

CHAPTER 1

INTRODUCTION

A. Background of Choosing the Subject

Pureklolon (2016) said that mass media has undergone various transformations and is now divided into three types: print mass media, electronic mass media, and online mass media. Print mass media includes magazines, newspapers, and tabloids, which are in the form of writing printed on sheets of paper. Electronic mass media includes radio, television, and film, which are in the form of audio or visuals. Online mass media can be accessed through sites on the internet. These categories of mass media are defined by differences in physical form, technological mechanisms, and the nature of content they offer, each playing a pivotal role in the evolving landscape of communication.

Mass media that is very flexible and favored in the current context is online mass media. The public can access news by connecting to the internet network only by using the fingertips on the device that they own or just using a PC without having to bother carrying paper copies. Online mass media is not limited to publication schedules and fee distribution, such as print mass media, as well as unlimited broadcast time, such as electronic media. Santana K (2017), "The internet combines the advantages and compensates for the shortcomings of traditional (print) media in journalism

practice. Images, sound, and text news are integrated into one and can be stored in an accessible online database for repeated access."

So, the contemporary preference for online mass media underscores its remarkable flexibility and accessibility. Online mass media has revolutionized the way the public consumes news, enabling easy access through internet-connected devices or personal computers eliminating the need for physical copies. This medium transcends the constraints of traditional print mass media, such as publication schedules and distribution fees. It even surpasses the limitations of electronic media by providing unlimited access and storage capabilities. In essence, online mass media stands as a powerful and unlimited source of information that can be conveniently retrieved anytime and from anywhere, making it an invaluable resource in the modern information age.

The influence of mass media is very strong because society knows the world from mass media (Romli, 2017; Nurhadi, 2017). However, the mass media provides information about the real world promptly and selectively, which will certainly provide an image of the environment or an object that is biased and lame (Sukendar, 2017). Therefore, it is necessary to understand further and study the truth of real reality and news broadcast by the mass media so that the public can receive valid information. Eriyanto (2002) stated, "The media is not what it is portrayed to be; it reports what it is, a mirror of reality. The media, as we see, actually constructs reality in such a way."

Understanding more deeply the news written by the media of various countries based on linguistic corpus becomes more interesting to study. The perception of the concept of the meaning of the most apparent words deserves further in-depth research to get a broader picture of how the media reports a phenomenon.

The concepts and representations of the meaning of the words that are apparent most frequently are very important to understand. In addition, by using corpus linguistics approaches and critical discourse analysis, the discourse framing (positive or negative) of a media will be clearer and can be expressed more deeply.

The integration of the use of corpus linguistics into discourse analysis has been carried out by Paul Baker (2008). Corpus linguistics, which uses a lot of qualitative approaches coupled with descriptive explanations in the discourse analysis tradition, is expected to be able to provide comprehensive analysis results.

In addition, the use of corpus linguistics in discourse analysis can provide at least two benefits. The first is to limit researcher bias (Koller and Mautner, 2004; Baker, 2006). Previously, discourse analysis often received criticism because it was considered that it only selected data that was relevant to the researcher; by using the corpus, the researcher ultimately did not select data as he pleased. The second benefit is strengthening the reliability of representation

depicted by a media, where the appearance of words that are repeated with high frequency can then shape or influence society (Stubbs, 2001).

The media has reported that spyware hit Indonesia, not only in Indonesia but also in other countries. This phenomenon is because the state of Israel is suspected of carrying out spyware attacks on several countries, one of which is Indonesia. News related to spyware is interesting because it involves two countries: suspected perpetrators of spyware and victims of spyware. The news that is written also varies in its alignment, as can be seen from the grammar and external factors that influence it.

In terms of data selection, the researcher took data from various online media sources, such as *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*. This media was chosen because of its ability to report various information, which is very updated, and the number of readers in the three media. The high public interest in reading news in the three media shows a belief, if not a mere preference, towards sources of information and media that report on a phenomenon.

The Jakarta Post was chosen because, according to a report by SimilarWeb (<https://www.similarweb.com/top-websites/indonesia/news-and-media/>), a digital market intelligence company, *The Jakarta Post* is among the most popular news websites in Indonesia, with a high number of monthly visits and pageviews. The report for March 2021 states that *The Jakarta Post* ranks second

among the most visited news websites in Indonesia, with 21.7 million visits and 67.6 million page views during that month.

Aljazeera was chosen because according to a report by the Reuters Institute for the Study of Journalism, *Aljazeera* English is the fourth-most-watched international news channel in the UK, behind *BBC World News*, *CNN International*, and *Sky News* 2020. Similarly, a report by the Global Soft Power Index 2021 ranks *Aljazeera* as the fourth-most influential news brand globally, behind *CNN*, *BBC News*, and *The New York Times*. The report notes that *Aljazeera's* reach and impact are particularly strong in the Middle East and Africa.

The Times of Israel was chosen because it has won several prestigious awards, including the B'nai B'rith World Center Award for Journalism Recognizing Excellence in Diaspora Reportage and Commentary, the Simon Rockower Award for Excellence in Jewish Journalism, and the Society of Professional Journalists Sigma Delta Chi Award for Breaking News Coverage. *The Times of Israel* also has a significant readership and impact, with over five million monthly unique visitors to its website and a strong presence on social media platforms. The news organization covers a wide range of topics.

The topic of researching news framing from a country becomes interesting to see alignments and relations between the countries. The research used a corpus approach to provide more details on relations between countries that were not explained in critical discourse analysis. Norman Fairclough's

approach—concordance and collocation corpus- helped in detailing the text dimensions. Then, the background of the country becomes a detailed part of the sociocultural discourse.

B. Problem Formulation

This research aims to answer three main problems, namely:

1. What are the textual features of Indonesian spyware news based on Norman Fairclough's analysis with Corpus Antconc 4.1.4.0?
2. How are the discourse practices of Indonesian spyware news based on Norman Fairclough's analysis?
3. How are the sociocultural practices of Indonesian spyware news based on Norman Fairclough's analysis?

C. Objectives of the Study

Based on the formulation of the problems above, the researcher proposes the following objectives:

1. To analyse the textual features of Indonesian spyware news based on Norman Fairclough's analysis with Corpus Antconc 4.1.4.0.
2. To find out the discourse practice of Indonesian spyware news based on Norman Fairclough's analysis.
3. To analyse the sociocultural practice of Indonesian spyware news based on Norman Fairclough's analysis.

D. Scope of the Study

The study focuses on finding the text dimension of Indonesian spyware news on three reputable online news, such as *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*. The news entitled “Senior Indonesian officials targeted by spyware last year”, “Indonesian officials targeted by spyware: sources”, and “Israeli spyware said used to hack senior Indonesian officials last year” are the data in this study. The data was limited on September 30th, 2022. This study examines the news through Fairclough’s three-dimensional model which consists of text, discourse practice, and sociocultural practice based on corpus analysis with AntConc 4.1.4.0.

E. Review of the Related Studies

This section will review several studies that have similar topics in critical discourse analysis. The studies that the researcher takes are presented below.

First, considering the previous study in this analysis, the researcher studies Kyung Hye Kim, who analysed 'Examining US news media discourses about North Korea: A corpus-based critical discourse analysis' (2014). This article discusses how the media constructs North Korea in US media. Using CNN, Newsweek, and The New York Times objects, this research has many perspectives. The method used is a combination of quantitative and qualitative methods, using the corpus software Word Smith Tools version 5.0 to find collocations and qualitative methods in the use of critical discourse analysis to

see how the media contributes to shaping social construction. Unfortunately, this research only uses a few of the critical discourse analysis methods, so it only focuses on the corpus, especially its collocations.

In addition, they are based on research entitled 'Collocation and Concordance of Presidential Candidates in News Text and Column Comments on Cyber Media (Critical Discourse Analysis Based on Corpus Data)' (2021) by Fadhilah Juwita Lestari. In the process of doing this, this research uses a qualitative approach with critical discourse analysis, which collaborated with the method of linguistic corpus analysis. The data in the study were analysed qualitatively and described descriptively. It discusses the analysis of news text and column comments on cyber media during the 2019 presidential election. The study focuses on the collocation and concordance of presidential candidates, examining the language used in discussions about the candidates and their positions. The researcher uses critical discourse analysis and corpus data to analyse the discourse surrounding the presidential candidates. However, the specific content or results of the study are not provided in the search results.

Then, the article that the title is 'Corpus-based Critical Discourse Analysis of Reporting Practices in English News Reports on Public Health Events in China and the United States' (2023) by Zhuoran Li¹, Ronghui Zhao, and Baocui Lou. This research included a qualitative method with a critical discourse analysis approach collaborating with corpus and using AntConc software 3.3.5. The study also employed corpus linguistic techniques to identify discourse strategies used in news reports and explore the ideological roots

behind the differences in reporting practices between China and the United States.

Research on the field meaning of the word is not something new. A thesis written by Salahshour (2017), for example, has studied this word in the New Zealand daily newspaper using the corpus linguistic method, where the corpus data consists of approximately 700,000 words. According to him, the word immigrant has many colloquialisms with several metaphorical words such as influx, inflow, and wave. Thus, immigrants are likened to a group of people who come to a place (country) massively. However, research using the word immigrant has been widely studied. Also, in this study, the analysis of immigrant representation does not use a larger data set that includes both regional and metropolitan newspapers and is only one-sided.

Also, an article written by Irham (2022) from Sultan Hassanah Bolkih Institute of Education/International and Comparative Education (ICE), University Brunei Darussalam, Brunei, examines words immigrants and foreign workers in the Indonesian media. As a result, the word immigrant does not always have a negative meaning, whereas, in English-language media, the word has a negative meaning. The researcher expresses this depending on the word that follows it. In different places, different collocations of meaning are created. The weakness of this research is that there have been many studies that discuss the words immigrant and foreign worker, and there is also a lack of data studied.

The difference with this research is that this research uses all of

Fairclough's three-dimensional components, and keywords are taken based on the most words from the three news stories analysed using corpus with the software AntConc 4.1.4.0 for analysing the data. The objects are taken based on the latest news. The words studied are also based on the words most used in the news research but are not words that are conjunctions.

F. Method of Research

This method is included in the qualitative research. According to Kumar (2011), qualitative research is a method to understand the situations and perceptions of a group of people. There are ten characteristics of research by Kumar. First, subjectivity is qualitative research, which acknowledges the subjective nature of reality. Second, contextual understanding is qualitative research that seeks to understand phenomena within their natural context. The inductive approach is qualitative research typically uses an inductive approach, where theories and hypotheses emerge from the data itself. Flexibility qualitative research is often flexible and adaptive.

The researcher may adjust their research design data collection or focus based on emerging insights during the study. Rich Description is qualitative research that aims for rich and detailed descriptions of the phenomena under investigation. This matter involves capturing the depth and complexity of human experiences, behaviours, and social interactions. The sixth, holistic perspective is qualitative researchers often take a holistic approach, considering the entire context rather than isolating specific variables. This approach allows

for a more comprehensive understanding of the research topic. Participant Involvement is qualitative researchers often engage directly with participants, interacting with them to gain a deeper understanding of their perspectives. This involvement may take the form of interviews, observations, or participant observations. *Emergent Design* is a qualitative study that may have an emergent design where the research evolves as new insights emerge. This flexibility allows the researcher to explore unexpected avenues and refine their focus during the study. Small Sample Size Qualitative research typically involves smaller sample sizes than quantitative research. The emphasis is on depth and richness of data rather than statistical generalizability. The last is open-ended data analysis, which is qualitative and involves exploring and coding data in an open-ended manner. This process allows for the identification of patterns, themes, and categories that may not have been anticipated at the outset of the study.

So, this research is categorized as qualitative research because the data of this study is observed from written text to explore how the media constructs social media through the news.

1. Data and Source of Data

a. Data

The main data of this research are words, phrases, clauses, and sentences from text of news entitled ‘Senior Indonesian officials targeted by spyware last year’ on *The Jakarta Post* website published on September 30th, 2022; ‘Indonesian

officials targeted by spyware: sources' on the *Aljazeera* website published on September 30th, 2022; and 'Israeli spyware said used to hack senior Indonesian officials last year' on *The Times of Israel* website published on September 30th, 2022.

The data used in this study is secondary data, namely in the form of news text from online mass media, which is then processed using AntConc software. The data in this study is categorized as secondary data because it is not obtained directly by researchers as a first hand. It is in accordance with Rukajat's statement that the research data was divided into two parts, namely primary and secondary data. Primary data is obtained by researchers firsthand, either from interviews or direct observation. At the same time, secondary data is data obtained from the results of literature studies or documentation (Rukajat, 2018).

b. Source of Data

The source of data in the form of news text in this study was taken from the official online mass media website, *The Jakarta Post*, *Aljazeera*, and *The Times of Israel* with the same news related to spyware attacks on Indonesia.

1) *The Jakarta Post*:

<http://www.thejakartapost.com/indonesia/2022/09/30/senior-indonesian-officials-targeted-by-spyware-last-year.html>

2) *Aljazeera*:

<https://www.aljazeera.com/economy/2022/9/30/indonesian-officials-targeted-by-spyware-sources>

3) *The Times of Israel*:

<https://www.timesofisrael.com/israeli-spyware-said-used-to-hack-senior-indonesian-officials-last-year/>

In this study, data was collected using literature study techniques. Technique Literature study, according to Zed (2004), has four main characteristics. First, the researcher deals directly with the text, not the data directly from the field. Second, data is ready-to-use in the sense that researchers are not going anywhere. The third is that data is a secondary source or obtained second-hand. It is not the result of direct observation in the field. The fourth is that the condition of library data is not limited in space and time.

2. Method of Collecting Data and the Data Classification in Table

a. Collecting data steps by Corpus of AntConc 4.1.4.0 by Laurence Anthony

Based on Marlin (2018) the steps of collecting data are as follow:

- 1) Searching the news: In the first step of data collection, the researcher used script observation by searching for news articles about spyware that attack in Indonesia on various news websites. The researcher looked for relevant news articles and selected one from *The Jakarta Post*, *Aljazeera*, and *The Times of Israel* website for analysis.

- 2) Downloading the news: After the news was discovered, the researcher downloaded it.
- 3) Reading the news: the researcher read the news from *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*.
- 4) Finding the data: After reading the news, the researcher attempted to find the data within the script of the news article by *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*.
- 5) Categorizing the data: The researcher then applied Corpus with a software AntConc 4.1.4.0 by Laurence Anthony to classify the data. The data was categorized into two types: collocation and concordance.

The concordance as follow: The high frequency of words and Key Word in Context (KWIC) or concordance.

Type	POS	Headword	Rank	Freq	Range	NormFreq	NormRange
The			1	142	1	69099.757	1.000
And			2	65	1	31630.170	1.000
To			3	63	1	30656.934	1.000
Of			4	44	1	21411.192	1.000
In			5	37	1	18004.866	1.000
That			6	32	1	15571.776	1.000
Indonesian			7	31	1	15085.158	1.000
Officials			8	30	1	14598.540	1.000

A			9	29	1	14111.922	1.000
Were			10	28	1	13625.304	1.000

Then, the collocation from the high frequency of words from the concordance is as follows:

High Frequency Word	Followed Word
..officials	Were
..officials	Attack

b. Collecting Data Steps by CDA of Norman Fairclough

In collecting the data, the researcher applies the observation method (*simak*) by Sudaryanto. Observation method (*simak*) is a method that is done by scrutinizing or observing the use of language (Sudaryanto, 2015). The basic technique used in observation method (*simak*) is tapping technique (*sadap*), in which the researcher first explores the text that will be used as data. After finding the text that will be used as data, the researcher uses non-participative observation technique (*simak bebas libat cakap*), that is paying attention to the contents of the text by reading carefully. Then, a note-taking technique was used to classify the data, such as words, phrases, clauses, and sentences. The steps of collecting the data can be described as follow:

- 1) Visiting the official news site which based on Indonesia and other countries, namely *The Jakarta Post*, *Aljazeera*, and *The Times of*

Israel.

- 2) Selecting the news text on *The Jakarta Post*, *Aljazeera*, and *The Times of Israel* regarding the issue spyware that hit Indonesia.
- 3) Reading the news text which have been selected to gather the data.
- 4) Classifying the data with different codes adapted by previous studies of Nursagi (2020), then researcher designates code ‘A’ as data from *The Jakarta Post*, ‘B’ as data from *Aljazeera*, and ‘C’ as data from *The Times of Israel*.
- 5) Arranging the data including words, phrases, clauses, and sentences into the table in order from *The Jakarta Post*, *Aljazeera*, and *The Times of Israel* data.

The following is the examples of data classification from the news text on *The Jakarta Post*:

Representation (Text Dimension)				
Observed Elements	Media	Data	Analysis	Code
The representation of person, group, activity, or action	<i>The Jakarta Post</i>	More than a dozen senior Indonesian government and military officials were targeted last year with spy software designed by an Israeli surveillance firm, according to nine people	The headline in the media <i>The Jakarta Post</i> stated who the victim was, when, and who was the perpetrator of the spyware.	6.A.1

		with knowledge of the matter.		
--	--	----------------------------------	--	--

Table 1. Data Classification

Notes on data code:

(table number).(news source code).(data sequence number)

*Code A = Data from *The Jakarta Post*

*Code B = Data from *Aljazeera*

*Code C = Data from *The Times of Israel*

3. Method and Technique of Analyzing Data

1) Data Analysis Steps

After going through the data collection process, the next step is processing data. In analysing the data, this study applies the identity method (*padan*) by Sudaryanto. The identity method (*padan*) determines the identity of the determining lingual unit whose determinants are outside the language, apart from the language, and are not part of the language in question (Sudaryanto, 2015). The type of identity method uses in this study is the referential identity method. According to Sudaryanto (2015), the referential identity method is a method where the determining tool is a fact outside the language designated by the language, such as noun is a thing; verb is action. The steps are as follows:

- a) Reading the data source which is a news text.
- b) Arranging the data into vocabulary features, grammatical aspects, representation, relation, and identity to gain understanding how *The Jakarta post*, *Aljazeera*, and *The Times of Israel* display discourse in text dimension.
- c) Examining the text production, the consumption, and distribution to perceive the ideology built by *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*.
- d) Categorizing the situational, institutional, and social under sociocultural practice to realize the social context correlated to discourse in *The Jakarta Post*, *Aljazeera*, and *The Times of Israel*.
- e) Concluding the data analysed.

G. Significance of the Study

The researcher hopes the result of this research will be useful, both theoretically and practically. Theoretically, the results of this study are expected to contribute references related to language studies, especially critical discourse analysis corpus data. In addition, the results of this study are also expected to improve and strengthen the role of language in assisting analysis in the field of social sciences.

This research will provide additional options for conducting discourse analysis research. Language research and discourse analysis do not always

necessarily have to be done using qualitative methods. Numbers and statistics can also help the researcher analyse the discourse.

Practically, this research is expected to make people aware of the construction of reality built by the mass media so that society can be more clever in filtering news from the media.

This research will assist researchers and students in obtaining references for the use of the English language corpus, which is available and accessible free of charge. Thus, they will not encounter many obstacles in studying linguistics, especially for researchers of comparative language, diachronic, and synchronic linguistics.

H. Presentation

This undergraduate thesis is presented in four chapters. Chapter one consists of the background of the study, problem formulations, objectives of the study, scope of the study, review of the related studies, significance of the study, method of the research, and presentation. Chapter two consists of the theoretical approach, framework, and construction. Chapter three consists of the data analysis. Last, chapter four consists of a conclusion and suggestions.