such information is required and crucial to assist the authorities in taking preemptive measures to increase public confidence in vaccination. Hence, this study intends to investigate the linguistic elements that are relevant to describe sentiments related to vaccination in Malaysia, set against the background of the rising threat of COVID-19 and how vaccination could potentially mitigate its adverse effects.

Secondly, a small number of studies (Sattar & Arifuzzaman, 2021 & Luo et al., 2021) stated that there is a lack of data throughout the course of a longer period to see how opinions on vaccination shift over the time as both researches only captured a glimpse of public opinion on the COVID-19 vaccination during its early stages. As the COVID-19 vaccination opinions may change over time, it is necessary to examine them in a prolonged length of time. By taking this into consideration, the researcher has decided

to investigate the transition of the sentiment toward COVID-19 vaccination in 3 different periods in Malaysia so as to bridge the gap.

In addition, while most SA research on COVID-19 vaccination relies on information gathered from social media platforms like Twitter and Facebook (Feizollah et al., 2022; Klimiuk et al., 2021; Liu & Liu, 2021; Nezhad & Deihimi, 2022; Ong et al., 2022; Park & Suh, 2023; Saleh et al., 2023; Sattar & Arifuzzaman, 2021; Sutrave et al., 2021; Wong et al., 2023; Yousefinaghani et al., 2021), Bučar et al. 2018 stated that the SA community has paid considerably less attention to the news genre. The researcher was able to find some related studies in the news genre (Fatah et al., 2022, Ghazali et al., 2020 & Nor & Zulcafli, 2020) however only one study did look at the vaccination context (Chen & Liu, 2023). Luo et al. (2021) also stated that as vaccination is frequently a debated topic on social media, it is unclear if the insights gleaned from social media users' sentiments can be applied to the entire community thus corpus from other sources are needed to increase external validity.

This study is likely to provide insights into how opinions on vaccination in newspapers have changed over time, as no previous researchers have studied this subject within this domain. In that event, the researcher has determined to focus on the domain of newswire that is centered around vaccination in Malaysia on the ground that none of the past scholars had ever worked on a topic of vaccination in a newspaper in a linguistics context. It is hoped that findings from this study may encourage the use of linguistic components to improve SA research and the development of more sentiment systems for real-world applications in order to gather a broad sense of public opinions on specific issues, given the importance of SA. To sum up, the aim of this research is not to establish this method as the only ways to analyze sentiment via a linguistic viewpoints; rather, it is

to inspire further discussion and better approaches to explore sentiment. As a result, this research may pave the way for future linguistics research in the field of SA.

2.6 Summary

This chapter reviews previous research on three related fields. They are Sentiment Analysis, Sentiment Analysis in the domain of Computer Science and COVID-19 vaccination. As the major theme of this study is Sentiment Analysis, the related topic is examined further by looking at previous studies on sentiment concerns, specifically COVID-19 immunization in Malaysia. This discussion also points out a gap in earlier studies that the results from the current study could fill, particularly by concentrating on a newspaper domain from a linguistic point of view in the context of Malaysia and on how anti-vaccination sentiments have evolved over time. Therefore, this study has great potential in providing a thorough understanding of aforementioned problems and contributing to the areas indicated above.

CHAPTER 3: DATA & METHODOLOGY

3.1 Introduction

This chapter contains in-depth descriptions of the research methods used in data analysis. Data collection is discussed in terms of the data size, procedure and duration of the data collection. The analytical framework and approach are explored in detail under subsequent subheadings. For corpus-based approach, two tools are adopted in the present study, namely NVivo and Wmatrix. Frequency lists are generated usingboth of the tools. While Sentiment Analysis is performed using NVivo, UCREL Semantic Analysis System (USAS) is performed using Wmatrix. Following that, two news values from DNVA are chosen to be the focal points in the sentiment analysis to identify the linguistic elements based on NVivo data. Next, the data analysis procedures explain the steps taken to generate findings and lastly, the chapter concludes with a brief explanation of how ethical considerations have been addressed in this study.

3.2 Data selection

A selection of articles from The Star newspaper make up the data sample. The Star is a Malaysian tabloid newspaper published in English. The three guiding principles of press independence, political balance, and best tradition were established when The Star was founded in 1970. According to data from Survey Research Malaysia, The Star has grown by an astounding 68% since the publication's inception. Moving forward, The Star launched The Star Online, Malaysia's first news website, as its first electronic gateway in 1995 and also the first national newspaper to offer an online version, featuring a wide variety of material, from local and national news to lifestyle articles and community tales. By the decade's close, The Star went global where it joined the Asia News Network (“Our Story So Far,” 2020). According to the Audit Bureau of

Circulations, it has the largest circulation in Malaysia hence it was chosen as the data source for this study because of its standing among Malaysians ("The Star is top," 2016).

For this research, a newspaper platform was chosen as it plays a significant role in shaping the readers' viewpoints and broad concepts. Not only that, they benefit from this online news channel which gives them real-time news about the latest occurrences. Over and above, the newspaper platform is the most cost-effective way for individuals to learn about vaccination (Padmaja et al., 2014). The Star Online, the web version of the most influential English daily in Malaysia, has more than 25 million visits on average each month (Sivanandam, 2020). In general, The Star Online is a top internet source for news in Malaysia. On that account, The Star has been identified as the sole source of data for this study. Plus, news articles are well-organized and concisely presented and through this form of media, one can quickly gain useful information as people recognize that the media serves as a bridge between them and the government, hence its immense influence on their thoughts.

The data were gathered for a whole month of March for every year from 2020 to 2022 (see Figure 3.1 below), namely from the time before the COVID-19 vaccination program in Malaysia began, when the vaccination first started and one year after the vaccination had begun. March 2020 marked the first outbreak of the pandemic in Malaysia and the development of a vaccine was still ongoing at that time. In March 2021, the Malaysian government started the flagship vaccination program, namely the nationwide COVID-19 immunization campaign (PICK) in stages, after the vaccines became available and vaccination efforts had begun in other parts of the world. Finally for March 2022, it was one year after the national vaccination program had commenced, hence it is an appropriate time for a review of the people's perceptions toward the vaccination program. One of the objectives of the study is to explore the changes of the

sentiments over time, which is why a total of 3 different months from 3 distinct years of 2020 to 2022 were selected as the timeframe from which the data are collected.

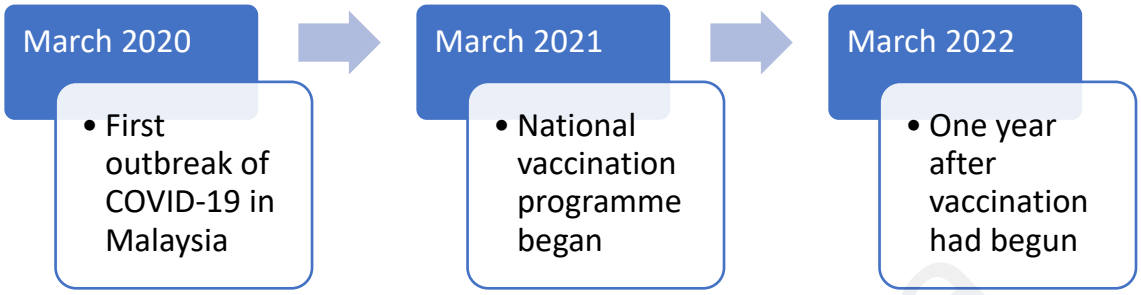

Figure 3.1: The timeline of data gathered

Through the u-Pustaka portal which is a national digital library that provides free access to digital content to all Malaysians, the data were collected via PressReader which is a platform that offers e-newspapers and e-magazines in 60 languages from 100 nations across the world. This source of data is simple to use because it is free and offers extensive news archives from the past to the present. To ensure that only news reports pertaining to the keyword were retrieved, "COVID-19", "vaccine" and "vaccination" were employed as the search terms on The Star Online portal, accessed on the PressReader app via u-Pustaka online databases. All of the related news reports were collected and Microsoft Excel was used to prepare the data for analysis. The collected news articles were screened and manually selected in order to avoid duplicated and irrelevant materials such as news articles in a sport or entertainment section. Besides that, the unnecessary clutter, such as marketing commercials, social network information and more were removed from the processed data during this stage. By doing so, it allowed the user to focus on only the most related and important aspects of the news articles, like the news title, date of publication, the name of the author and the body of the article. After all the processes, the selected news articles were then converted to plain text (.txt) format to ensure that NVivo and Wmatrix could read them.

Ultimately, a total of 5 articles were collected for the whole month of March in 2020, 146 articles in 2021 and 105 articles in 2022 (see Figure 3.2 below). The lack of articles in March 2020 is understandable as a result of limited news of vaccines in Malaysia. Firstly, Malaysia was not a vaccine producer and secondly, most of the vaccines were still in the early stage of testing and development. Additionally, two months before and after March 2020 were surveyed and the number of news articles relevant to COVID-19 vaccination did not differ much from March 2020. Hence, March 2020 was retained as the first period during which data would be collected. Altogether, a sum of 256 news stories were gathered for a whole month of March for every year from 2020 to 2022 as a corpus for this research.

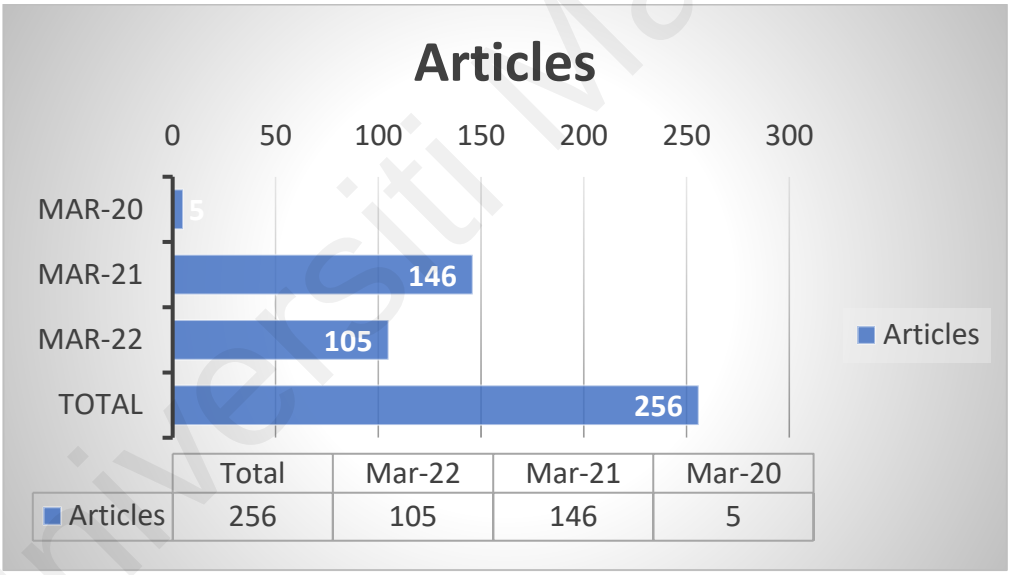

Figure 3.2: News articles collected from 2020 to 2022

3.3 Research design

This research adopts a mixed method study of corpus-based approach whereby both quantitative and qualitative data would be collected and analyzed. Nevertheless, it is largely a qualitative study focusing on linguistic elements of which sentiments are constructed in news reports, based by frequency counts in corpus software which provides the quantitative perspective. With this approach towards data collection and analysis, the

sentiments can be examined more closely through the qualitative lens while the quantitative tool helps to generate an overview of the extent of such sentiments. Such setup combines the strengths of both approaches while reducing the drawbacks of each (Creswell & Creswell, 2017).

3.3.1 Corpus-based approach

Corpus linguistics (CL) is a rapidly growing branch of linguistics that comprises the computer-based assessment of very large quantities of digital form texts. The term corpus comes from the Latin word corpus, which means "body," hence a corpus is a collection of writings (Baker, 2010). CL is much more than a systematic procedure; it is a field in which technical progress and theoretical growth coexist (Hunston, 2011). Plus, it has the ability to uncover patterns that would otherwise go unnoticed by human analysts via rapid and accurate procedures (Baker, 2010; McEnery & Hardie, 2011). A study in CL requires quantitative methods to obtain results that are both descriptive and explanatory. The descriptive analyses are derived via computationally looking for specific language traits in a corpus. Corpus analysis also provides linguistic study with a high level of dependability and validity by utilizing enormous volumes of natural sources data (Thornbury, 2010).

Gries (2012) stated that there are some researchers who regard CL as a theory, while others regard it as a methodology. According to McEnery and Hardie (2011) a scholar who regards CL as a methodology would adopt a corpus-based approach (CBA) to investigate a theory or hypothesis, usually one that has been developed in the literature, to justify, disprove, or refine it. According to Tognini-Bonelli (2001), a scholar who regards CL as a theory would collect data from the corpus and discover language phenomena without making prior assumptions or expectations. In short, theory does not pre-exist in this method; rather, it emerges from the corpus (Biber, 2012; Meyer, 2014).

That said, corpus-driven approach (CDA) is viewed as a theory in CL. In a nutshell. a contrast has been noted between corpus-driven approach (CDA) which is considered as a theory and corpus-based approach (CBA) which is viewed as a methodology in CL (Baker, 2010; McEnery & Hardie, 2011; Biber, 2012; Meyer, 2014).

Many scholars, on the other hand, concur that it is an approach rather than a theory that can be used in a variety of language studies thus the term "corpus-based" applies to all corpus linguistics (McEnery et al., 2006; McEnery & Hardie, 2011; Meyer, 2014). In that event, the researcher regards CL as a methodology and this study will be using a corpus-based approach (CBA) to analysing the data because it is in line with the research objectives and process of this study.

3.3.2 Discursive News Value Analysis (DNVA)

DNVA aims to investigate how semiotic resources are used to construct news values and enhance the existing literature on news sharing with a particular linguistic focus. Bednarek and Caple (2014) mentioned that the method of analyzing news values in a text allows us to examine how an event is "sold" to us as news(worthy) and how the audience is made to believe something is newsworthy using language, imagery, layout, typography, and other means. To reiterate, we see news values in terms of what makes actors, events, and topics "newsworthy," and we are not concerned with broad elements that influence the choice of news. News values are viewed as values that exist in and are formed through discourse, and the key study focus is in how texts build newsworthiness. Linguists can use DNVA to examine how news values are created throughout discourse in news texts. Therefore, it must be taken into account in any linguistic study of news texts (Bednarek, 2016).

In this research, DNVA was adopted as the theoretical background to examine the linguistics characteristics of news report that represent sentiments towards COVID-19 vaccination in Malaysia. With a view to serve the goals of the study, Table 3.1 below outlines the Bednarek and Caple's (2017) framework of DNVA which is divided into a number of news values and followed by the definition of the news values.

Table 3.1: News values and definition

Consonance	The event is discursively constructed as (stereo) typical (limited here to news actors, social groups, organizations, or countries/nations)
Eliteness	The event is discursively constructed as of high status or fame (including but not limited to the people, countries or institutions involved)
Impact	The event is discursively constructed as having significant effects or consequences (not necessarily limited to impact on the target audience)
Negativity	The event is discursively constructed as negative, for example as disaster, conflict, controversy, criminal act
Personalization	The event is discursively constructed as having a personal or 'human' face (involving non-elite actors, including eyewitnesses)
Positivity	The event is discursively constructed as positive, for example as a scientific breakthrough or heroic act
Proximity	The event is discursively constructed as geographically or culturally near (in relation to the publication location/target audience)
Superlativeness	The event is discursively constructed as being of high intensity or large scope/scale
Timeliness	The event is discursively constructed as timely in relation to the publication date: as new, recent, ongoing, about to happen or otherwise relevant to the immediate situation/time (current or seasonal)
Unexpectedness	The event is discursively constructed as unexpected, for example as unusual, strange, race

Additionally, in order to focus on and broaden the investigation of the linguistic traits uncovered from this corpus to represent sentiments toward COVID-19 vaccination in Malaysia, the analytical framework based on the linguistic devices by Bednarek and Caple (2017) was adopted to provide the context for this study which is shown in the Table 3.2 below.

Table 3.2: Coding manual adopted for linguistic analysis

News Value	Linguistic Device	Example
Positivity/Negativity	References to negative/positive emotion and attitude	**concerns** about even remote chances of Ebola exposure; **fury** as primary head takes week off in term to fly to Caribbean; a move that has **outraged** local politicians; amid signs of **panic**; 'First hydrogen bomb test' **condemned**;
	Negative/positive evaluative language	Corbyn's **shambolic reshuffle**; a **violent thug** who had no interest in Islam; **shoddy** financial advice; the **brilliant** astrophysicist;
	Negative/positive lexis	Boy, 8, one of 3 **killed** in bombings at Boston Marathon; scores **wounded**; Western black rhino declared **extinct**; Flint residents **protest** high bills for 'poison' water; 13 migrants **drown** as boat capsizes off Malaysia; Nigeria has been declared officially **free of Ebola**;
	Descriptions of negative (e.g. norm-breaking) or positive behaviour	Hospitals **don't** have enough beds, and there **aren't** enough ambulances; Treasurer Joe Hockey **has broken his promise** to balance the budget by 2019; [Canadian Prime Minister] Trudeau, who last year **unveiled a cabinet with an equal number of men and women "because it's 2015**

In this research, two news values from Table 3.1 which are related to the objective of this study have been chosen for analysis. In other words, the news value of positivity and negativity from the Bednarek and Caple's (2017) framework of DNVA will guide the analyses and discussions. In order to analyze the news value, Table 3.2 showed that there are linguistic devices for determining each of the news values and there are a total of 4 linguistic devices for the news value of positivity and negativity. The four linguistic devices involved are emotion and attitude, evaluative language, lexis and behavior.

3.4 Data analysis tools

A corpus alone would not be sufficient for analysis when it comes to linguistic research. As a result, corpora are frequently employed in combination with analytic tools that can do linguistic features like tallying, categorization, and visualization where the outcome of the results needs to be generated by researchers. This study uses CBA which

is a technique that primarily involves readings of corpus to generate ideas, concepts, or a pattern at the hand of the researcher's interpretations of the corpus. NVivo and the Wmatrix tools are used to analyze the linguistic elements which describe the sentiment of the corpus of an online news report on COVID-19 vaccination. Both of the corpus-based tools are useful in providing detailed information about the texts and extracting features of linguistic elements one by one in the texts. Not only that, data interpretation was inductively conducted based on an analysis of the occurrences the data.

3.4.1 NVivo

NVivo is a program for analyzing data on a computer also recognized as computer-based qualitative data analysis software. NVivo is extremely useful in automating the research process by arranging, handling, converting, and extracting both raw and interpreted sources of data (e.g. text, images, and audiovisuals) in a structured, organized, and sequential manner (Wong, 2008). In addition, the software eliminates a large number of manual procedures, giving the user sufficient time to analyze trends, discover themes, and draw conclusions. NVivo's finding tools enable the user to go down into the data at a specific level. As a result, the accuracy of the analysis process can be improved by verifying some of the user's observations of the data (Welsh, 2002). In the data analysis chapter, NVivo results will be presented and discussed. Furthermore, the software is utilized to identify vaccination-related language usage by visualizing data frequency list and the auto code features for Sentiment Analysis.

A word cloud is a word count search that helps one to locate widely used terms or expressions by generating a word cloud. All of these visual representations of data are a good tool for learning about what's happening with the data as well as a valuable end product. According to the online manual of NVivo, the tool can also detect the frequency of words and in the latest versions of NVivo, which is NVivo 11, the “Automated

Insights" function that has been added allows researchers to perform auto-coding for Sentiment Analysis throughout various datasets. Moreover, a grading system is used throughout the procedure which consists a total of four polarities (Very Negative, Negative, Positive and Very Positive). Each word that consists of sentiment has a predetermined score and those within the neutral range will not be coded. In cases where there are two different sentiments exist in a sentence, it will become a mixed sentiment as the software will tag both the positive and negative polarities because the words are examined individually throughout the analytic process. Moreover, the sentiment coding is stored alongside the nodes and relationships in NVivo and the sentiments work in a similar method to traditional codes in a way that they can be manually analyzed by the selected text.

3.4.2 Wmatrix

Wmatrix is a program used in a web-based corpus processing environment to examine the relevant linguistic aspects of the corpus data. Wmatrix has a number of useful corpus capabilities, including grammatical category tagging, also known as part-of-speech tagging (CLAWS), a semantic domain tagging, also known as The UCREL semantic analysis system (USAS), a lemmatizer (LEMMINGS) and the ability to construct frequency lists, compare them statistically and calculate concordances which is designed for quantitative text analysis (Rayson, 2009). The mixture of these features enables a macroscopic analysis where it explores the aspects of entire texts or types of language to a microscopic level where it provides insights into the use of a specific language trait and as a result, linguistic elements that should be examined further are suggested (Rayson, 2008). Wmatrix works on a variety of web browsers and it can run on all major operating systems too. In a nutshell, it is available to everyone that has a good Internet connection and a computer.

This semantic domains (USAS) is one of the features included in the Wmatrix application where the main semantic analysis tool (SEMTAG) assigns semantic tags describing the general context range of keywords from a lexicon of individual terms and a collection of multi-word groupings (Rayson, 2003). The lexicon comprises almost 37,000 words, with over 16,000 multi-word entities in the layout list. Furthermore, there are 21 primary semantic domains and more than 232 subsidiary categories identified. The major semantic categories are denoted by letters, whereas subcategories of the categories are denoted by numbers. (Rayson et al., 2004).

One of the major semantic categories, USAS Category E: Emotion is definitely one of the semantic domains that is important for the researcher to delve into as emotions are significantly linked to ideas, feelings, behavioral reactions, and a level of delight or discomfort. They also assist us in creating and navigating relationships, making crucial life decisions, and determining how we react to situations. Thus, it is meaningful to work on this domain in this corpus as it is also aligned with the objective of this research which is to find out the sentiment towards vaccines in Malaysia. Plus, this Emotion category in the semantic domains has the positive and negative polarity features in Wmatrix and it helps significantly in deducing the sentiment with the help of these features. Figure 3.5 below shows the 21 primary semantic domains in Wmatrix.

A General and Abstract terms	B The body and the individual	C Arts and crafts	E Emotion	F Food and farming
G Government and public	H Architecture, Housing and the home	I Money and commerce in industry	K Entertainment, Sports and games	L Life and living things
M Movement, location, travel and transport	N Numbers and measurement	O Substances, Materials, objects and equipment	P Education	Q Language and communication
S Social actions, states and processes	T Time	W World and environment	X Psychological actions, states and processes	Y Science and technology
Z Names and grammar				

Figure 3.3: Main semantic domains in USAS

3.5 Analytical procedures

The following figure shows how the process of research questions is operationalized and the flow for corpus analysis, from applying both corpus tools to interpreting the results and classifying the themes for the previous three different periods.

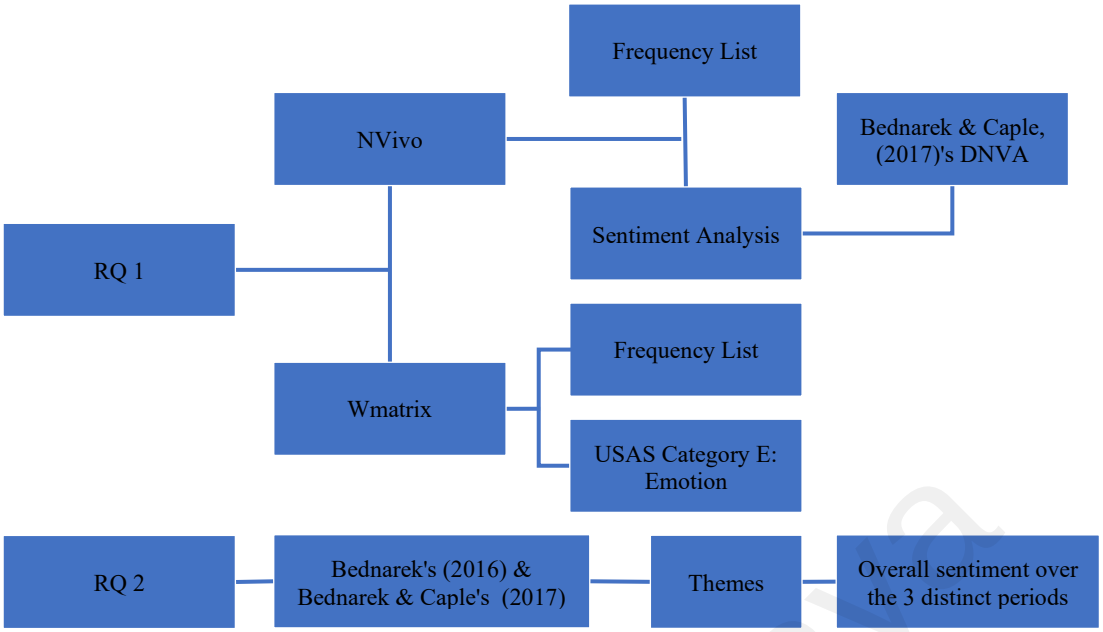

Figure 3.4: Flowchart of Data Analytical Procedures

In this study, the language aspects identified using corpus-based tools are more directly relevant to the current investigation. The research began by analyzing the corpus that were obtained as stated in 3.2. To answer Research Question 1, both corpus tools of NVivo & Wmatrix were used to generate the Frequency List to get the overview of the significant issues. Not only that, Wmatrix which has the features of the keyness function in USAS was used to figure out which groups of words are peculiar to the text. To shed light on how a particular linguistic characteristic is used, Wmatrix was used to derive the semantic domain distribution. The classification of semantic domains of category E: Emotion generated a list which were analyzed qualitatively to identify the linguistic elements or terms that act as indicators or used to express vaccination-related sentiment. In order to determine emergent issues regarding the vaccine, both of the interpretations from Frequency List(NVivo) and USAS categories obtained in keyness function (Wmatrix) were utilized.

Next, the auto code features for Sentiment Analysis which has a total of 4 polarities (Very Negative, Negative, Positive and Very Positive) were employed to

identify the linguistic elements of the corpus. The results retrieved from the Sentiment Analysis of NVivo were then analyzed by the researcher, who manually coded them into the category of the linguistic devices shown in Figure 3.4 in order to avoid irrelevant sentences such as sentences discussing other than COVID-19 vaccination related.

In other words, to better grasp the linguistic elements that represent the sentiments of the reported voices, SA were generated and extracted from NVivo and the linguistic elements related to positive and negative news value and their linguistic devices were manually coded. Linguistic devices help to express, improve, or are "connected to" specific news values. It often refers to a device that "cite" or "invoke" news values which 'improve' the news value by discussing specific examples.

In the interest of answering Research Question 2, the definition of the news value of positivity and negativity explained in the framework of Bednarek and Caple's (2017) and Bednarek's (2016) were used as a reference to group the issues and sentiments into themes over the past three years. Table 3.3 below lists the themes that were used under each framework.

Table 3.3: Positivity and Negativity News Value in both frameworks

Framework	Bednarek and Caple's (2017)		Bednarek's (2016)	
News Value	Positivity	Negativity	Positivity	Negativity
Themes	Scientific breakthrough	Disaster	Benefits	Damaging & destroying
	Heroic act	Conflict	Helping	Violence & anger
		Controversy	Good evaluation	Crime
		Criminal act		Death

The conceptualization for each news value is provided in the Table 3.3 which is consistent with their own discursive approach. Using the news value in Table 3.3, one can determine certain events, however depending on the various contexts in which the data is being used some situations might be left out and vary from the table. According to Bednarek and Caple (2017), their fundamental point is to avoid the notion that news values are either 'inherent' in situations or merely reflected or incorporated in language itself. Instead, they emphasize the importance of discourse in the formation and reinforcement of news values. Accordingly, this research used both frameworks outlined in Table 3.3 as points of reference to contextualize and exemplify the breadth and depth of this study. By analyzing the compiled news articles, linguistic resources were inductively discovered. By creating and using the researcher's own conceptualizations, not only does it help to link the observed relationship between variables, illustrate how these elements fit together to provide coherent results, it also highlights and brings out several new themes that the frameworks fail to account for.

In order to determine emergent issues regarding the sentiments, both of the interpretations from Sentiment Analysis (NVivo) and USAS (Wmatrix) categories obtained were compared to tease out significant features or themes. The key issues with regard to the sentiments of COVID-19 vaccines and concerns which underlie the people's opinions were scrutinized and discussed. By the end of the analysis, the issues are manifested and the people's opinions for each of the issues appeared. In sum, both NVivo and Wmatrix were utilized as the corpus tools to address Research Question 1 and the framework of DNVA was utilized for both Research Question 1 & 2. The linguistic features from Research Question 1 led in addressing Research Question 2 by comparing the data throughout three unique time periods. As this study takes on a corpus-based approach, the categorization scheme was not built on the basis of a theoretical framework. For this study, inductive coding was applied to describe the process of analyzing data.

3.6 Ethical considerations

Thanks to the development of information and communication technologies, the Internet has become an increasingly promising medium and research site for scholars. In this study, the data are taken from the newspaper portal that is The Star. According to Fiesler and Proferes (2018), while internet users anticipate privacy for their public data, they willingly and regularly place it in the public internet domain, hence it is stated that online data is open to all. Not only that, in huge databases that produce a useful digital representation of human reality that academics require for their studies, informed consent cannot be acquired. Nonetheless, methods of preserving informant data, particularly through collaboration and communication are critical to managing informant damage reduction (Nycyk, 2022). With that being said, online news portals, as open access resources, are meant for public use, therefore, no specific consent or authorization is needed.

3.7 Summary

This chapter includes details on the data selected, research design, corpus linguistics software employed, the framework utilized and the procedures for data analysis. Both research questions will be addressed using the corpus analysis tools, which are Wmatrix and NVivo, respectively. While Sentiment Analysis is generated using NVivo and UCREL Semantic Analysis System (USAS) in Wmatrix, Frequency List is done using both of the tools. The linguistic elements in the sentiment-related analysis lines were identified by two news values that are adopted in the analytical framework based on the linguistic devices by Bednarek and Caple (2017). To determine the major key themes of the study for the previous three years, the analysis of the news value will be further examined. This section also goes into great depth on how the data are analyzed in order to respond to the research questions. Last but not least, the final section in this chapter

also clarifies any ethical concerns. The analyses and discussions ofthe findings will be presented in the following chapter.

CHAPTER 4: FINDINGS

4.1 Introduction

This chapter presents the findings of the NVivo and Wmatrix analysis. As noted in Chapter 3, two corpus analysis tools were adopted in the analysis of the first and second research questions, which were NVivo and Wmatrix respectively. In addition to it, DNVA was applied as the theoretical background to set the backdrop of this research. For the first research question, Frequency List, USAS features and Sentiment Analysis were performed to generate the overall recurring words that were mentioned on the samples to discover the prominent categories.

In the direction of answering Research Question 1, the framework of DNVA is adapted to analyze linguistic elements used in this corpus. In exchange for that, it requires an analytical framework that can be used to news discourse in order to analyze how news values are developed across the corpus. The researcher has adopted Bednarek and Caple's (2017) analytical framework to contextualize this study, aiming to investigate linguistic traits in a Malaysian COVID-19 vaccination sentiment corpus, discussed in section 4.3.

Figure 4.1 is the flowchart of the process in analysis of Research Question 1. There are four different sections: Frequency List, Sentiment Analysis, USAS and Linguistic Devices. All of these sections comprise subsections that presents data from three differentperiods, namely March 2020, March 2021 and March 2022. Hence, analysis from Section to Section 4.3 correspond with Research Question 1.

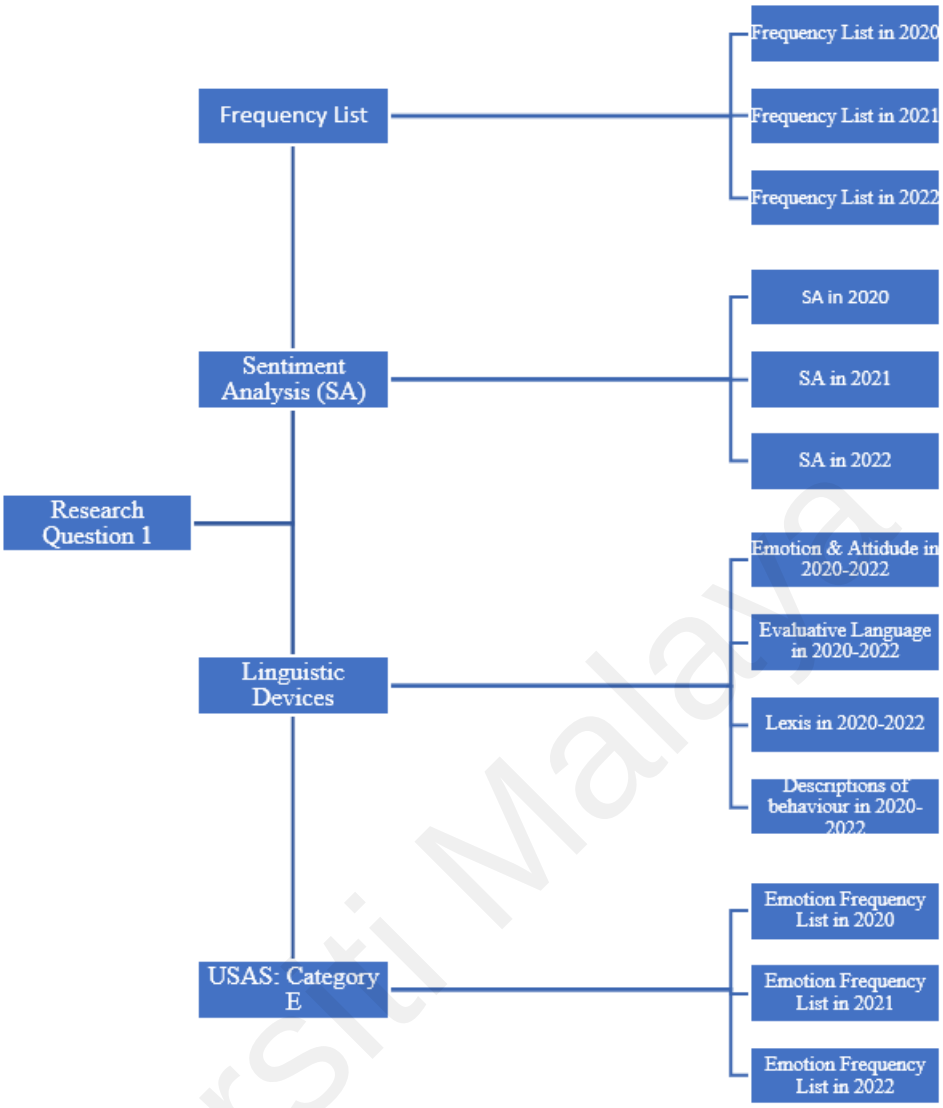

Figure 4.1 Flowchart of Data Analysis Section for Research Question 1

Figure 4.2 is the flowchart of the process in analysis of Research Question 2. The analysis will be elaborated in Section 4.4 below. In order to answer this research question, the data gathered from the analysis of linguistic devices in Section 4.3 has contributed toward generating relevant themes in the corpus. Moreover, the themes derived from data provide useful insights regarding the change in sentiments during the 2020-2022 timeframe. Themes inspired by Bednarek and Caple (2017) & Bednarek and Caple (2016) and themes that the researcher inductively discovers will be elaborated.

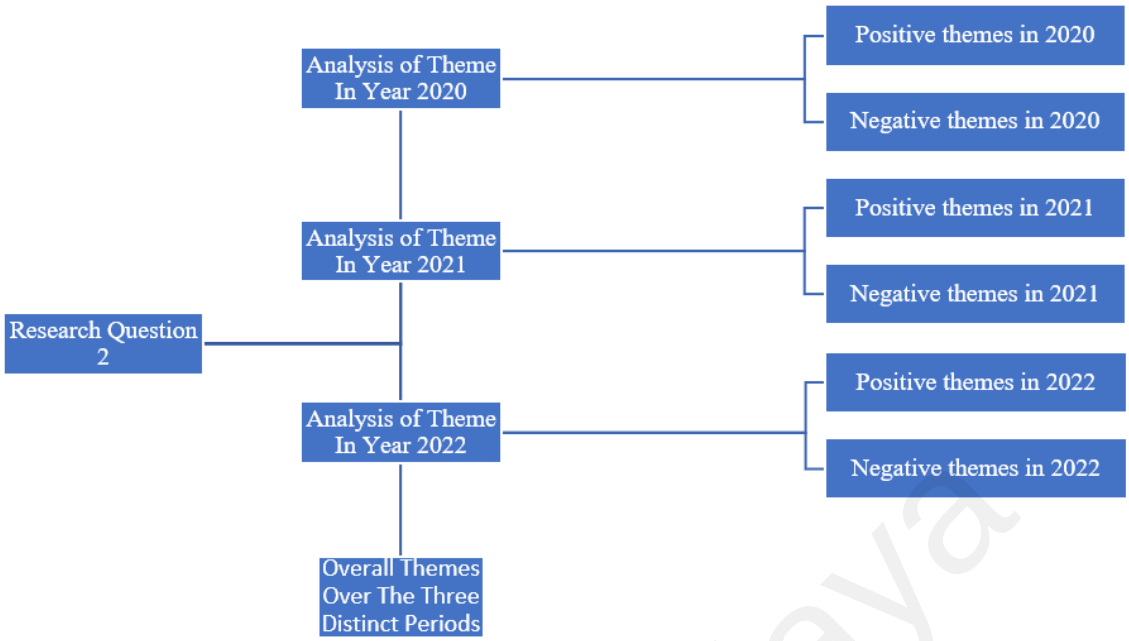

Figure 4.2 Flowchart of Data Analysis Section for Research Question 2

In short, answers and thorough analysis for Research Question 1 is provided in section 4.2 to 4.3 while section 4.4 is for Research Question 2. The discussions are also provided for both research questions in the section 4.5.

4.2 Frequency list and Sentiment Analysis

Frequency list was generated by using NVivo and Wmatrix software in order to see the overall recurring words that appeared most frequently in the sample texts. This step is important as it is crucial to consider a concept or issue in the text entirety. It can also aid in understanding the situational context and obtain general overview of the data based on the high frequency word list. Analyzing word frequency in articles is vital to pinpoint recurring terms, helping identify key themes and categories. By tallying how often words appear, this method offers a quick overview, aiding comprehension of the text's context and highlighting essential concepts. Besides that, in order to investigate the overall emotional undertone of the corpus over the previous three years, the Sentiment Analysis feature in the NVivo program was used to generate 4 types of polarities which are very negative, moderately negative, very positive and moderately positive for this

research. This program carries out a comprehensive semantic parsing of words and generally classifies them based on the positivity and negativity of their lexical meanings. This classification makes it easy to grasp the content's emotional tone and offers insightful information about the general sentiments around the research topic. This procedure improves the analysis's applicability to the research question while also making data interpretation easier. With that being said, it is crucial to incorporate both of the analysis in this section to enhance its relevance to the Research Question 1 before delving into Section 4.3.

For the results in NVivo, the frequently list has excluded some of the parts of speech like preposition, adverb etc while for Wmatrix, the researcher has to clean up the frequently list by focusing on nouns and adverbs and exclude some of the part of speech like adverb, preposition, pronoun, conjunction, interjection and articles. Figure 4.3 is the results and outcomes of the frequency lists while Figure 4.4 below present the polarity chart of the Sentiment Analysis for each year from 2020 to 2022.

4.2.1 Overall frequency list of NVivo & Wmatrix from 2020 to 2022

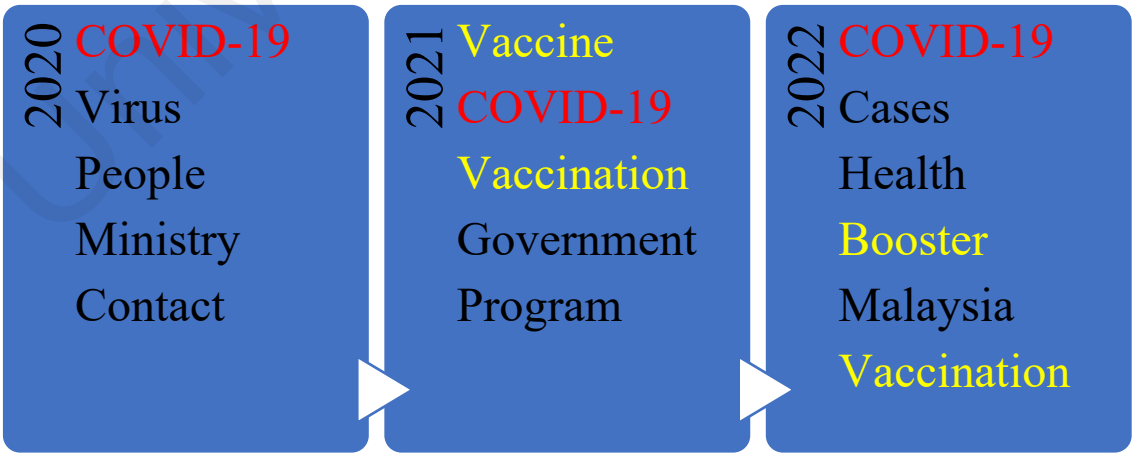

Figure 4.3 Overall Frequency List from 2020 to 2022

Based on Figure 4.3, there have been significant changes of events in the last three years. COVID-19 is one of the most recurrent words for all three years which is not a surprise since this research indeed revolves around sentiments toward the virus. In 2020, it is noticeable that in the beginning of the outbreak, the public only started to get to know about COVID-19 which was defined as illness caused by a novel coronavirus and how the ministries involved were responsible for briefing the public about this health crisis and started working on contacttracing in Malaysia.

Following that, as the outbreak continued for a year until 2021, it is possible to say that vaccine is the word of the year as it is universally acknowledged that the word is widely used and also the subject of attention in the world. The term "vaccine" can elicit sentiments and discussions in a manner it never could have before the pandemic. For some, it is a symbol of hope and health as it has the connotation of returning to the normal life prior to the pandemic but at the same time for others, it was a representation of a politicized issue and center of discussions about political affiliation, safety and other controversies. The amount of available information on vaccinations might be daunting; however, it was important for the public to discuss the issue. Following that, the government was pushing for the agenda of a massive vaccination program which would serve as the key to end the disaster by achieving herd immunity in Malaysia.

In the year 2022, the recovery process from a once-in-a-lifetime pandemic after fighting it for almost two years was evident. In actuality, even after almost two years, questions remain concerning the disease as no one can predict with absolute confidence whether the illness will subside or whether other varieties will appear but this implies that everyone is concerned about their own health. There are still ongoing cases of COVID-19 and the health authorities in Malaysia started giving booster doses to people who had been fully vaccinated at the end of 2021. In fact, health authorities were advising the

public to get booster shots as a form of protection as the vaccinations' efficacy deteriorates with time. Moreover, the shift from pandemic to the endemic phase is an exit plan that will allow Malaysians to resume practically normal living prior to the pandemic which has become one of the joys for all Malaysians in 2022.

All of the major events above explained the results generated in Figure 4.3 and it can be concluded that both the Frequency List in NVivo and Wmatrix has more or less the same result and it is encompassing COVID-19-related terms. By using the frequency list, the overall summary of the data is attainable to view a better picture in light of the related words for all three years.

4.2.2 Overall Sentiment Analysis from 2020 to 2022

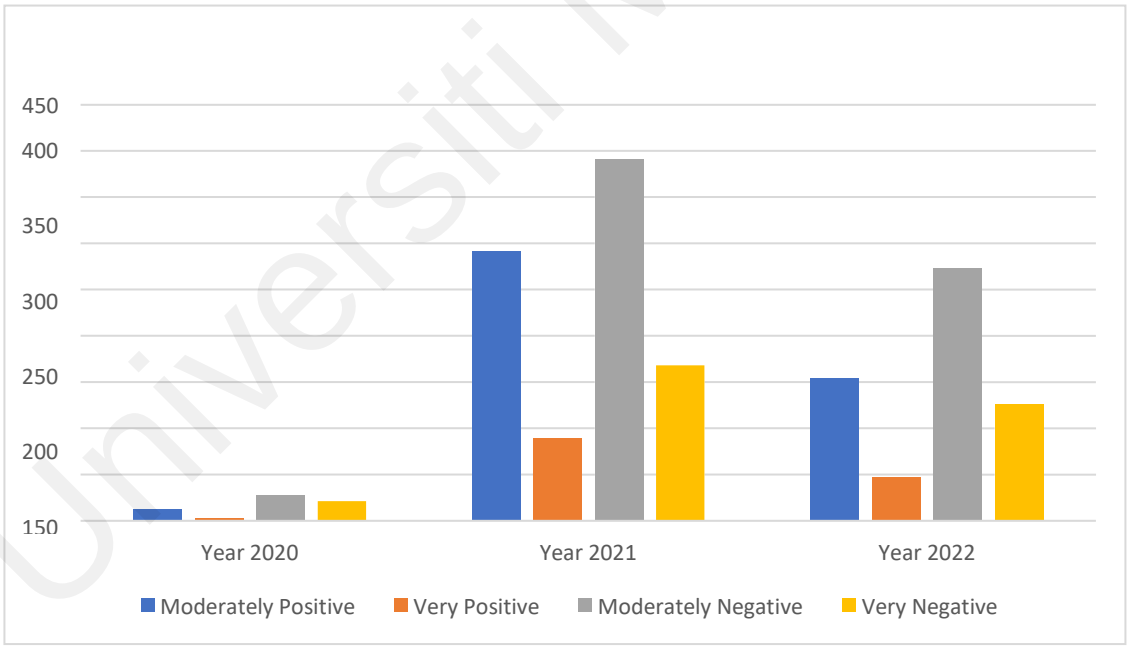

Figure 4.4 Overall Sentiment Analysis from 2020 to 2022

The big differences of result in the Figure 4.4 are due to the difference in the amount of news articles collected for each year in which March 2021 has the highest amounts of articles collected. The overall three years of Sentiment Analysis showed that

the sentiment towards vaccines in Malaysia is moderately negative as it has a consistent result of being the first place throughout the year. Not only that, the polarity of very positive is consistent as the last place in the chart throughout all three years. However, the polarity of moderately positive come-through as the second place for both year 2021 and 2022 indicate that there were positive emotions among the public during that time, reflecting the moods were changing and people were feeling more hopeful with the implementation of vaccination programs. Unlike in 2020, where the first and second places for polarity were moderately negative and very negative, suggesting an overall unpleasant public reaction to the circumstance. For that reason, the overall sentiment towards vaccines in Malaysia for the past three years is negative based on the data generated by NVivo.

4.3 Linguistic devices

In this paragraph, the overall results of Sentiment Analysis from NVivo were analyzed based on the linguistic devices by Bednarek and Caple (2017) for all three years on all polarities. In other words, the results retrieved from NVivo were then analyzed and were manually coded into the category of the linguistic devices in order to look deeper at the linguistic elements for each year and also to avoid irrelevant sentences such as sentences discussing other than COVID-19 related. Figure 4.5 is the flowchart for the data analysis for this section.

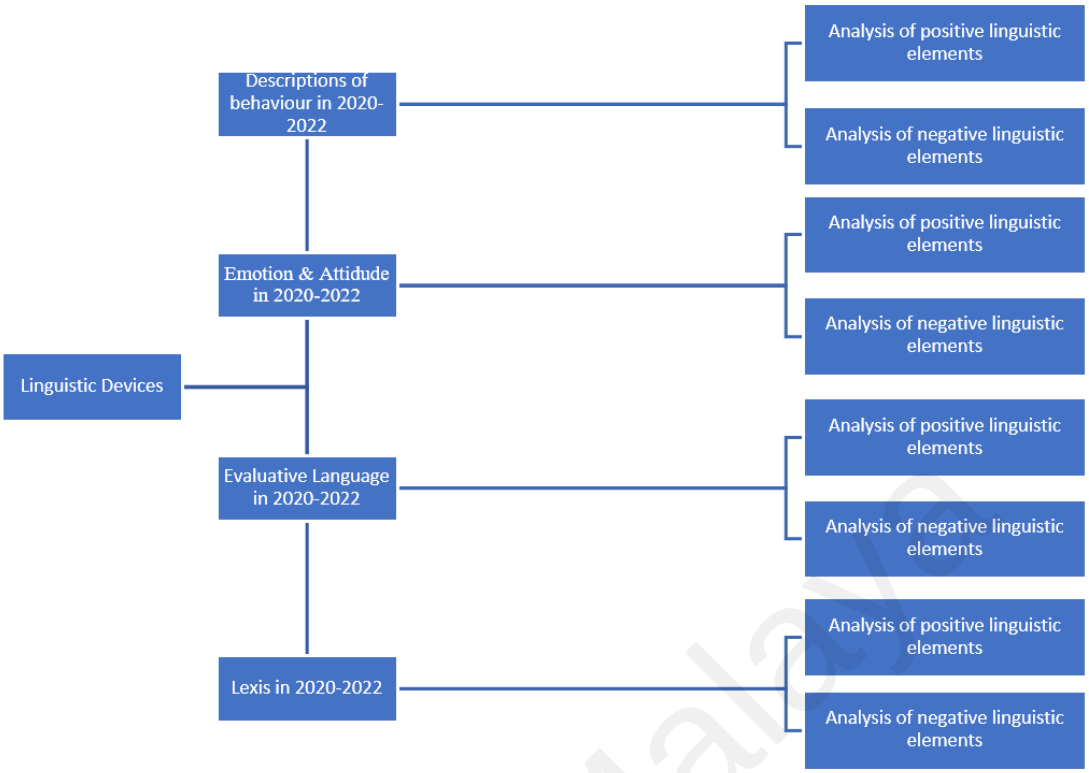

Figure 4.5 Flowchart of Data Analysis of Linguistic Devices

The results of an investigation using the results from NVivo for all three years based on Bednarek's linguistic devices for each polarity to examine the linguistic components are shown in Table 4.1 and a thorough data analysis will be shown in the next section.

Table 4.1: Data Analysis of Linguistic Devices

Year	Linguistic Devices	Linguistic Elements / Polarity			
		Very Negative	Moderately Negative	Very Positive	Moderately Positive
2020	Emotion &Attitude	-	-	-	-
2021		• hesitant, worrying • lack confidence	• Concern • worry	• safer • more confident	• faith
2022		• Afraid • Worry	• Sad • Skeptical	• hope	• Confident • safer
2020	Evaluative Language	-	-	-	• useful, effective
2021		• Horrible	• Discriminatory	• no red flags	• most effective

		• **inconclusive**	• **difficult**		• **could**
2022		• **Terrible** • **Not easy**	• **Inconvenience** • **severe**	• **Receptive** • **significant**	• **aggressive**
2020	**Lexis**	• **under control**	-	-	-
2021		-	• **Deaths** • **Lack, unable to register**	• **approval, safe** • **effective, reducing, good, preventing**	• **safe, approved** • **assurance, rare**
2022		• **death, risk** • **unvaccinated or unboosted, higher risk**	• **Unrelated,** • **Fatality,** • **higher**	• **significant, fewer** • **high, lower**	• **high, lower** • **increase, reduce**
2020	**Descriptions of behavior**	-	-	-	-
2021		• **Spreading** • **Misleading headlines**	• **Spreading** • **Jump the queue**	• **Ensuring** • **updated**	• **Debunks** • **Raise awareness**
2022		• **Prefer** • **spread**	• **Avoid** • **refused**	-	• **Contributes** • **Want, protected**

4.3.1 Analysis of linguistic devices of emotion and attitude

This section will be focusing on the category of emotion and attitude and for this category. Below are the coding results from the findings of Sentiment Analysis in NVivo and the bolded words represent Malaysians' both unfavorable and favorable opinions and views of COVID-19 vaccination in Malaysia from 2020 to 2022.

The examples of negative linguistic aspects are “hesitant”, “worrying”, “concern”, “afraid”, “sad” and “skeptical”. Examples as collected in the news report are shown below:

1. But there are some who are **hesitant**, **worrying** about the possible side effects of the jab, or adverse events following immunization (AEFI).
2. “I heard some people **lack confidence** in the vaccine and believe that it can cause other diseases; That’s totally not true.

3. "**Concern** has been expressed that allowing private clinics and hospitals to offer vaccinations when there is a shortage of vaccines will exacerbate the problem.
4. "I **worry** that this group may be left out of the COVID-19 vaccination program."
5. "I am **afraid** to take the booster after the horrible experience," he said, adding that said he was afraid the booster jab might give him something bad.
6. "I feel **sad** seeing the vaccination rate for children aged five to 11 is only at 30%.
7. Mohd Khairy Zakaria, who has received two doses of vaccine, said he is **skeptical** about taking the booster shot as his friends had told him about the negative side effects of the vaccine.

The highlighted text below emphasizes Malaysians' negative views of COVID-19 vaccination concerns in Malaysia. These lexical elements suggest that issues are a factor in or a cause of catastrophe, chaos, and conflict. The bolded items below underline the unfavorable feelings and impressions towards the vaccine as something that might be dangerous and staking their life for. The word implied the anxiousness, uncertainty and discomfort of the public due to concerns about vaccination safety and effectiveness as shown in example 1 and 2 which suggested the overwhelming negative sentiments towards the vaccine. Not only that, the uneasiness of the availability of the vaccine extracted from example 3 and the restlessness when one will not be able to receive the vaccine as stated in example 4 suggested the overwhelmed negative sentiments towards the vaccination program. Moving forward, these lexical components imply that problems contribute to or are the root of catastrophe and chaos. The relevant word that is associated with the dejected emotion against the circumstances of COVID-19 vaccination rate as stated in example 6 suggested the negative sentiments in relation to the vaccine scenario.

The examples of positive linguistic elements are “safer”, “confident”, “faith”, “lucky”, “privileged” and “hope”. Examples as collected in the news report are shown below:

1. It is not solely for me or business delegations to go abroad, but once we are inoculated, we feel **safer**.
2. The rakyat will be **more confident** once more frontliners had received their vaccine,” he said.
3. “I have **faith** in the vaccine and just like other types of vaccines in the past, its side effects will be manageable,” he says. “For me, I have **faith** in vaccination but I do know of others who are not so convinced.
4. “I feel **lucky** to be vaccinated,” she said, adding that she also felt **privileged** she would be starting work in the COVID-19 ward soon.
5. “We **hope** the government will help bear the cost of vaccination as most SMES are still recovering from the economic impact of COVID-19 lockdowns,” he said.

The top words that were found in the data indicate the issues surrounding COVID-19 vaccination which highlight the positive thoughts and perceptions people have about the vaccination as something that might save their life from the disease. The word implied the assurance, acceptance and certainty of the public towards effectiveness of the vaccination are shown in the example above. Not only that, the desire wanted the best for the people in the country and also for themselves too and the adoration towards the health care system extracted from example 5 suggested the overpowered positive sentiments towards the vaccine and the situation. Besides that, the eagerness to know more and to wait and receive the vaccine extracted from the data and the affirmation that the vaccine is safe after they receive the vaccine suggested the overwhelmed positive sentiments

towards the vaccine. Related words that support the claims are “eagerly wait” and “no pain” (see Appendix A).

4.3.2 Analysis of linguistic devices of evaluative language

This section will be focusing on the category of evaluative language. Thiscategory of linguistic device consists of expressions that reflect a viewpoint or evaluationand are considered to be evaluative or critical. Below are the analysis of positive and negative linguistic devices from 2020 to 2022.

The examples of negative linguistic aspects are “horrible”, “inconclusive”, “discriminatory”, “terrible”, “not easy”, “inconvenience” and “severe”. Samples taken from the news story are displayed below:

1. Raven rubbished claims circulating online about **horrible** side effects and pain right after getting the vaccine injection.
2. Although there appears to be growing evidence that vaccination reduces infectivity, the ability to completely stop transmission is still **inconclusive**.
3. “It would be **discriminatory** if employers make it a job requirement to be vaccinated for existing and future employees.
4. Mamat Yaacob, 73, said that it was **difficult** for him to travel to the designated vaccine centres to register for the vaccine.
5. Businessman Mohd Kamal Shah, 51, said he had **terrible** side effects from his first two shots.
6. However, unlike the homeless who are housed at centres, it is **not easy** to get the homeless out on the streets to get their booster shots,” he said, adding that the state Health Department’s medical officers had gone to the ground to talk to and convince the homeless, especially the elderly, to get their booster shots.

7. “It is definitely an **inconvenience** to return again after three months for another Sinovac booster and to possibly pay for it,” said Poh at the vaccination centre in Persatuan Ang Si Chong Soo recently.
8. On **severe** side effects following the booster shot, Dr Yau said some of these could be linked to a pre-existing condition.

The words in bold above indicate the sentiments expressed which carry the nuances of irritation due to the side effects as shown in example 1 with the word “horrible”. Not only that, doubts towards the effectiveness of vaccines in example 2 were implied in the word “inconclusive. There are also remarks about the process of vaccination explaining the strenuous situation that they need to face, with words like “difficult” and “discriminatory”. Moreover, the words in bold above showed that meanings can be conveyed differently and terms like “inconvenience” and “not easy” reflect the perception of getting vaccinated as shown in example 6 and 7. There are also critical stances towards the after effects of the vaccine that fall in example 5 and 8 using words like “terrible” and “severe” to convey their viewpoint.

The examples of positive linguistic elements are “useful”, “effective”, “effective”, “could”, “no red flags”, “receptive” “significant” and “aggressive”. Samples from news report are as stated below:

1. He said that this would be **useful** while awaiting vaccines and more **effective** treatment modalities to be introduced, on top of stepped-up measures at social distancing.
2. And there have been **no red flags** as far as AEFIs are concerned,” he explains.
3. “Vaccination is the most **effective** way to break the chain of COVID-19.

4. “After I had my second dose (of COVID-19 vaccine) yesterday, I still **could** work a full day today,” he said yesterday.
5. And there have been **no red flags** as far as AEFIs are concerned,” he explains.
6. “After I had my second dose (of COVID-19 vaccine) yesterday, I still **could** work a full day today,” he said yesterday.

The linguistic elements such as “useful” and “effective” were used as praises of the vaccine as it is used advantageously for the society, thus expressing a positive sentiment towards the vaccine. The use of words like “effective”, “could” and “no red flags” conveys the thoughtful perspective on ending the pandemic by affirming the potency of the vaccination and some of the phrases from the corpus that support this idea are all in the example above. The statements that represent an assessment towards the vaccination coverage and also the vaccination shot in Malaysia use words like "significant", "receptive" and "aggressive" which showed support towards the initiative by the government. Therefore, the samples above indirectly express the positive sentiment towards vaccination in Malaysia.

4.3.3 Analysis of linguistic devices of lexis

This section will be focusing on the category of lexis. This category of linguistic device which refers to the full collection of all words that can be used in a language has a mix of both positive and negative sentiments towards issues related to COVID-19 vaccination in the corpus. The study of positive and negative language devices of lexis from 2020 to 2022 is shown below.

The example of the negative linguistic aspects is “under control”, “deaths”, “lack”, “unable to register” “death”, “risk”, “fatality” and “unrelated”. Below is the example extracted from news report:

1. And they can point the way to medicines, vaccines and public health strategies that might bring a runaway crisis **under control**.
2. “Out of one million people vaccinated, only 11 had adverse effects compared to 3,500 **deaths** in one million infections,” she said.
3. The **lack** of a phone and home address could leave the destitute out in the cold and **unable to register** for COVID-19 vaccination.
4. Khairy also said that 105 COVID-19 **deaths** were reported on Wednesday, and that senior citizens who were unvaccinated or had yet to receive their boosters were most at **risk** of contracting the disease.
5. “Unvaccinated or ‘unboosted’ elderly folk with comorbidities have a higher **risk** of dying if they contract COVID-19.
6. She also said 22 out of 45 deaths involving recipients of the COVID-19 booster dose that were reported to the NPRA were found to be **unrelated** to the vaccine.
7. Health director-general Tan Sri Dr Noor Hisham Abdullah said yesterday that the COVID-19 **fatality** rate among the elderly, especially those who have yet to complete their vaccination or receive a booster shot, is higher than other age groups.

In 2020, one of the terms from the corpus that lends credence to the idea of the ongoing pandemic situation is “under control” where it expressed vaccines as one of the ways to keep the circumstances in check. In 2021, the corpus contains a number of words that support the claims to loss of human lives, like “death” in the example of 2 and 4. Following that, it also describes a situation of some under privileged people not being able to register for the vaccination which were hinted at by the words like “lack” and “unable to register”. Regarding COVID-19 vaccination sentiments in 2022, the corpus includes several terms like “death”, “fatality” and “risk” that lend credence to the assertions of the side effects of not being vaccinated in the example of 4, 5 and 7. The

corpus also confirm that the fatality involved in this certain period doesn't involve those who were getting the vaccine shot which were implied by using the word "unrelated".

The examples of the positive linguistic elements are "approval", "safe", "effective", "reducing", "preventing", "assurance", "rare", "significant", "fewer", "high", "lower", "increase" and "reduce". Samples taken from the news story are displayed below:

1. "The vaccines given conditional **approval** by our National Pharmaceutical Regulatory Agency (NPRA) are very **safe**.
2. We are aware that the COVID-19 vaccines are **effective** at **reducing** hospital admissions and clinical infection, but data on how good they are at **preventing** transmission are still uncertain.
3. The government, he pointed out, had given **assurance** that the possibility of an extreme reaction to COVID-19 vaccine was very **rare** and that it would provide compensation should this occur.
4. He announced in Mandarin that the vaccine he received was **safe**, as were others **approved** for use in Malaysia. In a 24-second video posted on Twitter, Khairy joked that the Sinovac vaccine had also improved his command of Mandarin.
5. "Yes the situation is definitely not as bad now especially with our **significant** vaccination coverage alongside **fewer** deaths but there is still uncertainty over a potential outbreak.
6. "With **high** vaccination rates and a good proportion of the adult population having booster doses, we notice that COVID-19 related hospitalization and ICU usage were **lower** during the Omicron wave.
7. "Booster shots will **increase** one's antibody levels and help **reduce** the risk of severe COVID-19 infection.

There are a handful of terms in the corpus relating to the assertion or the affirmation provided by the government and healthcare sector regarding vaccination such as “assurance”, “safe”, “approved”, “approval” and “rare” in example 1,3 and 4. Words that hinted at the efficacy of the vaccination were “effective”, “reducing” and “preventing” which adequately describes how vaccination protects people from infection hence giving a positive sentiment towards the vaccination. There are also a handful of verbs and adjectives in the corpus relating to the decreasing of the fatality rate. By using the word such as “significant”, “high” and “increase” to describe the use of vaccination, the corpus showed that it came with terms like “fewer”, “lower” and “reduce” in the sentences which indicated the benefits of the vaccination as it will help preventing in getting infected by COVID-19. On that account, it is portrayed as a positive sentiment toward vaccination in this period of time.

4.3.4 Analysis of linguistic devices of descriptions of behavior

This section will be focusing on the category of descriptions of behavior. This category of linguistic device which highlights Malaysians' unfavorable manner regarding the cause of strife and turmoil as the issue of propagating fake news and others is not new in our globalized world. Not only that, it also highlights actions that were considered as mistreatment or committing an act of physical abuse or violence of any kind related to COVID-19 vaccination in the corpus. The following section depicts the research of positive and negative linguistic devices of descriptions of behavior from 2020 to 2022.

The examples of negative linguistic elements are “spreading”, “misleading headlines”, “jump the queue”, “prefer”, “spread”, “avoid” and “refused”. Here are a few instances from the news article:

1. Action can be taken under the Sedition Act against those found **spreading fake** news on the COVID-19 vaccine, warns top cop Tan Sri Abdul Hamid Bador.
2. Fadillah, who is also Petra Jaya MP, also suggested the media refrain from using "**misleading headlines**" on COVID-19 vaccines as this could push people away from getting immunized.
3. Individuals **spreading fake** news and instigating people not to take the COVID-19 vaccine will face action under the Sedition Act 1948, says Tan Sri Abdul Hamid Bador.
4. Fines will be imposed on those who **jump the queue** to get their COVID-19 vaccine, says Health Minister Datuk Seri Dr Adham Baba.
5. Most Sinovac vaccine recipients **prefer** to get the same brand for their booster shot, citing "less side effects".
6. Medical Practitioners Coalition Association of Malaysia president Dr Raj Kumar Maharajah said people are still hesitant with regard to taking their booster shots as there is too much fake news being **spread** about the vaccines.
7. "I came for my booster to **avoid** losing my 'fully vaccinated' status in Mysejahtera.
8. Khairy also said that some parents have **refused** to bring in their children for vaccination under the National COVID-19 Immunization Program for Children (PICKIDS) despite having appointments to do so.

In 2021, these linguistic elements suggested that the actions taken are either the cause or factor in the chaos by circulating the fake news regarding vaccination is not new in our globalized world and has become a huge danger to different industries and institutions. Just because a story exists on the internet does not imply it is true. The internet is wonderful, but it can also be used to distribute false information and material. There is no real scientific data to back up these misleading assertions and when

misleading information is spread, it may take on a life of its own and have catastrophic implications. It can result in false allegations, health worries, and possibly devastating fake articles. Related words that supported the claims in this corpus were “spreading” and “misleading headlines” in example 1, 2 and 3. Not only that, moving ahead of other individuals waiting in line for anything rather than waiting one's turn is one of the negative descriptions of behavior found in the corpus of this certain period. “jump the queue” hinted at the analysis as it describes the action of some people that wished to get vaccination earlier by cutting the queue which was perceived as a negative sentiment in this corpus. In 2022, the bold sentence emphasizes Malaysians' actions which these words imply that the public's behaviors have caused friction among the citizen by comparing and choosing their own vaccine instead of receiving the vaccine that is provided by the government as shown in example 5 and 8 by using terms like “prefer” and “refuse”. This negative action was brought in light as the refusal of some citizens to opt for the booster shot due to the side effects as shown in example 6 with the statement is backed up by a number of terms from the corpus, including “refuse” and “unverify” (see Appendix A). Moreover, for the sake of maintaining the vaccination status on MySejahtera, some of the Malaysians were willing to get the booster shot as shown in the example 7 with the words “avoid” which is perceived as a negative description of behavior in this corpus.

Aside from that, the examples of positive linguistic elements are “ensuring”, “updated”, “debunks”, “raise awareness”, “contribute”, “want” and “protect”. The following are excerpts from the news story:

1. Therefore, the government is always mindful of **ensuring** that the country is guaranteed a sufficient number of vaccines through international consultations.”

2. “This is to **ensure** that the effectiveness and safety of the vaccine are **always updated** and the ‘benefits over risk’ of this vaccine remains positive,” he said during the ministry’s COVID-19 press conference here.
3. Prof Dr Abhi Veerakumarasivam **debunks** some popular misconceptions and helps viewers to get a better understanding of the COVID-19 vaccines.
4. “Therefore, we will be having nationwide roadshows to **raise awareness** about the advantages of taking the COVID-19 vaccine.
5. When someone reports an AEFI, he or she is providing vital information that is needed to monitor vaccine safety, and this information is also used to report on vaccine safety to authorities, which further **contributes** to the success of immunization programs, though it should be noted that the acceptance of such reports does not mean or suggest that the vaccine caused the adverse event.
6. We **want** everyone to be **protected** and it is also for their own benefit to be safe and fully vaccinated,” he said.

In 2021, there are some terms that exhibit assistance in a time when help is desperately required or mindful gesture to ease the situation regarding the vaccine in example of 1 and 2 that were suggested with terms like “ensuring” and “update” which distinguish as a positive sentiment as it is implying that the people reacted favorably to the scenario. Other situations found contributing factors in solving the chaos like in the example of 3 and 4 that were hinted with words like “debunk” and “raise awareness” in order to instill public trust in the effectiveness and safety of vaccinations. There are further statements that support the assertions, which are supported by a number of phrases from the corpus like “guarantee”, “protect”, “focus”, “fulfilling”, “provide”, “setting” and “speed up” (see Appendix A) which are further viewed as positive linguistic devices in this corpus. For 2022, there are also factors contributing to unity among Malaysians with assurance statements by the medical officers regarding the success of the vaccination

program and expressed their desire to safeguard everyone, hinted with words like "contribute", "protect", and "want" which suggested that the public sentiments in response to the situation were positive.

4.3.5 Analysis of USAS features category E: Emotion

Table 4.2 is the summary of the frequency list for all of the three years generated by Wmatrix for the semantic category E: Emotion and the analysis of the results is explored and divided into few paragraphs below the table.

Table 4.2: Data Analysis of USAS Features Category E: Emotion

2020			2021			2022		
Word	No	Sentiment	Word	No	Sentiment	Word	No	Sentiment
anxiety	2	Negative	confidence	28	Positive	relaxed	10	Positive
worrisome	1	Negative	rest	14	Positive	hit	9	Negative
assaults	1	Negative	concerns	13	Negative	worried	7	Negative
terror	1	Negative	concerned	12	Negative	prefer	6	Positive
comforting	1	Positive	trust	10	Positive	concern	6	Negative

2020 is dominant with negative emotions with words shown in Table 4.2. "Anxiety" which is in the first place is due to the continued spread of false information, particularlyon COVID-19, which the police are involved too hence causing Malaysians to worry. The words "worrisome" and "terror" are about the transmissions of the virus while assault and comforting are about the process of making and getting the vaccine. In conclusion, the result clearly demonstrated that these linguistic elements suggest that COVID-19 is a factor in or a cause of catastrophe, confusion, and chaos in Malaysia and the overall sentiment towards the COVID-19 pandemic situation in 2020 is negative.

The overall emotions in 2021 are positive as the positive emotion word "confidence" has the highest frequency in the Wmatrix emotion list in the corpus. Most

of the articles explained that the public figures and religious leaders need to build, instill and boost the confidence in the vaccination so that the public participate in the vaccination program. This finding is in accordance with Fatah et al.'s (2022) finding which demonstrates the reporting that COVID-19 vaccination in Malaysia is noteworthy from the religious perspective as it shows that religion is a powerful factor that can alter public attitudes on vaccines. The data here is similar to the data analyzed in 4.3.1 in which the word "confidence" is described as positive sentiment. Besides, the word "rest" in the corpus does not concern with the vaccination and correlates with the study.

The negative emotion word "concerns" and "concerned" has the third and fourth highest frequency in the Wmatrix emotion list in the corpus for year 2021 as it occurred 13 and 12 times respectively in the news reports. Looking into the corpus, it is mainly concerning lack of vaccines, about homeless communities being left behind, increase of COVID-19 cases and being scared of the side effects of the vaccine. Not only that, it is also about the uncertainty of the vaccine, worrying about crossing borders between two countries, wanting to get the vaccine shot, and worrying about the slow pace of the public in registering the vaccine and for that reason it is referred to as negative emotion. The data here are identical to the data that were investigated in 4.3.1, where the term "concern*"* is associated with the situation of grief and dissatisfaction about the COVID-19 scenario and is expressed as a sentiment of negativity. The positive word "trust*"* has a total of 15 counts which in the corpus are mainly focusing on asking the public to put trust in the government's national immunization program and in the vaccine as they are the most trusted health authorities. The information presented here is the same information examined in 4.3.1 as both of them were analyzed as the positive sentiments.

Below are the examples to back up the statements.

1. Now we say we give it to leaders first to instill **confidence** and be an example to the people, and still there are grouses.
2. And as regular **concerned** citizens, my family and I, and especially our aged parent of nearly 90 years, would like to be vaccinated as soon as possible.
3. " Malaysians should also **trust** our Health Ministry personnel who are highly qualified to administer the vaccinations, " he said.

In conclusion, the emotion category extracted from Wmatrix implied that in 2021 the overall emotions are a mix of positive and negative sentiments. According to the data extracted here, the linguistic elements that consist of the positive sentiments are slightly higher than the linguistic elements that express negative sentiments. This is because the efforts and initiative by the government to ensure and convince the public concerning the vaccine is identified as a positive sentiment while the hesitation of the public in the vaccine is suggested as a negative sentiment in the corpus. Additionally, all of the emotion lists produced from Wmatrix that are revealed to include both positive and negative emotions, such as “confidence”, “trust” and “concern” have the same emotion evaluated from 4.3.1.

In March 2022, the frequency of negative emotions is slightly higher than positive emotions. The word “relaxed” which is the most recurrent word in this year is generally about relaxing the local regulations like the restrictions and SOPs which are followed by the complete border opening in this year. The negative emotion word “hit” has the second highest frequency which is mainly linked to achieving a particular figure which is about the COVID-19 cases and how the virus has affected us. The word “relax” and “hit” mentioned in the corpus wasn't addressed in the analysis of the linguistic device in the

part before hence it has nothing to do with vaccinations and connects with the study. The negative emotion word “worried” has the third highest frequency in the news reports which mostly refers to their health after vaccination where parents are worried about the side effects for their children and grandparents. It also linked with how the public is concerned in sending kids to school with larger crowds and so it was interpreted as a negative sentiment. The data presented here is the same data examined in 4.3.1 as both are interpreted as a negative sentiment.

The positive emotion word “prefer” has the fourth highest frequency in the corpus is mainly focusing on the public's preference in choosing the vaccine brand instead of what was given or chosen by the government and also having a skeptic attitude as they are unsure of their actions and simply observe what takes place without responding. In the corpus, the public has voiced out that they were keen to choose the type of vaccine that they are fond of. Even though the word “relaxed” which is classified as the positive words in the Category E in Wmatrix, however the corpus has shown that the word “prefer” has the connotation of a negative emotion as the public would want and compare one thing more than another. In this case, it is different with the data analyzed in 4.3.1 as it is considered as negative emotion and attitude while the data here is interpreted as positive emotion. The negative emotion word “concern*”* has the fifth highest frequency in the news reports. In relation to their health following vaccination, the corpus showed that parents were being concerned about the potential ill effects for their family. The data also implied the issues of adverse effects of the vaccine and the company worried about the quarantine requirements for foreign workers. The evidence given here is the same as the data analyzed in 4.3.1 since both interpretations indicate the word as a negative sentiment.

Below are some examples to support the claims.

1. " As parents, I am sure we all are equally **worried**. But I won't stop my son from going to school. "
2. Most Sinovac vaccine recipients **prefer** to get the same brand for their booster shot, citing " less side effects ".
3. Poh, who lives with four family members, said his daughter followed suit by taking Sinovac for all three doses as she had the same **concerns**.

As a result, the emotion category derived from Wmatrix suggested that in 2022, there are a mix of positive and negative emotions, with the negative emotions having a significantly high number of recurrent terms compared to positive emotions in this corpus. It can be seen that the negative sentiment is slightly higher compared to positive sentiment in this year. Taking into account after 2 years of constant battle with the COVID-19, the actions and attempt by the government to evaluate the controls used throughout the phase of transition from pandemic to endemic is identified as a positive sentiment. Not only the public is still having doubt about the vaccination as they have their personal preference for the type of vaccine they want to take but the public is having it hard in living the new norm which suggested a negative sentiment in the corpus. Moreover, all of the emotion list derived from Wmatrix that is revealed to have positive emotion like “relaxed” and negative emotion like “worried” and “concern” has the same emotion analyzed from 4.3.1 except for the word “prefer” which this term is associated with negative emotion.

In short, the emotion list derived from Wmatrix in 2020 depicted that it is overwhelmed by negative emotions with 4 out of 5 of the top words being generated as negative. In contrast, the emotion list in 2021 demonstrated that positive words have

higher recurring rate compared to negative words and for that reason, this implies that the negative emotion has slightly lower occurrences compared to positive emotion in 2021. Last but not least, the emotion list generated from Wmatrix in 2022 showed that there were overwhelmingly negative feelings, with 3 out of the top 5 words being negative emotions. However, after looking closer into the context of the data, it is observable that one of the words from the top 5 in the list that is interpreted as positive emotion is actually a negative emotion given to the circumstances of that year. On that account, the researcher has decided to indicate that the general emotion in this year is negative. Overall, the sentiment towards issues concerning COVID-19 vaccination in the year 2020 and 2022 displayed significantly negative sentiments while 2021 is the only year that has a balance of positive and negative emotion which the negative sentiment is just slightly higher than the positive sentiment throughout the year.

4.4 Themes

As explained in the 4.1, this section is to answer Research Question 2 for this study. The analysis of the themes for three distinct years from 2020 to 2022 will be divided into three different parts below.

4.4.1 Analysis of themes in 2020

Table 4.3: Analysis of positive and negative themes in Year 2020

Positive Themes	Examples
Benefit	He said that this would be **useful** while awaiting vaccines and more **effective** treatment modalities to be introduced, on top of stepped-up measures at social distancing.
Helping	And they can point the way to medicines, vaccines and public health strategies that might bring a runaway crisis **under control**.
Negative Themes	Example
Disaster	Transfusing plasma from recovered patients, which contain antibodies that are effective against the virus, is an unconventional therapy used to treat **severe** or **critical** cases of COVID-19 at a

	time when vaccines and specific medications against the disease are not yet available, according to the latest guideline on the diagnosis and treatment of the novel coronavirus released by the National Health Commission.

Table 4.3 above is the result that has been analyzed and compiled with a few examples to back up the assertion for each theme based on the reference of the conceptualization and the linguistic elements related in the previous data. From what the data demonstrated, the positive expressions were classified based on two circumstances depicted in the positivity news values, which were occurrences connected to benefit and helping. Following that, the negative expressions were categorized based on one of the scenarios indicated in the negativity news values: disaster. There was just one sample under each theme found in this study.

Disasters are abrupt accidents and natural disasters that result in significant property damage or fatalities. As it causes severe mental suffering, the coronavirus disease is referred to as a pandemic disaster. The coronavirus disease has been linked to a wide range of sentiments like stress, anxiety, grief and worry and it's normal to have these emotions during the COVID-19 pandemic. Following that, the news of a COVID-19 has typically been accompanied with groans and worries of safety precautions, but the most recent announcement of antibodies within that time period may really count as good news hence providing a positive sentiment in the events of benefit. Since the situation is alarming, there are instances that medical officers will give assurance regarding the outbreak. With the support given above by describing vaccines might ease the situation as a consequence it brings a positive view towards vaccination.

4.4.2 Analysis of themes in 2021

Table 4.4: Analysis of positive and negative themes in Year 2021

Positive Themes	Examples
Benefit	The **quicker** Malaysians, especially front liners, get vaccinated, the **easier** it would be for us to get our economy back on track.
	This can be achieved through vaccination, which significantly **reduces** the spread of the disease and also protects those who cannot be vaccinated (contraindicated).
Helping	**Addressing** these concerns and other questions to build trust and acceptance is of utmost importance.
	Like his first jab on Feb 24, Muhyiddin said the procedure was **painless** and **quick**.
Vaccine Acceptance	He said people should **not** have any doubt about the vaccination as the vaccine has been proven **safe** and **effective**.
	"I **hope** I can take the vaccine soon so that I can feel safer.
Negative Themes	Examples
Disaster	"**Concern** has been expressed that allowing private clinics and hospitals to offer vaccinations when there is a **shortage** of vaccines will exacerbate the problem.
	The process of obtaining them is very **difficult** as the amount of vaccines produced for the world is **insufficient** despite the manufacturing companies' promise to produce a billion doses of vaccine until 2021.
Death	In a WhatsApp message that went viral, it was claimed that the nurse suffered sluggishness and breathing difficulties **after** her vaccination on March 12 and was found **dead** at home on March 17.
Crime	Individuals **spreading fake** news and instigating people not to take the COVID-19 vaccine will face action under the Sedition Act 1948, says Tan Sri Abdul Hamid Bador.
	Action can be taken under the Sedition Act against those found spreading **fake news** on the Covid-19 vaccine, warns top cop Tan Sri Abdul Hamid Bador.
Controversy	As the race to vaccinate Malaysians against Covid19 gathers pace, homeless people **worry** about being left behind as most do not own smartphones and have no permanent addresses.
	Not having a smartphone or a phone number is one of the issues hampering efforts to vaccinate the homeless community living in the heart of the city and the government has to find ways to address the problem, says MCA Youth chief Datuk Nicole Wong.

Vaccine Hesitancy	"Currently, we also **don't know** if the vaccines will be effective against any new mutations of the virus, and as such, adherence to SOP is still needed even after the population is inoculated," he added.
	Firstly, there are still chances for some vaccinated individuals to **not be successfully** immunized.

According to the data, the negative expressions were classified based on five events portrayed in the negative news values, which were disasters, crime, death, controversy and vaccination hesitancy. Then, the positive expressions were categorized in accordance with the three scenarios which were benefit, helping and vaccine acceptance that were included in the positivity news values. Following that, there are two new themes (vaccine acceptance and vaccine hesitancy) this year that were developed based on the corpus. There were 2 samples under the negative themes of disaster, 5 under crimes, 11 under vaccine hesitancy, 1 under death and 14 under controversy. While in positive themes, there were 6 samples under the positive themes of benefits, 35 under helping and 30 under vaccine acceptance (See Appendix C).

In 2021, the restlessness of the situation to face COVID-19 has led to the rise of fake news on vaccination and cutting the line to get the vaccination. Moving forward, unfortunately, even after a year of fighting the virus, the side-effects after getting vaccinated are common. When a fatality happens soon after receiving a vaccination, family members and others may understandably wonder if the immunization was to blame. The examples of death-related cases in 2021 shown above in the table indirectly supporting the claims. It is not an exaggeration to say that the spread of fake news regarding vaccination on social media somehow affects people's decision in getting vaccines as shown in the example above. In addition, the proliferation of alarming adverse effects from vaccinations influences people's choice to receive vaccinations too as shown in the table. Although getting COVID-19 vaccine is free in Malaysia, the data above

shows that homeless people are struggling in getting them due to lack of smartphones. It is important to stand up for the access to the COVID-19 vaccine as a right and not a privilege. However, the option to refuse immunization does not absolve one from repercussions. The recognition of the COVID-19 vaccine has been considered as good news within this time period as vaccines can protect us against COVID-19 by lowering our risk of infection. Consequently, it implied a positive sentiment towards vaccination hence accomplishing in spreading an affirmative statement in order to influence people in getting vaccinated. Trust in academic institutions and healthcare providers has aided the willingness of the public to get the COVID-19 vaccine.

4.4.3 Analysis of themes in 2022

Table 4.5: Analysis of positive and negative themes in Year 2022

Positive Themes	Examples
Benefit	"Booster shots will **increase** one's antibody levels and help **reduce** the risk of severe COVID-19 infection.
	Dr Sharifa, however, noted the nation's significant vaccination coverage does **help** to prevent cases from recurring here.
Helping	Furthermore, the high vaccination rate among Malaysians, thanks to the **aggressive** National COVID-19 Immunization Program, has helped to substantially diminish the risks posed by the virus at the macro level.
	Dr Kuljit advised Sinovac recipients **not** to **panic**, but to schedule their appointments as soon as possible.
Vaccine Acceptance	After being informed about the home vaccination drive, the 63-year-old was more **receptive** about getting his booster shot.
	"I believe Sinovac will be **safer**, hence my husband and I prefer it to be given to our sons.
Expectant	"We **hope** the government will help **bear** the cost of vaccination as most SMES are still recovering from the economic impact of COVID-19 lockdowns," he said.
	With the vaccination of teachers and parents, we **hope** that this will make schools a safer environment for children as they will be spending at least half a day at these premises.
Negative Themes	Examples

Death	Still, the Health Ministry has constantly **stressed** the need for older folk to get their boosters as most **deaths** have occurred among those aged 60 and above.
Risk	"Unvaccinated or 'unboosted' elderly folk with comorbidities have a higher **risk** of dying if they contract COVID-19.
	Thus, she said they had higher **risks** of mortality compared to those aged 18 to 59 even though both groups were fully vaccinated.
Vaccine Hesitancy	"I am **afraid** to take the booster after the horrible experience," he said, adding that said he was afraid the booster jab might give him something bad.
	Most Sinovac vaccine recipients **prefer** to get the same brand for their booster shot, citing "less side effects".
Controversy	Phee said there were about 300 homeless people in the heart of George Town currently, adding that there was no initiative to pick up homeless people from the streets using mobile units or vehicles.
	"I came for my booster to **avoid** losing my 'fully vaccinated' status in Mysejahtera.

According to the findings in the chart above, the data analysis revealed that the four events that were shown in the bad news values— risk, death, controversy and vaccine hesitancy were used to categorize the negative expressions. The positive expressions were then classified according to the four scenarios that were contained in the positivity news values, which were benefit, helping, vaccination acceptance and expectant. Following that, four new themes were developed based on the corpus which are vaccine acceptance, expectant, risk and vaccine hesitancy in this year. There were 1 sample under the negative themes of death, 6 under controversy and risk and 17 under vaccine hesitancy while there were 4 each under benefit and helping, 5 under vaccine acceptance and 2 under expectant (See Appendix C).

The death cases of COVID-19 are unavoidable for this month due to the omicron variant. The statement above has indirectly mentioned that death will be inevitable for older folk who still haven't got their booster shot and urge them to get their booster shot. After the initial vaccination shot given in 2021 started to wane over time, COVID-19 booster vaccination doses are now being made available in 2022. Not only that, the

difficulty in getting the booster shot has been one of the reasons for the vaccine hesitancy in 2022 as one of the individuals expressed the process of getting the shot as a nuisance. In addition, one of the causes of the vaccine hesitation is that one can compare and select their own brand of vaccination instead of using the one that the government provides hence leading to delaying in getting the vaccine as shown above. Besides that, due to the current prevalence of the omicron variety, the exposure of the danger is unavoidable. Example above also described that those who are vaccinated and aged 60 and above claimed that their mortality risks were increased hence are urged to get booster shots. Furthermore, example above exposed the waning immunity, which is the gradual loss of antibodies where for many illnesses and immunizations, the precise duration of protection varies hence affecting the rise of COVID-19 death rate in the period. Likewise, the data above concluded that booster shots of COVID-19 vaccination may lower the chance of contracting an infection of another variant and developing a serious disease. In the year of 2022, one of the controversies found in the corpus was that there is a claim that underprivileged people such as the homeless people were not reached out and were not given the access of the booster shot as shown above.

Moving forward to positive themes in 2022, vaccinations can protect us against COVID-19 by reducing our chance of infection. People are predicted to be more eager to get vaccinated as a result of the information availability, which is thought to help them comprehend the pandemic and vaccinations better and sufficiently. In addition, the COVID-19 immunization had a significant influence on the course of the pandemic in Malaysia. The examples indicated the unwavering support from the government, higherup and also from the public that demonstrate how it helps to the vaccination campaign's success. It is also intended to convince and increase the people’s knowledge about the vaccines, which should increase people's desire to receive vaccination. Last but not least, expectant is a sensation of excitement that something positive is about to happen

and above are some instances that describe the anticipation of the public towards the pandemic's situation in the year 2022 after vaccination.

4.4.4 The overall themes over the three distinct periods

Table 4.6: Data Analysis of the overall themes

Year	2020		2021		2022	
Sentiment	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**
Themes	**Benefit**	**Disaster**	**Benefit**	**Disaster**	**Benefit**	**Death**
	Helping		**Helping**	**Crime**	**Helping**	**Risk**
			Vaccine Acceptance	**Vaccine Hesitancy**	**Vaccine Acceptance**	**Vaccine Hesitancy**
				Controversy	**Expectant**	**Controversy**
				Death		

Table 4.6 above is the summary of themes for the past 3 years and it is mostly influenced by the different occasion and activity that happened in each year. The color difference in the table above is to differentiate between the themes that were inspired by Bednarek and Caple (2017) & Bednarek and Caple (2016), which is in black color and the one that inductively generated by the researcher is in red color. It is notable that the theme of helping and benefit which express the positive expressions exist throughout the whole three years, while the theme of disaster exists in 2020 and 2021 and the theme death exists in 2021 and 2022. It is distinguished that four of these themes are heavily connected and relevant to the situation of the pandemic COVID-19. Besides that, the theme of vaccine acceptance, vaccine hesitancy and controversy come into sight in both the years of 2021 and 2022 which can be assumed that there were a lot of hurdles that went through across the year compared to 2020 which does not have any of these themes.

In 2020, for the events of the disaster, the coronavirus illness has been related to a wide range of emotions from anxiety to fear and is something that everyone feels when

one’s in danger or threatened hence it's a natural reaction and these feelings are typical during the COVID-19 pandemic. On the other hand, the positive expressions were classified to the themes of benefit and helping are presented in the positivity news values as it is associated with words such as “recover”, “effective” and “reduce”. Around that time, the scientist announced the discovery of antibodies. Despite the fact that the situation is concerning, the medical officer's statement regarding the antibody and how effective it is towards the disease at the specific time generates positive sentiments for both of the themes.

In 2021, the rise of COVID-19 cases and confirmed deaths has been continuous in this particular month as the vaccination drive has just been launched less than 2 weeks from this month. “Concern”, “difficult” and “insufficient” are some of the words that are related to the theme that depict the difficulties and struggle that they had encountered over the years in the midst of the outbreak. Along with the events of crime, the emergence of fake news regarding vaccination on social media, particularly during the COVID-19 pandemic, has become a huge danger to Malaysia's numerous industries and institutions. Scammers' techniques are growing more complex, and they are continuously seeking for new ways to prey on naive and vulnerable people. It’s one of the situations that couldn’t be prevented given the situation at that time as everyone was panicking and defenseless at that time. Fatah et al. (2022) also asserted that it is crucial for Malaysian media to encourage constructive debate and dispel incorrect information and preconceived notions by placing a strong focus on the available scientific facts. Saleh et al. (2023) also mentioned that comparisons to influenza and its vaccine, along with discussions on conspiracy theories, were prominent topics with negative sentiment.

The public health efforts as well as viewers' faith and credibility in the media might be significantly impacted by the media's distortion of truth. Besides that,

hospitalizations and fatalities from the coronavirus are unavoidable hence the word like "death" that is associated with such a situation is expressed as the negative expressions in this year which leads to the creation of the negative theme. Furthermore, the theme of controversy also emerged in this year despite the scientific accomplishment, the path from vaccine development to herd immunity against COVID-19 remains ambiguous with policy problems that need a cooperative response from the public. In Sutrave et al.'s (2021) study, one of the developing topics in negative sentiments related to the COVID-19 vaccine is concerns regarding the safety and distribution of the vaccine. A conflict occurs when inaccessible healthcare and privilege exists as it is referred to when an individual or organization has an undeserved advantage over others. As the COVID-19 outbreak has vividly depicted, individuals with less privilege frequently had to pay the extreme price during times of crisis hence terms associated with the scenarios such as "worry", "discriminatory", "left out" and "lack" has inspired the creation of the theme.

Given the initiative of the government to kick off the massive vaccination program in Malaysia, there are some that ponder over getting the shot as it is correlated with their own health. Kumar et al. (2016) stated that many experts believe that addressing vaccination reluctance at the community level is the most effective strategy. They think it can be done by increasing transparency in policy decision-making prior to immunization programs, informing the public and healthcare professionals about the rigorous processes followed before the introduction of new vaccines. However, given the situation in Malaysia, vaccine hesitancy is an especially pressing issue in light of the COVID-19 pandemic this month as COVID-19 booster vaccine doses are now offered after the effects of the original immunization injection administered in 2021 began to fade over time. "dangerous", "worrying", "lack of confidence", "pain", "horrible" and "concerning*"* are words that unconsciously link with the vaccination which express the negative sentiments of not wanting to get vaccinated and there are more claims that back

up the assertions by using the linguistic elements like “horrible” and “lack” (See Appendix C). Correspondingly, Melton et al.’s (2021) results revealed that during the LDA topic modelling, terms suggesting vaccine hesitancy were discovered.

Positive news values, on the other hand, were divided into themes such as benefit, helping and vaccination acceptance, which are associated with terms such as “effective”, “acceptance”, “debunk”, “raise awareness”, “speed up”, “reduce” and much more (See Appendix C). It is also claimed by Sattar and Arifuzzaman’s (2021) study that the public are more pleased about taking COVID-19 shots than they are about certain vaccines' side effects. Not only that, it also leads to the circumstances that demonstrate vaccine approval and getting support from the public which the words that support the claims were “confident”, “lucky”, “safe”, “faith”, and “grateful” that tally with Syed Alwi et al.’s (2021) study, thus indicating that Malaysians have a high degree of acceptability for the COVID-19 vaccination.

In 2022, although there are themes that express negative expressions which overlap with those in the previous year such as death, vaccine hesitancy and controversy, there are also new themes that emerge in this year. The theme for death is uncommon when one is discussing the pandemic. After two years of battling the pandemic, the social and economic repercussions of the pandemic are numerous and varied, ranging from school closures to destroyed industries and millions of job losses. The fact that COVID-19 threatens to expand inequality around the world and among other things is still as depressing as ever. Although there is no generally ‘accepted’ number of hospitalizations and fatalities that society will tolerate, there are some scenarios that most people will try to avoid which is getting infected and impacted by the disease. In this particular month, the spread of Omicron variant is bound to cause an increase of cases and death. The theme for controversy can’t be ignored as it emerged in this year too as it cannot be denied that

there are lives that were being impacted and there were difficulties encountered for post-covid such as to cope with the new norm. The example of the new norm would be the mandatory wearing of mask in public, work from home, online meeting and exercising social distance in order to avoid and decrease the spread of COVID-19 and there are people who find it difficult to face all of these new norms. Yet, this runs counter to Sattar and Arifuzzaman's (2021) research where the scholar examined the public's healthy lifestyle following immunization and discovered that positive sentiment outnumbers negative sentiment as people are enthusiastic about preserving hygiene while simultaneously being optimistic about social distance.

Following that, there are instances of vaccine hesitancy for the booster vaccine which were offered in 2022 as they wanted to choose the type of vaccines based on what they heard or were keen on. It was divided in opinion that some vaccines had harsher side effects and they wanted to avoid it. The word that describes the situation and supports the claim of the statement is the use of the word "prefer". After the COVID-19 booster vaccine doses were offered to the adults in 2021 which also continue to be offered in 2022, the COVID-19 vaccine has now been authorized by the Malaysian government for children aged 5 to 11 years which was already implemented a month before this particular month in 2022. With that being said, the example of words that describe the situation of the refusal of the booster shot and shot to children by the parents are "fear", "afraid", "scared", "hesitated", "terrible", "skeptical" and more. This aligns with the findings of Park and Suh (2023), revealing a persistent negative public opinion, especially in sentiments related to the Pfizer vaccine, spanning the entire year since the initiation of COVID-19 vaccination. Following that, the new theme that appeared in this year which is risk is due to the current prevalence of the omicron variety which has the potential of a new risk. Hence, the exposure of the danger is unavoidable and new policies were implemented during that time. In this case, some Malaysians are receiving the additional

vaccine shot despite not being really eager for the sake of maintaining their vaccination status on MySejahtera. For that matter, Mohamed et al.'s (2021) research discovered that people were more likely to get immunized if they were female, in a younger age group, and had higher level of education.

On the other hand, there are themes that express positive expressions that overlap with last year such as benefit, helping and vaccine acceptance. There is a new theme that emerges this year which is expectant. The success of the pandemic in Malaysia can be seen due to the public efforts to break the cycle of infections, which resulted from a timely and comprehensive approach that was implemented from the very beginning. Additionally, the use of contact tracing techniques which has been crucial for the active detection of positive patients, the cooperation given by the vast majority of Malaysians to get vaccinated and adhere to the restrictions, all the frontliners' sacrifice in saving lives lead to the success of shifting the pandemic to an endemic phase in Malaysia. "Contribute", "success", "aggressive", "safer" and more are the words that represent the situations that express positive sentiments. As a result, the term "hope" is used to represent the public's anticipation of the pandemic's position in 2022 as well as the feeling of exhilaration that something good is going to happen. Similar to Sutrave et al.'s (2021) results, some of the emerging themes in supportive tweets about the COVID-19 vaccine include the public's gratitude for the leadership's efforts to hasten vaccine development, appreciation for pharmacies' contributions to the community's access to COVID-19 doses, and a sense of security among the general public as a result of herd immunity that the vaccine may bring.

4.5 Discussions

Based on four analyses presented earlier, namely 1) Frequency Lists, 2) Sentiment Analysis, 3) Linguistic Devices 4) USAS Features Category E:Emotion, it is clear that

the sentiment surrounding vaccination is seen as more negative than positive. Both sides, however, have opposing interpretations of vaccination, which are examined in the following section. The discussion towards Research Question 1 will be in section 4.5.1 while section 4.5.2 is for Research Question 2.

4.5.1 Linguistic elements

The word frequency study of the corpus related to COVID-19 vaccination in NVivo and Wmatrix found that nouns are the most frequently used words in the corpus. Some nouns that appeared with the node word vaccination, such as COVID-19, virus, vaccination, booster, government, program and cases which exhibit these recurring words are prevalent and general in relation to COVID-19 vaccination news. These nouns also depicted the general situation that happened in that certain year which also acted as a representation word by aiding in the discovery of the connections and patterns among them as shown in Section 4.2.

Sentiment Analysis in NVivo is an analysis wherein textual input is automatically analyzed, interpreted, and given different sentiment indicators by a computer. It is detected that NVivo analyzes the sentiment of words on their own by finding the sentiment expressions in the source material which then encodes them in sentences. Moreover, the data should typically be seen from the perspective of time as the sentiment very likely varies with time. The findings of the analysis for the past three years indicate the overall sentiment in the corpus is negative according to NVivo statistics which is a totally opposite with the findings of Melton et al. (2021), Saleh et al. (2023), Sattar and Arifuzzaman (2021) and Sutrave et al. (2021) as all of the them exhibited more positive sentiment than negative sentiment towards vaccine-related topics. Having said that, the words found in the corpus pertaining to vaccinations in Malaysia consist of more frequent words with negative connotations compared to positive connotations.

It is evident from the analysis of linguistic devices that there are both positive and negative linguistic components, which motivate the researcher to look for the factors that led to the use of language to convey the respective sentiments. In the category of emotion and attitude in one of the linguistic devices, usually the use of adjectives is more frequently found in the corpus to convey their emotion. Some of the factors that lead to the usage of negative adjectives such as “hesitant” and “concern” is connected to the fear of a vaccine’s harmful impact and confidence and readiness of the public to receive vaccinations which is linked to the use of positive adjectives like “safer” and “hope”. Simultaneously, it is similar to Saleh et al.’s (2023) findings where while there is an overall positive shift in public perception, identified concerning trends, particularly within specific topic and demographic clusters, indicating potential areas of COVID-19 vaccine hesitancy. Not only that, Melton et al.’s (2021) findings indicate that discussions about the COVID-19 vaccine or experiences with receiving the vaccine generally receive positive public sentiment in spite of certain keywords and topics that suggest some hesitancy among the public.

Moving forward to the next category of linguistic devices which is the analysis of the evaluative language, the usage of adjectives, adverbs and modal verbs is more frequently observed in the corpus in order to indicate the value of particular objects or topics. Since each individual has a different viewpoint on certain issues, each individual's evaluative perception will differ. Some of the causes of the use of negative adjectives such as “horrible” and “discriminatory” is related to the skepticism over the efficacy of the vaccine while the use of the adverb of “not” is linked to the remark explaining the strenuous vaccination's process. In essence, the results correspond to Sutrave et al.’s (2021) finding that the negative sentiments are due to worries about the effectiveness and serious side effects of vaccines and also Wong et al.’s (2023) finding which showed that the doubts and challenges surrounding the COVID-19 vaccination have prompted a

number of serious issues in Malaysia. Moreover, there are positive adjectives such as "effective" and "significant" expressing the potential of vaccination to benefit society as a cure, the usage of the modal verb "could" and adverb "no" is connected to the remark of demonstrating the effectiveness of the immunization.

In the analysis of lexis, some of the negative verbs like "death" are frequently used to back up the assertions about the loss of human lives while there also contains a small number of positive verbs and adjectives referring to the reduction of the mortality rate such as "increase" and "low". Last but not least, the use of verbs is widespread in the analysis of descriptions of behavior for example by utilizing some of the negative verbs like "spread" and "avoid" to indicate the situation of disseminating false vaccination information and refusing to receive an extra vaccination shot due to adverse effects. Some of the reasons for the use of positive verbs like "debunk" and "contribute" are to demonstrate the situation of some officer alleviating the fake news surrounding the vaccine in order to promote public trust in the effectiveness and safety of vaccinations and explaining the factors that contributed to the success of the vaccination program. In concordance with Nasyaya et al.'s (2023) study where it showed that the factors such as public trust in accurate information, the influence of community leaders, transparent data, and effective communication by health authorities and the government play pivotal roles in shaping a positive view of vaccination.

The findings of the emotion list generated according to rank frequency order from Wmatrix illustrated the lexicon that is used in conjunction with one another associated with emotion. It shows that both 2020 and 2022 exhibited more words with negative emotion such as "terror", "anxiety", "worried" and "concern" which are associated with fear. In relation to this, it is parallel to Nor and Zulcafli's (2020) study where the majority of the collocates represent the worry, anxiety, and uncertainty that the most of

Malaysians. On the other hand, 2021 comprises more positive emotion words such as “confident” and “trust” which are associated with optimism. Nonetheless, it is implied that the vaccination has an impact on Malaysians in all facets of life, including health, society, psychology, and other areas as the data interpret the sentiment towards vaccination as both positive and negative throughout three years. In actuality, the data came upon more unpleasant and challenging sentiment compared to favorable sentiment. Besides that, the findings of this list strengthen the claim and analysis of linguistic devices in the category of emotion and attitude as most of the linguistic elements are found in Section 4.3.5 which is a proof that the words are in the right category with their respective sentiment. Despite everything, the usefulness of the instruments is demonstrated by the quantitative analytical help provided by automated semantic analysis, which facilitates the researcher's interpretive findings.

4.5.2 Changes in sentiments

The overall themes suggested that from 2020, it consists of more convincing linguistic elements and consequently portrayed as positive sentiment as people were enlightened with the news of the pandemic being designated as a worldwide emergency which brings up the vaccination news and how persuasive it is against the disease at the moment. Moving forward to themes in 2021 where it represent more negative themes compared to positive themes as a large number of linguistic elements were observed that linked with various occasions. Previous study showed that the main focus of the public’s worries while talking about vaccines is the intimidating and inexpedient strategy used to distribute information about vaccinations in Malaysia (Fatah et al., 2022). Some of the events occurred that influenced the negative conception are situation of hospitalizations and deaths as a result of the vaccination shortage, the rise of fake news of vaccination, the possible adverse effects of the vaccination which are some of the factors contributing to the public's rejection to be vaccinated and the lack of vaccine access for

underprivileged groups, who are more susceptible to COVID-19 due to their living arrangements or jobs. Past studies showed that the majority of individuals showed positive sentiments like trust and eagerness, showing faith in medical professionals and immunizations. As encouraging findings from the clinical studies were published, people expressed their happiness as they thought that a vaccination would stop their suffering since COVID-19 has had severe effects on many facets of human existence (Sutrave et al., 2021). In connection with this, some of the events found in this corpus associated with positive themes are higher-ranking authorities and well-known people publicly endorsing vaccination and stressing its safety in an effort to persuade the public and the abundance of data on the effectiveness of vaccines accessible in the news report.

Last but not least, the overarching themes for 2022 indicated a balance between topics concerning the use of linguistic elements to represent occurrences that elicit both positive and negative sentiments. Some of the incidents involving urging booster shot as most deaths have occurred among elderly person, the inconvenient vaccination procedure, influenced by unfounded reports linking immunizations to deadly consequences and death, single out certain vaccination brands, taking the booster shot as a means not to lose the vaccination status and many more that shaped this negative themes. Additionally, it is aligned with the result of Nasyaya et al.'s (2023) study showed the failure to address misinformation seriously may lead to the emergence of negative opinions, apathy, and public rejection of the COVID-19 vaccine.

Past study exhibits that some of the themes seem to be directly linked to a discussion of the vaccine's effectiveness, safety issues, and possible adverse effects (Melton et al., 2021). In addition, some of the instances from the corpus that are interpreted as positive sentiments in 2022 are information availability regarding the efficacy of vaccine, government and health officer consistent support which led to

effective immunization programs, the persuasion of medical officers, celebrities, religious leaders and higher-ups to instill and boost the effectiveness and safety of vaccine and much more that guide in forming positive themes (see Appendix C). Syed Alwi et al. (2021) also asserted that in spite of high incidence of vaccine acceptance, it is still vital to allay the worries of the vaccine's doubters by fostering confidence in the safety and efficacy of the vaccination through accurate vaccine information.

4.6 Summary

The results of the data analysis have been presented in this chapter 4. The first research question was addressed, and there were both positive and negative linguistic elements to the linguistic devices, with the amount of negative linguistics elements are lightly outweighing the positive linguistics elements in this study. There are several examples in the data that show both of the sentiments towards the COVID-19 vaccine in Malaysia. In the corpus, the use of adjectives, adverbs, and modal verbs is more commonly used to emphasize the importance of certain subjects pertaining to the effectiveness of the vaccine, vaccine adverse effects, the issue of spreading misleading vaccination information and others. To address research question 2, the linguistic element analysis was further explored. All language components were then analyzed to generate the primary core themes in order to investigate how sentiments change over time. The overall themes suggested that at the beginning, 2020 contained more compelling linguistic features and was depicted as positive sentiments throughout the year. Following that, in 2021, the data showed more negative themes than positive ones, as a great number of language features associated with various occasions were identified such are the hospitalizations and deaths caused by vaccine shortages, the growth of vaccination-related fake news, and the potential detrimental effects of vaccination. Last but not least, the main themes for 2022 suggested a balance of issues using language elements to

portray events that provoke both positive and negative sentiments. Chapter 5 will conclude the study by summarizing key findings presented in this chapter.

CHAPTER 5: CONCLUSION

5.1 Introduction

The researcher wraps up this study on the Sentiment Analysis of news reports in Malaysia regarding the COVID-19 vaccination in this chapter. The study employs a corpus-based approach to analyze the linguistic components that were employed to convey the sentiment towards vaccination in the news report and offers some insights into the understanding of public opinion on vaccination. In this chapter's subsequent sections, the researcher summarizes key findings (Section 5.2), discusses the implications of the research (Section 5.3), reflects on the limitations of this study (Section 5.4), proposes ideas for future research (Section 5.5) as well as providing a conclusion for the research (Section 5.6).

5.2 Summary of key findings

The goal of the current study was to analyze the language used in the news story to convey the sentiment regarding the COVID-19 vaccination. Both of the research questions raised in Chapter 1 have been addressed, serving the intended goal.

The most important findings from Frequency List in NVivo and Wmatrix were the top frequency list in 2020 focusing on the origin and explosion of COVID-19. Next, at the top frequency list in 2021 is the word “vaccine” which is associated with the government’s effort in launching a huge vaccination program. Last but not least, the top frequency list in 2022 disclosed the rising incidents of COVID-19 cases due to new variants and the continuing efforts by the healthcare sector in distributing booster doses to Malaysians. Besides that, the most crucial conclusions of Sentiment Analysis studies in NVivo during the last three years revealed that the sentiment is relatively negative and it can be perceived that there were many linguistic elements that were associated with

negative sentiments hence the sentiment is consistently negative throughout the year. It is also possible that since the news reports collected in this study were mainly focusing on infectious disease, this indirectly leans toward negative sentiments due to the language used. Conclusions drawn from USAS characteristics in Wmatrix were linguistic elements in 2020 and 2022 being mainly negative due to concern expressed in the news report about the vaccination while the opposite was observed in 2021.

Furthermore, there are four categories of linguistic devices analyzed in this research. First, the main findings for the analysis of linguistic devices of emotion and attitude revealed that some of the negative linguistic elements reflected the public's apprehension, uncertainty, and discomfort as a result of worries about vaccination safety and efficacy. Likewise, the concern over the vaccine's accessibility and the anxiety about not being able to acquire the vaccine both hinted and aggravated the negative sentiments towards the vaccination program. Concurrently, there were terms that signified the public's assurance, acceptance, and certainty over the efficacy of the immunization. In addition, the curiosity to learn more, the anticipation of receiving the vaccine, and the confirmation of the vaccine's safety upon administration all pointed to the overwhelming positive sentiments towards the vaccine were expressed in this category. Besides that, the analysis of linguistic devices of evaluative language presents a thoughtful viewpoint on eradicating the pandemic by emphasizing the effectiveness of the vaccine owing to the ability to treat, it can be utilized in society for the better. After everything, doubts on the usefulness of vaccines were conveyed on account of the side effects.

Moving on to the third category, primary conclusions for the analysis of lexis was mostly involving the lives and the risks that the public must deal with while facing the virus. The corpus also contains a small number of verbs and adjectives that refer to the decline in fatality rates and highlight the advantages of immunization in reducing

COVID-19 infection. Last but not least, the analysis of linguistic elements used to represent the category of descriptions of behaviors exhibited that there is a variety of actions and characteristics by the public that have sparked conflict among locals. Since it is human nature to desire more, people could believe that it is acceptable to do so without realizing the effects of their unending and destructive desires will make them more frantic in their life.

Next, the most significant findings for the analysis of themes to answer Research Question 2 were the event upon the announcement of antibodies during that time period, the news regarding vaccination to counter the pandemic and linked to cases that were treated during a period when vaccines were not available yet. What the researcher can conclude from this section is that there was no prejudice towards vaccination as it was still in the early stages of the pandemic and there was a lack of information regarding vaccination. It can also be observed that the public was eager and keen for the vaccination as it was the only way to stop the spread of COVID-19 and protect themselves at that time.

In 2021, the major important findings were the vaccination's contribution to putting an end to the outbreak and the unequivocal support of celebrities and high-ranking government officials, which greatly contributed to the case for the COVID-19 vaccination's safety. Besides that, the challenging conditions that the general public had to endure while waiting for vaccinations, death-related incidents were linked to vaccinations, and also related to crime-related events like spreading false information about vaccination and cutting the queue to receive vaccinations were found in the corpus. What is instructive is that it is a common thing for the receivers of the news to frequently forward and distribute them to their loved ones, friends, or even chat rooms. As a result of the rapid dissemination of news without fact-checking, confidentiality and authenticity

are being disregarded. Sharing without taking ownership is insane as it only takes a split second for someone to be persuaded to believe messages, rumors, or even dubious information without even verifying the information. In this case, it is clearly seen in this study how the spread of fake news regarding vaccination has taken a toll on the vaccination programs in Malaysia.

In 2022, the main key findings were related to an event describing the effectiveness of the vaccine which is believed to increase the public's desire to receive the injection as a consequence of the information's accessibility. Additionally, the proliferation of false information, the challenging nature of the vaccination process, the increased demand for the booster shot in favor of maintaining the immunization status in the MySejahtera app, opposition to getting the booster shot due to unfavorable side effects, and numerous other factors were revealed in the corpus. Last but not least, these issues contrast with those in vulnerable situations, such as the homeless, who were not contacted or given access to the vaccination.

5.3 Implications

This research enriches literature concerning the viewpoint of Malaysians in the newspaper by analyzing the language used in the context towards vaccination in three distinct periods. It also fills a gap in the research by collecting data over a longer period of time to see how views on vaccination change with time.

This research is significant because it enables us to get a broad sense of how the general population feels about certain issues and in this case the issues of vaccination in Malaysia. The analysis expected to reveal findings that assist the necessary parties in developing and implementing more suitable and flexible vaccination regulations, for example the freedom to choose the brand of vaccine, the access for the vaccination and

how to foster public trust and participation which may result in more empathetic and engaged responses from the public.

Not only will these findings assist in identifying significant problems such as the healthcare system being inaccessible to the homeless, implementing preventative measures to increase people's confidence in vaccinations, involving public figures and leaders to promote the vaccination program, they will also be a crucial element in battling the COVID-19 pandemic.

Lastly, this research demonstrates the effectiveness of DNVA as the theoretical framework as it enriches the current literature on media coverage with a specific linguistic emphasis. Considering the main area of interest in this work is how the language is used to express sentiment, linguists can utilize DNVA to investigate how speech in news texts creates news values and consider DNVA in any linguistic analysis of news texts as recommended by Bednarek (2016).

5.4 Limitations

The approach can provide a list of themes from three distinct time data sets for comparative purposes since it can conduct statistical comparisons across corpora. However, the approach adopted in this study has several drawbacks.

The researcher only focuses on a limited time frame (the month of March in 2020, 2021 and 2022) which may not accurately reflect the sentiment of the entire Malaysian blogosphere throughout the year and it only indicates the sentiment during that specific period.

As noted in Section 3.2, one of the limitations of the study is the limited number of news articles pertaining to the research gathered. The insufficient search phrases were

used to find news articles pertaining to the research which may restrict the number of news stories obtained for this study.

In addition, the data utilized in the research is confined to The Star, and other newspaper platforms may have contributed to a wider data set for the investigation. The corpus also solely includes public opinions, with no consideration of actual facts.

Last but not least, the limitation of the result of Sentiment Analysis in NVivo. The researcher is aware that the result of the overall Sentiment Analysis in NVivo has an overlap of polarity in some sentences which one of the sentences in the corpus can be categorized as two different polarities in the software. With that being said, the data is based on the algorithm of the software since these techniques merely examine words without taking into account their semantics.

5.5 Suggestions for further research

In addition to the vaccine debate, one may also look at the COVID-19, standard operating procedure (SOP), herd immunity, quarantine, and other topics. Further research on these topics can make use of the same dataset or additional data from various newspaper platforms. In addition, one could consider using other search terms such as “coronavirus”, “pandemic”, “dose” or “booster” for future research. Moreover, the current study only focused on a specific time frame and an online newspaper platform; future research should look into different time frames and other platforms. A comparative study of the same issue in the two distinct languages is also suggested.

Following that, it is advised to use SA software other than NVivo to explore linguistic components to improve SA research which as a result may open up opportunities for future linguistics research that will benefit SA.

5.6 Conclusion

If the news in 2020 concentrated on how COVID-19 spread over the world, disrupting daily life for most people and overwhelming healthcare providers, the news in 2021 has so far centered on stopping the pandemic through the vaccination program followed by new COVID-19 variants going around in 2022. Vaccine is a word with different connotations. While it may represent hope and health for some, at the same time, it may represent death and uncertainty for others. The current study tried to investigate the sentiment towards vaccination from the standpoint of the general public. Malaysians were concerned about the vaccination program, primarily because it is one of the largest immunization programs implemented by the government in the country's history, and it has become a hot topic in society. As a result, the current study has provided some insights and understanding of the general public's perception of vaccination, as well as the need of identifying serious concerns and implementing preventative actions. In addition, the use of corpus linguistic approaches in the current study has added to the increasing body of SA studies.

REFERENCES

Abdul-Mageed, M., & Diab, M. (2012). AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 3907–3914. http://www.lrecconf.org/proceedings/lrec2012/pdf/1057_Paper.pdf

Alabid, N. N., & Katheeth, Z. D. (2021). Sentiment analysis of Twitter posts related to the COVID-19 vaccines. *Indonesian Journal of Electrical Engineering and Computer Science*, *24*(3), 1727-1734. https://doi.org/10.11591/ijeecs.v24.i3.pp1727-1734

Allington, D., Duffy, B., Wessely, S., Dhavan, N., & Rubin, J. (2021). Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, *51*(10), 1763-1769. https://doi.org/10.1017/S003329172000224X

Amjad, A., Qaiser, S., Anwar, A., & Ali, R. (2021). Analysing Public Sentiments Regarding COVID-19 Vaccines: A Sentiment Analysis Approach. *2021 IEEE International Smart Cities Conference (ISC2)*, 1-7. https://doi.org/10.1109/ISC253183.2021.9562904

Awijen, H., Ben Zaied, Y., & Nguyen, D. K. (2022). Covid-19 vaccination, fear and anxiety: Evidence from Google search trends. *Social Science and Medicine*, *297*, 114820. https://doi.org/10.1016/j.socscimed.2022.114820

Baker, P. (2010). Corpus methods in linguistics. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 93-113). Continuum International Publishing Group.

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge and Kegan Paul Ltd.

Barkur, G., Prabhu, V., & Kamath, G. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*, *51*(2), Article 102089. https://doi.org/10.1016/j.ajp.2020.102089

Bednarek, M. (2016). Investigating evaluation and news values in news items that are shared through social media. *Corpora*, *11*(2), 227-257. https://doi.org/10.3366/cor.2016.0093

Bednarek, M., & Caple, H. (2014). Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond. *Discourse & Society*, *245*(2), 135-158 https://doi.org/10.1177/0957926513516041

Bednarek, M., & Caple, H. (2017). *The Discourse of News Values: How News Organizations Create Newsworthiness*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190653934.001.0001

Biber, D. (2012). Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 159–192). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199544004.013.0008

Bučar, J., Žnidaršič, M., & Povh, J. (2018). Annotated news corpora and a lexicon for sentiment analysis in Slovene. *Language Resources and Evaluation*, *52*(3), 895-919. https://doi.org/10.1007/s10579-018-9413-3

Chen, C., & Liu, R. (2023). Discrepancies in the portrayal of the COVID-19 vaccine in Chinese and US international media outlets: A corpus-based discursive news values analysis. *Global Public Health*, *18*(1), 2201315. https://doi.org/10.1080/17441692.2023.2201315

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5 ed.). SAGE Publications.

Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. *Proceedings of the Asia Pacific finance association annual conference (APFA)*, *35*, 43.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th international conference on World Wide Web*, 519-528. https://doi.org/10.1145/775152.775226

Esuli, A., & Sebastiani, F. (2006). Determining term subjectivity and term orientation for opinion mining. *11th Conference of the European chapter of the association for computational linguistics*, *11*, 193-200. https://aclanthology.org/E06-1025

Fatah, F., Omar, N., & Zamri, N. (2022). "Sinovac safe and halal": The newsworthiness of religion on vaccine reporting in Malaysia. *Search-Journal of Media and Communication Research*, *14*(2), 1-16.

Feizollah, A., Anuar, N. B., Mehdi, R., Firdaus, A., & Sulaiman, A. (2022). Understanding COVID-19 Halal Vaccination Discourse on Facebook and Twitter Using Aspect-Based Sentiment Analysis and Text Emotion Analysis. *International journal of environmental research and public health*, *19*(10), 6269. https://doi.org/10.3390/ijerph19106269

Fiesler, C., & Proferes, N. (2018). “Participant” Perceptions of Twitter Research Ethics. *Social Media + Society*, *4*(1), 2056305118763366. https://doi.org/10.1177/2056305118763366

Gao, H., Zhao, Q., Ning, C., Guo, D., Wu, J., & Li, L. (2022). Does the COVID-19 Vaccine Still Work That "Most of the Confirmed Cases Had Been Vaccinated"? A Content Analysis of Vaccine Effectiveness Discussion on Sina Weibo during the Outbreak of COVID-19 in Nanjing. *International Journal of Environmental Research and Public Health*, *19*(1), 241. https://doi.org/10.3390/ijerph19010241

Ghazali, W. N. W. M., Mohamed, S., Nasir, N. S. M., & Yusoh, M. H. (2020). The coverage of vaccination in the Malay newspapers: An exploratory study. *Asian Journal Of Applied Communication (Ajac)*, *9*(1), 351-366. http://irep.iium.edu.my/id/eprint/82145

Gries, S. T. (2012). Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: Towards more and more fruitful exchanges. In J. Mukherjee & M. Huber (Eds.), *Corpus Linguistics and Variation in English: Theory and description* (pp. 41–63). Brill. https://doi.org/10.1163/9789401207713_006

Hatzivassiloglou, V., & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. *Proceedings of the 18th conference on Computational linguistics*, *1*, 299-305. https://doi.org/10.3115/990820.990864

Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial intelligence review*, *52*(3), 1495-1545. https://doi.org/10.1007/s10462-017-9599-6

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. Routledge. https://doi.org/10.4324/9780203841686

Indurkhya, N., & Damerau, F. J. (2010). *Handbook of Natural Language Processing* (2nd ed.). Chapman and Hall/CRC.

Karami, A., Zhu, M., Goldschmidt, B., Boyajieff, H. R., & Najafabadi, M. M. (2021). COVID-19 Vaccine and Social Media in the U.S.: Exploring Emotions and Discussions on Twitter. *Vaccines (Basel)*, *9*(10), 1059. https://doi.org/10.3390/vaccines9101059

Klimiuk, K., Czoska, A., Biernacka, K., & Balwicki, L. (2021). Vaccine misinformation on social media - topic-based content and sentiment analysis of Polish vaccine-deniers' comments on Facebook. *Human Vaccines & Immunotherapeutics*, *17*(7), 2026-2035. https://doi.org/10.1080/21645515.2020.1850072

Kumar, D., Chandra, R., Mathur, M., Samdariya, S., & Kapoor, N. (2016). Vaccine hesitancy: understanding better to address better. *Israel journal of health policy research*, *5*(1), 1-8. https://doi.org/10.1186/s13584-016-0062-y

Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis. *Journal of medical Internet research*, *23*(5), Article e26953. https://doi.org/10.2196/26953

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial intelligence review*, *54*(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

Liu, B. (2015). *Sentiment Analysis : Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139084789

Liu, S., & Liu, J. (2021). Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. *Vaccine*, *39*(39), 5499-5505. https://doi.org/10.1016/j.vaccine.2021.08.058

Liu, Y., Ma, Q., Liu, H., & Guo, Z. (2022). Public attitudes and influencing factors toward COVID-19 vaccination for adolescents/children: a scoping review. *Public Health*, *205*, 169-181. https://doi.org/10.1016/j.puhe.2022.02.002

Luo, C., Chen, A., Cui, B., & Liao, W. (2021). Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China. *Telematics and informatics*, *65(C)*, Article 101712. https://doi.org/10.1016/j.tele.2021.101712

Malaysia to roll out covid-19 vaccination for kids aged five to 12 from Feb 3. (2022, January 21). *The Straits Times*. https://www.straitstimes.com/asia/se-asia/malaysia-to-roll-out-covid-19-vaccination-for-kids-aged-five-to-12-from-feb-3

Malaysia's vaccination rollout for teens among fastest in the world, says Khairy. (2021, October 22). *The Star*. https://www.thestar.com.my/news/nation/2021/10/22/malaysia039s-vaccination-rollout-for-teens-among-fastest-in-the-world-says-khairy

Mallapaty, S., Callaway, E., Kozlov, M., Ledford, H., Pickrell, J., & Van Noorden, R. (2021). How COVID vaccines shaped 2021 in eight powerful charts. *Nature*, *600*(7890), 580-583. https://doi.org/10.1038/d41586-021-03686-x

Marcec, R., & Likic, R. (2022). Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate Medical Journal*, *98*(1161), 544-550. https://doi.org/10.1136/postgradmedj-2021-140685

McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Routledge.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health*, *14*(10), 1505-1512. https://doi.org/10.1016/j.jiph.2021.08.010

Meyer, C. (2014). Corpus-based and corpus-driven approaches to linguistic analysis: One and the same? In I. Taavitsainen, M. Kytö, C. Claridge, & J. Smith (Eds.), *Developments in English: Expanding Electronic Evidence* (pp. 14-28). Cambridge University Press. doi:10.1017/CBO9781139833882.004

Mohamed, K., Rzymski, P., Islam, M. S., Makuku, R., Mushtaq, A., Khan, A., Ivanovska, M., Makka, S. A., Hashem, F., Marquez, L., Cseprekal, O., Filgueiras, I. S., Fonseca, D. L. M., Mickael, E., Ling, I., Arero, A. G., Cuschieri, S., Minakova, K., Rodriguez-Roman, E., Rezaei, N. (2022). COVID-19 vaccinations: The unknowns, challenges, and hopes. *Journal of Medical Virology*, *94*(4), 1336-1349. https://doi.org/10.1002/jmv.27487

Mohamed, N. A., Solehan, H. M., Mohd Rani, M. D., Ithnin, M., & Che Isahak, C. I. (2021). Knowledge, acceptance and perception on COVID-19 vaccine among Malaysians: A web-based survey. *PLoS One*, *16*(8), Article e0256110. https://doi.org/10.1371/journal.pone.0256110

Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, *53*(4), 675-679. https://doi.org/10.1016/j.dss.2012.05.022

Mountassir, A., Benbrahim, H., & Berrada, I. (2013). Sentiment classification on Arabic corpora: A preliminary cross-study. *Document numérique*, *16*(1), 73-96. https://doi.org/10.3166/DN.16.1.73-96

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. *Proceedings of the 2nd international conference on Knowledge capture*, 70–77. https://doi.org/10.1145/945645.945658

Nasyaya, A., Musdalifah, F. S., & Arianti, Y. (2023). What Indonesian Netizen Says about Covid-19 Vaccination? A Corpus Assisted Discourse Studies of YouTube's Comment. *Policy & Governance Review*, *7*(3), 298-315.

Nezhad, Z. B., & Deihimi, M. A. (2022). Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *16*(1), 102367. https://doi.org/10.1016/j.dsx.2021.102367

Nor, N. F. M., & Zulcafli, A. S. (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. https://doi.org/10.17576/gema-2020-2003-12

Nycyk, M. (2022). Ethical Use of Informant Internet Data: Scholarly Concerns and Conflicts. *Journal of Digital Social Research*, *4*(1), 1-22. https://doi.org/10.33621/jdsr.v4i1.88

Ong, S. Q., Pauzi, M. B. M., & Gan, K. H. (2022). Text Mining and Determinants of Sentiments towards the COVID-19 Vaccine Booster of Twitter Users in Malaysia. *Healthcare (Basel)*, *10*(6). https://doi.org/10.3390/healthcare10060994

Our Story So Far. (2020). https://50.thestar.com.my/ourstory.aspx

Over 36% of children in Malaysia given covid-19 vaccine. (2022, March 26). *The Star*. https://www.thestar.com.my/news/nation/2022/03/26/over-36-of-children-in-malaysia-given-covid-19-vaccine

Padmaja, S., Fatima, S. S., Bandu, S., Kosala, P., & Abhignya, M. (2014). Comparing and evaluating the sentiment on newspaper articles: A preliminary experiment. *2014 Science and Information Conference*, 789-792. https://doi.org/10.1109/SAI.2014.6918276

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, *2*(1–2), 1-135. https://doi.org/10.1561/1500000001

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using Machine Learning Techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79-86. https://doi.org/10.3115/1118693.1118704

Panicheva, P., Cardiff, J., & Rosso, P. (2013). Identifying subjective statements in news titles using a personal sense annotation framework. *Journal of the American Society for Information Science and Technology*, *64*(7), 1411-1422. https://doi.org/10.1002/asi.22841

Park, S., & Suh, Y. K. (2023). A Comprehensive Analysis of COVID-19 Vaccine Discourse by Vaccine Brand on Twitter in Korea: Topic and Sentiment Analysis. *Journal of Medical Internet Research*, *25*, e42623. https://doi.org/10.2196/42623

Patterson, N. J., Paz-Soldan, V. A., Oberhelman, R., Moses, L., Madkour, A., & Miles, T. T. (2022). Exploring perceived risk for COVID-19 and its role in protective behavior and COVID-19 vaccine hesitancy: a qualitative study after the first wave. *BMC Public Health*, *22*(1), 503. https://doi.org/10.1186/s12889-022-12900-y

Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. O'Reilly Media.

Qin, Z. (2017). *A framework and practical implementation for sentiment analysis and aspect exploration* (Publication No. 10670931) [Doctoral dissertation, The University of Manchester]. ProQuest Dissertations and Theses Global.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

Radu, A. (2021, June 8). Malaysia Aims to Speed Up COVID-19 Vaccination Campaign. *The Diplomat*. https://thediplomat.com/2021/06/malaysia-aims-to-speed-up-covid-19-vaccination-campaign

Rayson, P. (2003). WMatrix: *A statistical method and software tool for linguistic analysis through corpus comparison* [Unpublished doctoral dissertation]. Lancaster University.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549. https://doi.org/10.1075/ijcl.13.4.06ray

Rayson, P. (2009). Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. http://ucrel.lancs.ac.uk/wmatrix/

Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004). The UCREL semantic analysis system. *Proceedings of the beyond named entity recognition semantic labelling for NLP tasks workshop.* https://eprints.lancs.ac.uk/id/eprint/1783

Salah, Z., Al-Ghuwairi, A.-R. F., Baarah, A., Aloqaily, A., Qadoumi, B. a., Alhayek, M., & Alhijawi, B. (2019). A systematic review on opinion mining and sentiment analysis in social media. *International Journal of Business Information Systems*, *31*(4), 530-554. https://doi.org/10.1504/IJBIS.2019.101585

Saleh, S. N., McDonald, S. A., Basit, M. A., Kumar, S., Arasaratnam, R. J., Perl, T. M., Lehmann, C. U., & Medford, R. J. (2023). Public perception of COVID-19 vaccines through analysis of Twitter content and users. *Vaccine*, *41*(33), 4844–4853. https://doi.org/10.1016/j.vaccine.2023.06.058

Sattar, N. S., & Arifuzzaman, S. (2021). COVID-19 Vaccination Awareness and Aftermath: Public Sentiment Analysis on Twitter Data and Vaccinated Population Prediction in the USA. *Applied Sciences*, *11*(13), 6128. https://doi.org/10.3390/app11136128

Sivanandam, H. (2020, April 26). The Star Online a reliable go-to source during MCO. *The Star*. https://www.thestar.com.my/news/nation/2020/04/26/the-star-online-a-reliable-go-to-source-during-mco

Sutrave, K., Godasu, R., & Liu, J. (2021). Understanding the Public Sentiment and Discourse on COVID-19 Vaccine. *AMCIS 2021 Proceedings*, 24. https://aisel.aisnet.org/amcis2021/art_intel_sem_tech_intelligent_systems/art_intel_sem_tech_intelligent_systems/24

Syed Alwi, S. A. R., Rafidah, E., Zurraini, A., Juslina, O., Brohi, I. B., & Lukas, S. (2021). A survey on COVID-19 vaccine acceptance and concern among Malaysians. *BMC Public Health, 21*(1) https://doi:10.1186/s12889-021-11071-6

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*(1), 325-347. https://doi.org/10.1146/annurev-linguistics-011415-040518

Taboada, M., & Grieve, J. (2004). Analyzing appraisal automatically. *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07),* 158-161.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267-307. https://doi.org/10.1162/COLI_a_00049

The Star is top circulated English daily. (2016, December 5). *The Star*. https://www.thestar.com.my/news/nation/2016/12/05/the-star-is-top-circulated-english-daily/

Thornbury, S. (2010). What can a corpus tell us about discourse? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 270-287). Routledge. https://doi.org/10.4324/9780203856949

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. John Benjamins Publishing Company. https://doi.org/10.1075/scl.6

Tong, R. M. (2001). An operational system for detecting and tracking opinions in on-line discussion. *Proceedings of SIGIR Workshop on Operational Text Classification.*, *1*(6).

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the Association for Computational Linguistics (ACL)*, 417-424. https://doi.org/10.48550/arXiv.cs/0212032

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information*, *12*(5), 204. https://doi.org/10.3390/info12050204

Welsh, E. (2002). Dealing with Data: Using NVivo in the Qualitative Data Analysis Process. *Forum Qualitative Sozialforschung Forum: Qualitative SocialResearch*, *3*(2). https://doi.org/10.17169/fqs-3.2.865

Wibowo, D. A., & Musdholifah, A. (2021). Sentiments Analysis of Indonesian Tweet About Covid-19 Vaccine Using Support Vector Machine and Fasttext Embedding. *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 184-188. https://doi.org/10.1109/ISRITI54043.2021.9702871

Wiebe, J., & Mihalcea, R. (2006). Word sense and subjectivity. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 1065-1072. https://doi.org/10.3115/1220175.1220309

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational linguistics*, *30*(3), 277-308. https://doi.org/10.1162/0891201041850885

Wong, L. (2008). Data analysis in qualitative research: a brief guide to using nvivo. *Malays Fam Physician*, *3*(1), 14-20. https://www.ncbi.nlm.nih.gov/pubmed/25606106

Wong, L. P., Alias, H., Wong, Y. L., Megat Hashim, M., Lin, Y., & Hu, Z. (2023). COVID-19 Anti-Vaccine Sentiments in Malaysia: Narratives of Comments from Facebook Post. *Vaccines (Basel)*, *11*(4). https://doi.org/10.3390/vaccines11040834

World Health Organization. (2020). *Coronavirus disease (COVID-19) pandemic*. https://www.who.int/europe/emergencies/situations/covid-19

Xuan, H. N. T., Le, A. C., & Nguyen, L. M. (2012). Linguistic Features for Subjectivity Classification. *2012 International Conference on Asian Language Processing (IALP)*, 17-20. https://doi.org/10.1109/IALP.2012.47

Yan, C., Law, M., Nguyen, S., Cheung, J., & Kong, J. (2021). Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on Reddit. *Journal of medical Internet research*, *23*(9), Article e32685. https://doi.org/10.2196/32685

Yang, X., & Sornlertlamvanich, V. (2021). Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis. *2021 International Electronics Symposium (IES)*, 151-155. https://doi.org/10.1109/IES53407.2021.9594036

Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021). An analysis of COVID-19 vaccine sentiments and opinions on Twitter. *International Journal of Infectious Diseases*, *108*, 256-262. https://doi.org/10.1016/j.ijid.2021.05.059

Yousefpour, A., Ibrahim, R., Hamed, H. N. A., & Hajmohammadi, M. S. (2014). A comparative study on sentiment analysis. *Advances in Environmental Biology*, *8*(13), 53-68. https://link.gale.com/apps/doc/A385070617/AONE?u=anon~632382f9&sid=googleScholar&xid=cef102ba