

CHAPTER I INTRODUCTION

A. BACKGROUND

The 2016 U.S. presidential election is one of the most infamous and controversial political events in American history, in which both Republican presidential candidate Donald Trump and Democratic presidential candidate Hillary Clinton fiercely competed, obtaining voters from 50 states to claim the presidency in the following year. They typically promote their candidate on traditional media such as TV, radio, billboards, and newspapers, at a higher cost. Politicians have found cost-effective ways to promote their candidates to appeal to broader potential voters by promoting them via digital media such as social media. However, this led to disastrous consequences, as the election was bombarded by the slew of misinformation known as “fake news”, a term that Trump frequently uses.

Fake news, or hoaxes, consists of story reports based on real-life or fictional events using false and provocative narratives with a lack of sources and citations. They initially emerged as rumors or conspiracy theories through verbal communication and evolved with the introduction of physical media such as newspapers and magazines. News readers can mostly recognize the news as false without having to analyze sources thoroughly; thereby, it can be used for satirical intentions. Some news outlets also explicitly disclose to readers that their stories are for satirical intentions and should not be taken seriously.

When digital media such as Facebook, Twitter, Youtube, Google, etc., was emerged to the public, hoaxes have quickly evolved into the form of news articles.

Fake news is usually published by an individual or a group, but cannot specify the true identity, as they can publish the story under an alias, given the online anonymity nature. These fake news stories contain rumors or conspiracy theories from either the perpetrators or unreliable sources, as well as mimic the journalistic attributes and standards found in mainstream news sites, manipulating the readers into thinking the story shown in the article is genuine.

While the effects on the readers' point of view or critical thinking is generally small. However, fake news has a more significant impact on voters' perceptions, primarily in the political sphere and the 2016 U.S. election. The perpetrators from both political parties, the Democrats and the Republicans wrote news stories based on events in which the candidates were involved, be it controversial or common, with added false information or sources created or altered by them, as well as added certain provocative words or sentences to emphasize the urgency and manipulate the news readers, both the voters or the oppositions. The impact can cause real-life outcomes as some fake news stories are used to incite hate speech online and create societal conflicts against certain groups of races, ethnic, or religions with violence.

Despite fake news sharing some similarities to genuine news at first glance, it presents distinct linguistic features that drive the false stories to the general public. Hence, some fake news have linguistic errors which can be easily point out and do not meet the journalism standards as opposed to genuine news. In contrast to fact checking method, this level of analysis requires more advanced methods used by experts. According to a study made by Victoria L. Rubin, Y. Chen, Niall J. Conroy (2017), the authors made the table on frequently used linguistic features that can be

found on fake news articles or propaganda; 1) subjectives, 2) superlatives, 3) assertive adverbs, and 4) first-person and second-person pronouns, these features are used to exaggerate certain words and sentences while describing the events in the author's personal point of view. Whereas genuine news articles have assertive words that are used to describe events less vaguely in an objective point of view.

As stated by Teguh Budiharso (2009), linguistic features are elements that consist of sentence construction, grammar, and mechanical aspects of writing within the text or sentence. A text with appropriate linguistic features is utilized by characteristics of language of science. The text conveys statement that is accurate and exact, supporting the logic of the statement. The use of appropriate sentence structure and formal yet comprehensible vocabulary help reader identify propositions in the text more readily.

Linguistic features have six different parts used in physical written text, as classified by Hezili (2010) and T.Hasan (2018); (1) Lexical features (interjection, abbreviation, word combination, and diction), (2) orthographic features (word spelling, and capitalization), (3) grammatical features (ellipsis, passive/active voice, and personal pronouns), (4) discourse features (coherence, relevance, and paragraph structure), and (5) graphic features (images, charts and graphs, infographics, typography, layout, and color schemes).

In the present time, social media has become the staple digital media in society for instant information and fake news is constantly evolving and can be seen almost in any internet platforms, yet still retains many of the core linguistic features and

techniques from the 2016 election fake news stories, making this study relevant for contemporary and future references on fake news detection.

B. IDENTIFICATION OF PROBLEMS

Based on the statement shown above, the writer of this study has found the following issues:

1. Most common analysis methods used by non-experts have difficulties in identifying the linguistic features in fake news stories.
2. The semantic level between genuine and fake news articles is almost identical at first glance.

Although most fake news stories have linguistic feature errors that can be easily identified, some have successfully replicated the journalistic standards and language found in genuine news stories. This poses a challenge for identification, as even people relying on common methods may find it more difficult to distinguish between genuine and fake news in such cases, even with the most advanced identification system used in social media platforms.

C. SCOPE OF PROBLEM

This study focuses on the scope of specific issues mentioned above:

1. The sole focus of this study is to semantically analyze and manually extract linguistic features of the text within sentences of fake news articles found

on social media and other platforms in five different linguistic features, proposed by Hezili (2010) and F. Hasan (2018).

2. The writer oversees the distinct linguistic differences between fake news stories and genuine news stories.

D. RESEARCH QUESTIONS

1. What are the types of linguistic features used in the selected fake news articles?
2. What types of words, phrases, or sentences in fake news articles have an effect on public opinions?
3. How are the linguistic features in fake news articles different from those of genuine news?

E. OBJECTIVE OF THE STUDY

The main objectives of this study can be seen in the following;

1. To disclose the linguistic features of fake news,
2. To describe the types of words, phrases, or sentences in fake news articles,
3. To explain the linguistic differences between fake and genuine news articles.

F. SIGNIFICANCE OF THE STUDY

The writer is determined that this study will have both theoretical and practical significance for the respective parties:

1. THEORETICAL SIGNIFICANCE

This study presents a comprehensive examination of the linguistic features that characterize fake news, particularly in the context of the 2016 U.S. Presidential Election. The research conducts a semantic analysis of said issue and gives a broader understanding of how language is used to deceive, manipulate, and influence public opinion. This theoretical framework is crucial for the field of linguistics and communication studies, as it helps to differentiate the apparent boundaries between factual and misleading information based on linguistic features.

2. PRACTICAL SIGNIFICANCE

This study highlights the importance of media literacy in identifying linguistic features identification. Educators and media literacy programs can develop more effective detection tools and strategies for the general public to learn how to critically analyze news sources independently. This can encourage them to fully understand the linguistic differences between genuine and fake news content, therefore lessening the spread of online misinformation.

CHAPTER II

LITERATURE REVIEW

A. PREVIOUS STUDIES

A number of researchers have conducted studies about linguistic structures or any linguistic features related to fake news with different analysis methods, both human-based and computer-based. The studies were taken from six journal articles by the researchers as followed; S. Gard (2022), H. Allcott and M. Gentzkow (2017), L. Soetekouw and S. Angelopoulos (2022), M. Mahyoob, J. Algaraady, M. Alrahaili (2021), D.P. Panagiotis Kasseropoulos and C. Tjortjis (2021), A. Ari Iswara and K. Agus Bisena (2020), and one article by M. Lilleslåtten (2022).

1. S. Garg (2022) - Linguistic features-based framework for automatic fake news detection

The research explores the use of artificial intelligence in machine-learning algorithm from various social media platforms to identify the fake news' data and its pattern for timely decision-making as digital media grows and misleading content spreads extensively for years, which humans cannot manage the large volume of data received in its database. With its rapid advancements, AI can be used in both ethical and unethical ways. For instance, fact-checking websites like factcheck.org and PolitiFact verify news articles' authenticity and give accurate sources if needed, and a dedicated news outlet like MBFC News concerns on educating the public about media bias and deceitful news methods. However, previous studies suggest that many machine-learning models are equipped with a different set of textual and visual

detection features. Which led to a challenging task for researchers and people to identify fake news solely through the analysis of the news content (Lancaster et al., 2018, Jin et al., 2016).

Fake news in general is a common term since the introduction of mass media in previous centuries, it is created to intentionally deceive people and affect their opinion with misinformed content in the form of text, image, and video. High novelty and emotional appeal drive them forward the news (Kumari et al., 2022). Beyond from detecting fake news, this issue is also related lack of critical thinking and trust among people. The main reason why people get easily attracted to fake news is how it is written with appealing headlines and fake content. They are less concerned regarding sources or the authenticity of the news. There are approximately 56% of people are concerned about whether news outlets cover factual information or not, according to a news report by digital news report in 2020. The term was overlooked by the public until it gained attention during the 2016 US presidential election, in which fake news publishes six times faster than real news on social media, as shown by a UN report. At that time, fake news is created for various intentions, primarily defaming celebrities and changing public opinion for political gain, and simple entertainment. It also creates a huge impact on the country with the rise of bullying and violence, as well as fake news video propagation, which many experts have created their own methods of identifying fake news. Choi and Ko (2022) used domain knowledge and multi-modal data fusion for fake video identification, Sahoo and Gupta (2021) created a browser add-on for

early fake news. Another issue while detecting fake news is the uneven arrival of responses and the dynamic evolution of crowd opinions. Li et al. (2022) used progressive detection based on Kalman Filter-based approach to deal with this issue. Thus, people need to verify the credibility of a person sharing news on social media (Alrubaian et al., 2016).

S. Gard (2022) and the team conducted some research in which they examine different linguistic features that are effectively identify fake news, including complexity features, readability index, psycholinguistic features, and stylometric features. He then further chooses 26 important features and applies them to different machine learning models for implementation. Three different techniques are used to extract these features such as frequency-inverse document frequency (tf-idf), count vectorizer (CV), and hash-vectorizer (HV). The researcher then tested these models with different training dataset sizes and compared their accuracies. These four existing datasets (Reuter, Buzzfeed, The Random Political, and Mc_Intire) were used for the experiment, the proposed framework achieved higher accuracies with 90.8% in the Reuter dataset, while the Buzzfeed dataset obtained the highest accuracy of 90%. The Random Political dataset with an accuracy of 93.8%, and the Mc_Intire dataset with 86.9% accuracy.

2. H. Allcott and M. Gentzkow (2017) - Social Media and Fake News in the 2016 Election

The research, conducted by the two researchers, Hunt Allcott and Matthew Gentzkow (2017), proposed the large influence of social media as a sole source of political news and information, they defined fake news to be news articles that provide false information to anyone that favors their bias over the accuracy of the information it conveys. In the cases of the elections, multiple fabricated news articles are widely shared, one of which is from the now-defunct website denverguardian.com with the headline, “FBI agent suspected in Hillary email leaks found dead in an apparent murder-suicide.” In addition, fake news can be broken into several sub-types which can be misunderstood as factual; conspiracy theories, typically originated by people who perceive them as true, and difficult to verify the authenticity of the sources; satirical articles that are unlikely to be misconstrued as factual; unintentional reporting mistakes, such as a recent incorrect report where Donald Trump had removed a Martin Luther King Jr. statue from the Oval Office.

The researchers found that fake news has made a wide impact on voters’ political perceptions during the election. According to recent evidence, 62% of U.S. adults get news on social media (Gottfried and Shearer 2016), and they shared the most popular fake news on Facebook than the most popular mainstream news stories on a wider scale (Silverman 2016). Most people believe fake news stories (Silverman and Singer-Vine 2016) that tended to favor Donald Trump over Hillary Clinton (Silverman 2016). In addition,

numerous correspondents stated that Trump's winning was heavily influenced by fake news (Parkinson 2016; Read 2016; Dewey 2016).

Therefore, their findings have led to several reasons why social media platforms and the advertising networks are the sole political information nest for fake news. 1) Any fake news producers can enter the market and produce news at half the cost compared to traditional media, increases the relative profitability of the small-scale, short-term strategies often adopted by, and reduces the relative importance of the long-term reputation for quality; 2) The "bite-size" information format of social media makes identify article's sources much more difficult; 3) Facebook friend networks are ideologically segregated based on the respective ideological affiliation in their profile. Additionally, social media platforms such as Facebook and Google have faced some pressure from both consumers and civil society to reduce the severe spread of fake news that violate misleading content policies on their systems (Wingfield, Isaac, and Benner 2016). Facebook has made additional content moderation step on their system to identify fake news articles, by flagging false articles as "disputed by third party fact-checkers," showing less potentially false articles in users' news feeds, and notifying users that the story they shared is flagged by third party fact checkers (Mosser 2016).

Allcott and Gentzkow (2017) conducted a research on the American people's consumption of fake news before and after the election. They then collect the data from various sources including web browsing, a 1,200-person post-election online survey, and 156 false election-related news stories verified

by leading fact-checking websites in the three months before the elections. The survey concluded that only 14% of American adults considered social media as their “most important” source of election-related news, despite the low trust rate in information accessed through social media compared to traditional media. Furthermore, most fake news stories were widely shared and heavily leaned into Donald Trump, with 115 fake pro-Trump stories shared over 30 million times on Facebook, as opposed 41 fake pro-Clinton stories shared 7.6 million times.

3. L. Soetekouw and S. Angelopoulos (2022) - Digital Resilience Through Training Protocols: Learning to Identify Fake News on social media

The two researchers suggested that everyone of all ages, roles of skepticism, and levels of education can have a positive effect on their ability to recognize fake news on every social media platform with proper online experiments and training protocols. Thereby, numerous studies point out that humans lack the ability to recognize characteristics of fake news (e.g., Bond & DePaulo, 2006; Rubin et al., 2016). They achieved with the score of 54% on tasks in which they need to distinguish truth and deception without further training or tools, only slightly better than chance. However, there is no clear solution to counter the effect of fake news (Au et al., 2021), and with the limited use of certain tools to do so (Paredes et al., 2021).

Social media has become the epicenter of fake news, as general population heavily uses social media as their main information source (Lazer et al., 2018).

And with the rapid evolution of technological advancement such as deep fake, providing education about the impact of misinformation and tools to identify fake news is challenging for experts. The content is presented through algorithms based on the user's preferences. They are trained and built with biases inevitably by humans involved, despite they are primarily driven by data in the neutral side (Gillespie, 2014), let alone their inner processes are much complex and difficult to understand (Carlson, 2018). This grows filter bubbles and echo chamber from the increasing exposure of personalized content (Borges & Gambarato, 2019), which can lead to the reinforcements of existing beliefs and to intellectual isolation.

While fake news come in various forms and more complex aspects, there are two categories to counter the negative effect of fake news, one is detection and intervention on platforms, the other is empowering people. The first category involves the use of algorithms as an automated fake news counter, social media platforms like Facebook use their algorithm to enhance consumer engagement, ensuring users are exposed to quality content (Lazer et al., 2018), primarily diverse political content as opposed to merely content confirming their existing beliefs. This reduces the effect of echo chambers, a phenomenon caused by and reinforcing the polarized political opinions (Borges & Gambarato, 2019).

The second category involves training social media users and expert knowledge in countering fake news. Expert knowledge has been used for several decades (Fridkin et al., 2015). For instance, Facebook released a tutorial

on how to identify fake news (Brady et al., 2017). Furthermore, efforts to find the truth behind fake news stories have been made by fact-checkers (Hameleers & van der Meer, 2020). Fact-checking conducted by experts seems to have risen as a response to growing misinformation in politics (Fridkin et al., 2015). Recent studies show fact-checkers can potentially reduce polarization and help dealing with partisan identities (Hameleers & van der Meer, 2020) and that they affect people's evaluation of political messages' accuracy (Fridkin et al., 2015).

4. M. Mahyoob, J. Algaraady, M, Alrahaili (2021) - Linguistic-Based Detection of Fake News in Social Media

The study, conducted by the three researchers, investigated the linguistic features of fake news and systematic nuances between it and factual news in 16 linguistic attributes under three main linguistic features categories such as lexical, grammatical and syntactic features.

In recent years, fake news has been widely dispersed in social media platforms causing real-world effect both on users and news outlets, as stated by Allcott and Gentzkow (2017). It was highlighted during the 2016 U.S. presidential election campaign as a serious threat to journalism, democracy, freedom of expression, and the public's trust in governments. Thus, identifying fake content on social media is highly required for social and political fields. However, doing so is much challenging as it manages to replicate those of genuine news articles and deceive the people who do not authenticate for the

reliability of the contents and sources. Furthermore, with the limited tools require careful fact-checking and evidence collection.

The researchers obtained 20 different genuine news articles from PolitiFact and 20 randomly chosen fake news articles from social media platform such as Facebook to study their authenticities from three different perspectives; 1) writing style, 2) how fake news spread, and 3) how people participate in fake news and the role they can play in all these perspectives (Zafarani et al., 2019; Zhou & Zafarani, 2018). They analyze its writing style in 16 different linguistic attributes, including personal pronoun, proper pronoun, adverb, stative verb, toinfinitive, passive voice, reported speech, comparative adjectives, superlative adjectives, modal verbs, quotes, conjunctions, long sentences, interrogative, and negation. The researchers then compare the final data both fake and genuine news in two different datasets. The results revealed that reported speech, passive voice, negation are the top shared features used in the datasets, which these features are used less in genuine news articles. In contrast, authentic news uses proper nouns more than the other set.

5. D.P. Panagiotis Kasseropoulos and C. Tjortjis (2021) - An Approach Utilizing Linguistic Features for Fake News Detection

The two researchers explored the various advanced machine learning (ML) algorithms and to perform linguistic analysis on the text within the false content. They chose a style-based machine learning approach that primarily

rely on the textual information from news, in which the algorithms manually extract the lexical features and part of speech counts, combine Named Entity Recognition (NER) with the Frequent Pattern Growth (FP Growth) algorithm. The researchers then evaluate the performance of several ML algorithms to identify the best-performing linguistic features. The research revealed that convolutional neural networks (CNNs) could achieve more detection accuracy, despite the content differences between fake and genuine news is limited.

In the 21st century, known as the digital media, social media as a whole have become a staple for our everyday lives and drastically changed how we interact with other people, as high-speed internet and website deployment becoming accessible to everyone over the years. However, these benefits emerged misinformation and its rapid spread on the internet, which commonly known as, fake news.

The researchers used a dataset of news articles, extract and select appropriate features for fake news detection. Lexical features consist of character and word-level signals, while language features consist of sentence-level syntax and readability metrics. Syntactic features encompass parts-of-speech tagging and frequencies of function words. Domain-specific linguistic features pertain to news-specific elements like quoted words and external links. The results concluded that first-person singular pronouns are more common in genuine news, while fake news uses more third-person pronouns and modal verbs, reflecting uncertainty. Despite these apparent patterns, the overlap between fake and real news complicates differentiation. Support Vector

Machines (SVMs) achieved higher accuracy with text embeddings than with lexical features alone. The combination of NER and FP-Growth highlighted the structural similarity between fake and real news, with CNNs outperforming other algorithms in accuracy.

6. M. Lilleslåtten (2022) - Linguistic cues could be key to exposing fake news

The article provided an overview the collaborative effort between the linguists and computer scientists from University of Oslo in developing tools to identify any linguistic cues within the news articles in various languages, including Norwegian, English, and Russian. This research was called the Fakespeak project, led by Silje Susanne Alvestad.

The emergence of fake news was occurred during and after the 2016 U.S. presidential election, in which Russian preparators were involved through fake news. Public awareness of fake news has increased, due to the prominent political figures, such as Donald Trump, were known for disseminating false information. These two periods saw a surge the need for more effective tools and methods to detect fake news and understand how it created and published.

The linguists in this research analyze the texts for specific linguistic features, including verb tenses, pronouns, metaphors, and emotional language. These findings are then given to computer scientists, and integrate these features into AI models designed to identify fake news, resulting faster and reliable AI fake news detection tools. Although most of the data the researchers

obtained was in English, this project was faced with challenges for researchers to obtain another in Russian and Norwegian, hence they relied on fact-checking services to build their datasets. The goal of the project is to create faster and more reliable fake news detection, which potentially benefits various languages and improves global media literacy.

7. A. Ari Iswara and K. Agus Bisena (2020) - Manipulation and Persuasion Through Language Features in Fake News

The two reseachers of the study explored how the linguistic features are used in fake news as a primary tool to alter facts and deceive readers. This research used qualitative and descriptive methods to analyze the data they obtained from a popular local fact-checking in Indonesia, turnbackhoax.id, and determine the effects of linguistic features and their provocative influence.

With the wide spread of digital media to the general public, the online news portal has made information easily accessible and shareable, particularly through social media. However, this ease of access has also increased the spread of fake news. turnbackhoax.id cites alarming statistics published by Indonesian authorities, indicating the widespread presence of fake news websites and content on social media. Therefore, fake news detection is much needed.

The analysis revealed that fake news authors frequently use various linguistic features, such as acronyms and initialisms, word reductions, numbers and letters as substitutes for words, unconventional punctuation, and images.

According to a documented example by turnbackhoax.id, it highlights the use of unconventional spelling and punctuation to create a misleading narrative relating to the geographical location in Indonesia. They also found that fake news uses assertive, expressive, and directive speech acts to enhance its persuasive power. Assertive speech acts are used to present false information as factual, thereby linking the speaker to the purported truth of the proposition. Commissive speech acts involve promises or commitments intended to manipulate the reader's trust. Expressive speech acts convey emotions to engage the reader's sentiments, while directive speech acts include commands or suggestions to influence the reader's actions.

These studies mentioned above, highlight their own advanced analysis methods and the process of identifying the linguistic features of fake news;

1. S. Gard (2022) used artificial intelligence to identify different linguistic features in fake news, and the results later are sent to various training datasets.
2. H. Allcott and M. Gentzkow (2017) studied how social media is the sole influence in spreading fake news, held a post-election online survey to the Americans, and collected the results.
3. L. Soetekouw and S. Angelopoulos (2022) have found that everyone with different aspects can have a positive effect on their fake news observation skills by giving them proper online training and training protocols, as well

as studying how social media companies like Facebook provided a guide to their users in how to encountering fake news stories in the platform.

4. Mahyoob, J. Algaraady, M, Alrahaili (2021) analyzed the linguistic features of fake news and its systematic nuances with factual news in 16 linguistic attributes under three key features such as lexical, grammatical, and syntactic features, as well as three perspectives such as its writing style, how fake news spread, and how people participate in it. The three researchers concluded that three of the 16 linguistic attributes of fake news are the top shared features used in the datasets.
5. D.P. Kasseropoulos and C. Tjortjis (2021) explored the various advanced machine learning (ML) algorithms, and the researchers picked a style-based ML approach to perform linguistic analysis on the text within the false content. It highlighted that convolutional neural networks (CNNs) higher fake news detection accuracy.
6. M. Lilleslåtten (2022) overviews the collaborative project, known as the Fakespeak project and led by Silje Susanne Alvestad, in which linguists and computer scientists were involved in creating faster and reliable AI tools for fake news detection. The linguists were tasked to extract the linguistic features data from various fake news articles; the findings were then given to the computer scientists to integrate into the AI tools they developed, resulting in faster and more reliable AI fake news detection tools.

7. A. Ari Iswara and K. Agus Bisena (2020) used qualitative and descriptive analysis to identify how linguistic features in fake news articles are used to alter facts and manipulate readers. The research highlights the data they obtained from a popular and the findings of various linguistic features and speech acts that fake news authors often use.

These seven studies primarily used advanced computer-based analysis methods, which utilized artificial intelligence (AI) and machine learning (ML) models to manage and identify large fake news data from various sources at a faster rate, reducing human efforts and errors and saving research times to generate objective and accurate results. However, these methods require advanced hardware and software expertise to get the optimal database and results, making them less accessible for the general public. The AI and ML models also may struggle detecting sarcasm, satire, and nuanced misinformation within the text, and require regular updates to easily detect the new misinformation tactics. They also may inherit various biases from training data.

In contrast, this study used a human-based analysis method. It is fully accessible for the general public with basic analysis training and can benefit critical thinking and media literacy for potential human researchers by interpreting nuanced misinformation from tone, satire, sarcasm, hidden agendas, and contextual meanings. The researchers can quickly adapt to the new fake news contexts without needing retraining. Although this method requires more time and human labor to manually analyze fake news with a limited human reading capacity. The results may

get mistakes if not properly analyzed and vary based on context and experience, which subject to the researcher's personal, political, or cultural biases.

Both computer-based and human-based analysis method have advantages and disadvantages of their own in certain scenarios and may generate different results, dependent on the amount of data they analyze and context of the fake news stories. Human-based analysis method is suitable for this study due to its small database and research method.

B. THEORETICAL BACKGROUND

1. Semantic Analysis

Semantic analysis is the process of analysis that examines the syntactic structures of text to comprehend the meanings, ranging from words, phrases, clauses, sentences, and paragraphs, as well as the writing of it. According to V. Kanade (2022), the analysis process begins by studying and analyzing the dictionary definitions and meanings of individual words to fully understand their meanings, which refers to as lexical semantics. The words or elements then are parsed with their designated grammatical role, and their structure is processed to reduce any confusion caused by words with multiple meanings.

Semantic analysis has six key elements that are essential in understanding the language; hyponyms (refers a specific lexical entity having a relationship with a more generic verbal entity), meronymy (refers to the arrangement of words and text that denote a minor component of something), polysemy (refers to a word with multiple meanings), synonyms (refers to words with similar

meanings), antonyms (refers to words with opposite meanings), and homonyms (refers to words with the same spelling and pronunciation, yet with different meanings).

There are two distinct techniques that can be used for analysis, these refer to text classification and text extractor;

1. Semantic classification, implies text classification wherein predefined categories are assigned to the text for faster task completion, with various types such as topic classification (classifies text into preset categories based on the content type), sentiment analysis (detects positive, negative, and neutral emotional tone in the text), and intent classification (classifies the intent of the text presented).
2. Semantic extraction, a type of semantic analysis technique that extracts specific data from the text. It includes types such as keyword extraction (identifies relevant terms and expressions within the text and gives insights when combined with the above classification techniques), and entity extraction (identifies and extract entities within the text, such as names of individuals, organizations, places, and others).

This type of analysis method, particularly semantic extraction, was chosen to examine certain words, phrases, or sentence structures found in fake news articles that are considered provocative and vulgar, and they then will be extracted from the article and used as the main data of the analysis for different parts of linguistic features respectively.

2. Linguistic Features

Linguistic features are the various elements and characteristics of language designed to construct and convey meaning, which refer to the use of sentence construction, grammar, and mechanical aspects of it. In general, Linguistic features can be broken down into three large constructs: lexical, syntactic, and cohesion.

The linguistic features can be broken down into five different parts, according to F. Hasan (2018);

1. lexical features (interjection, abbreviation, word combinations, and diction),
2. orthographic features (word spelling and capitalization),
3. grammatical features (ellipsis, passive/active voice, and personal pronoun),
4. discourse features (coherence, relevance, and paragraph structure),
5. graphic features (images, charts and graphs, infographics, typography, layout and color schemes).

This particular subject was chosen for this study, as the writer suggested that there are distinctive linguistic features used in news articles. The sole reason why they are essential is to present the information in news articles well-articulated and comprehensible, while avoiding confusion among readers. For instance, one of the commonly used linguistic features in written language is paragraph structure, a well-organized structure can make a complex story readable and easy for readers. This also applies to news

articles, which it typically starts with the introduction, providing a summary of the story/problem, it jumps into the body of the article, presenting a detailed coverage of the story/problem, while also giving solutions if needed, and the article ends with the conclusion, presenting a detailed summary and aftermath of the story/problem.

The writer added that the linguistic features can be misuse for perpetrators to create fake news stories, and mimic the key visual elements to make news articles look genuine and credible, deceiving the readers into thinking information they present is factual. However, the write believed that fake news articles may use some distinctive linguistic features compared to genuine news articles, which can pinpoint them from certain words, phrases, sentences, and other visual elements such as images, charts and graphs, infographics, typography, layout and color schemes. For instance, a fake news article mimics the visual elements of those mainstream news articles, but poor grammar and paragraph structure organization may hint the readers it is false if they thoroughly read the entire story.

3. Fake News

Fake news is story reports based on current events that contain false or misleading information from vague sources. These news stories are usually published on the internet or other digital media such as social media platforms, the story they present can be used in a satirical or malicious intent. According to M. Mahyoob, J. Algaraady, and M, Alrahaili (2021), they stated that fake

news is a fictitious article that is deliberately fabricated to deceive readers and create psychological warfare. The term itself was emerged at the start of the printing press, but reappeared in the OxDford Dictionary in 2017. There are various types of fake news, as Balmas (2014) suggested it refers to satire news, which in its primary intention is for entertainment purposes.

The difference between fake and genuine news, as Americas High School (2024) stated, is that fake news persuades the readers to hold one particular ideological viewpoint based on the misinformation it presents, which can be used as propaganda. Whereas genuine news presents an objective and unbiased information or report about actual events to the readers from reliable, authoritative sources.

Fake news was chosen as the main subject of this study, as it is the main force that became 2016 U.S. presidential election the most controversial political event in American history. It was used to provoke the supporters' beliefs and views, as well as demeaning the opposition.