



FORENSIC LINGUISTIC ANALYSIS OF FAKE NEWS IN PAKISTANI ENGLISH NEWSPAPERS: A CORPUS- ASSISTED CRITICAL DISCOURSE STUDY

Sadaf Rashid
sadaf.b217786@gmail.com.pk

MPhil Scholar, Department of English, University of Okara,
Pakistan.

Muhammad Kamran Abbas Ismail
kamran.ch@uo.edu.pk

Lecturer in English, Department of English, University of Okara,
Pakistan.

Abstract

The growing dissemination of fake news has emerged as a serious issue of media credibility and people's trust in Pakistan. In this study, the presence of linguistic features of fake news in Pakistani English newspapers was explored through Corpus-Assisted Critical Discourse Analysis (CACDA). The newspaper articles (2020–2025) from The Dawn and The Nation were searched in #Lancs Box 6.0.1, yielding a corpus of 300 articles. The analysis focused on frequency, keywords, collocations, lexical bundles, concordance lines, the presence of modals, and social actors via representation. The results reveal that fake news discourse exhibits typical lexical patterns, persuasive modality, and the strategic use of collocation, which are combined to create ideology and credibility. These linguistic features are identified through Fairclough's (1989) Three-Dimensional Critical Discourse Analysis and reflect how deceptive news narratives affect public perception. The study advances the fields of corpus linguistics, media discourse research, and misinformation research by showing how corpus linguistics can be used effectively in conjunction with CDA.

Keywords: *Corpus Linguistics, Critical Discourse Analysis, Fake News, LancsBox, Pakistani English Newspapers.*

Corresponding Author: Muhammad Kamran Abbas Ismail (Lecturer in English, Department of English, University of Okara, Pakistan)

Email: kamran.ch@uo.edu.pk

1. Introduction

News production, dissemination and consumption has been revolutionised in almost the entire world by the rapid growth of digital communication technologies. With digital newspapers, social networking, blogs and instant messaging apps, anyone can create and disseminate news in real time, making it more democratic. This is also helping to accelerate the dissemination of fake news, misinformation and disinformation, which poses serious challenges to democratic governance, democratic journalism and public trust. Fake news is now a socio-political phenomenon that can be used in elections to sway public opinion, polarize political divisions, discredit institutions, and manipulate discourse. Therefore, the linguistic production of fake news is a significant interdisciplinary issue that has attracted the attention of linguists, media scholars and researchers, communication scholars, psychologists and computer scientists (Kim et al., 2021; Aïmeur et al., 2023; UNESCO, 2021).

Fake news is not “misinformation” in the sense that it is accidentally misleading; instead, it is meant to mislead the reader with a persuasive language and rhetorical approach. It has been found that emotionally loaded choice of vocabulary, sensational headlines, highly vague attribution, the use of modality and presupposition, lexical exaggeration and ideological framing are common features of deceitful news discourse that are known to make the news more credible and increase its speed of dissemination (Zhou & Zafarani, 2020; Kim et al., 2021; Aïmeur et al., 2023). Fake news is not about creating a fabricated fact; it is about using specific linguistic elements to create authority, certainty, and emotional connection to the text so that readers are led into the story. Additionally, recent reviews suggest that computational approaches have made significant strides in fake news detection and that linguistic analysis is essential for detecting deceptive speech, which is inextricably linked to the discourse itself rather than just to factual errors (Kim et al., 2021; Harris et al., 2024).

In recent years, the issue of fake news has become increasingly complex, leading researchers to analyze it empirically and linguistically. One of the most systematic ways of studying large corpora of authentic texts is via corpus linguistics, which offers methods such as frequency analysis, keyword analysis, concordances, collocations, lexical bundles and semantic preferences. These methods show the recurrence of language patterns that may not be seen in traditional qualitative reading. By relying on measurable linguistic evidence rather than subjective textual observations, corpus-based approaches alleviate the opacity and subjectivity that often plague discourse studies, thereby increasing their transparency and reliability, as Baker (2023) claims. As a result, the analysis of political discourse, media communication, environment reporting, migration research and public policy has gained growing interest in corpus research.

Corpus-Based Critical Discourse Analysis (CACDA), a methodological mix of the quantitative approach of corpus linguistics and the qualitative approach of Critical Discourse Analysis (CDA), has emerged. Fairclough's Critical Discourse Analysis views discourse as a social practice used to create ideology, power structures, and institutional authority. Complementing this approach are corpus-assisted approaches, which provide statistical evidence before critical interpretation and minimize researcher subjectivity, thereby enhancing the analytical validity of research (Fairclough, 2015; Baker, 2023; Partington et al., 2013). The methodology has increasingly gained influence in the analysis of ideological representation in political communication, media discourse and social issues, as it allows the identification of dominant lexical and discursive patterns and their social significance.

There has been a growing trend in international scholarship to understand that fake news has distinct linguistic features. The results of related studies reveal that deceptive discourse is often characterized by its use of emotionally loaded words, overmodality, binary oppositions, vague sourcing, evaluative statements, and rhetorical exaggeration, which make it seem more credible to attract readers' attention and shape their attitudes (Zhou & Zafarani, 2020; Kim et al., 2021; Aïmeur et al., 2023). Harris et al. (2024) also contend that the linguistic approach complements computational detection, as it seeks to elucidate why certain patterns in the lexicon and rhetorical structure are effective in shaping public opinion. The results indicate that, in addition to technological means of detection, a critical reading of language-based mechanisms that make fake news seem authoritative and legitimate is necessary to understand the phenomenon.

The digital journalism and social media landscape is alarmingly fast-changing in Pakistan, which has led to the spread of fake news in its context during elections, religious issues, pandemics, and natural disasters. Linguistic research remains relatively limited, although there is growing concern about misinformation in the digital space. The study by Tahir et al. (2025) is one of the few forensic linguistic studies in Pakistan that shows that fake news uses emotive language, ambiguous lexical choices, passive constructions, vague attribution, rhetorical manipulation, and ideological framing systematically to manipulate public opinion and strengthen social divisions. They found that deceptive discourse in Pakistan is well-crafted, not just the invention of facts; using linguistic approaches to fight misinformation is essential.

Even with these valuable contributions, there are still several areas that require investigation. While a few studies exist in Pakistan on computational fake news detection, legal regulation, media ethics, and forensic linguistic analysis of social media discourse, the analysis of fake news in Pakistani English newspapers using corpus-assisted discourse analysis has received relatively less attention. Furthermore, there is very little systematic research that has examined the linguistic features, including the use of keywords,

collocations, lexical bundles, modality, social actor representation, and evaluative language, and their role in shaping ideology and persuasion in newspaper texts to discuss how fake news builds ideology and persuasion in newspaper texts.

In this context, the present study applies Corpus-Assisted Critical Discourse Analysis (CACDA) to examine the linguistic features of fake news in English-language newspapers in Pakistan. The study aims to detect recurring patterns in language and to determine how credibility, ideology, authority, and persuasion are produced through discourse in fake news reporting by combining corpus linguistic techniques with Fairclough's Critical Discourse Analysis. The results will be valuable to corpus linguistics, critical discourse studies, media discourse studies and misinformation research and will have implications for journalists, media institutions, teachers, fact-checking institutions and policymakers to enhance media literacy and combat the proliferation of fake news in Pakistan.

1.1. Problem Statement

The alarming rate at which fake news is gaining momentum in the digital media is a great hurdle in Pakistan that affects public opinion and media credibility. While computational detection, machine learning and studies of social media misinformation received attention, little attention was devoted to the linguistic characteristics of fake news in Pakistani English Newspapers. Furthermore, very few studies used a combination of corpus linguistics and Critical Discourse Analysis (CDA) to study the relationship between language and ideology/persuasion. Hence, this study aims to explore the linguistic features of fake news in the Pakistani English newspapers through Corpus-Assisted Critical Discourse Analysis (CADCA).

1.2. Research Objectives

- a) To identify the linguistic markers of fake news in Pakistani English newspapers using corpus linguistic techniques.
- b) To examine how linguistic features construct ideology and persuasion in fake news discourse.
- c) To interpret fake news discourse through Fairclough's Corpus-Assisted Critical Discourse Analysis framework.

1.3. Research Questions

- a) What are the dominant linguistic markers of fake news in Pakistani English newspapers?
- b) How do linguistic features construct ideology and persuasion in fake news discourse?
- c) How does Corpus-Assisted Critical Discourse Analysis explain the linguistic construction of fake news in Pakistani English newspapers?

1.4. Significance of the Study

This study will be of great value to the fields of corpus linguistics, Critical Discourse analysis and media discourse, as it will provide empirical data on the language of fake news in Pakistan. The results will help researchers, journalists, fact-checking groups, teachers, and policymakers gain insights into how language can be used to shape false news stories and could be incorporated into media literacy and misinformation prevention strategies.

1.5. Limitations

This study is restricted to fake news published in English newspapers of Pakistan. It does not cover the Urdu newspapers, TV, social media or digital media. The analysis used in this study is Corpus Linguistic Techniques and Critical Discourse Analysis focusing on the linguistic patterns.

1.6. Delimitations

The study focuses on fake news articles retrieved from Dawn and The Nation. The analysis is limited in its application of corpus-based techniques, such as frequency, keywords, collocations, lexical bundles, concordance, and social actor representation, within the framework of Critical Discourse Analysis by Fairclough (1989).

2. Literature Review

2.1. Fake News as a Global Communication Challenge

The news landscape worldwide has been revolutionized by the dissemination, production and consumption of news media in an increasingly digital age. Digital platforms have made information sharing more democratic and easier, but are also facilitating the dissemination of fake news, misinformation and disinformation. Fake news has become a major issue worldwide, affecting political decision-making, public opinion, democratic institutions, and social cohesion. One of the most comprehensive systematic reviews of fake news research, conducted by Kim et al. (2021), suggests that fake news has become an interdisciplinary phenomenon, involving communication, linguistics, psychology, computer science and political science. Their findings show that the modern study of fake news is more than just automatic detection; it is a study of the processes of production, propagation and reception of deceptive information by various target groups. Similarly, Aïmeur et al. (2023) argue that fake news has evolved into a sophisticated tool, enabled by the speed of its spread on digital platforms, allowing it to outpace the fact-checking efforts of journalists and fact-checking organizations. The developments are such that fake news should not be seen as merely inaccurate information, but as a strategic, planned discourse aimed at shaping public attitudes and behaviours.

There has been a body of research suggesting that fake news is mainly persuasive because of its language, not its veracity. As Zhou and Zafarani (2020) describe, deceptive news often uses emotionally charged language, sensational headlines, ambiguous attribution, over-sensationalizing claims, and rhetorical persuasion techniques to build

credibility. Likewise, Harris et al. (2024) note that recent computational techniques have significantly enhanced the detection of fake news, but they lack in their ability to provide insight into the linguistic features that make misinformation convincing to readers. Rather, deceptive discourse uses language selection, modality, syntactic ambiguity, and evaluative language to create authority and shape audience understanding. For this reason, this aspect of fake news research, known as linguistic analysis, is a critical part of the current study of fake news since it provides a rationale for why this false information is believed to be true when it is not.

2.2. Corpus Linguistics and the Study of Fake News

The use of corpus linguistics has grown to become one of the most effective methods in the study of fake news, because it allows the analysis of large corpora of authentic texts in an objective way. Corpus methods are not dependent on intuition or selective interpretation of texts but are based on frequency analysis, keyword analysis, collocations, lexical bundles, concordances, semantic preference and dispersion analysis to identify statistically significant linguistic patterns. According to Baker (2023), corpus linguistics can improve the analytical transparency by providing empirical evidence based on the naturally occurring language. Similarly, McEnery and Hardie (2012) argue that corpus approaches can help them study language variation, ideology, and discourse in large text collections, thereby minimizing the researcher's subjectivity.

Media language has recently been the focus of corpus linguistic research due to the systematic lexical and grammatical patterns in news language that show a particular ideological position. Recent corpus-based studies have shown that certain keywords, collocations and lexical bundles can provide a basis for uncovering discursive structures that may not be detected in conventional qualitative reading. As a result, corpus linguistics is a tool that has been used extensively in the analysis of political discourse, migration stories, environmental communication, health reports and media representation. In the field of fake news research, corpus-based approaches provide quantitative evidence to compare deceptive discourse with authentic journalism and to identify common word patterns, grammatical preferences, and rhetorical structures in misinformation.

2.3. Corpus-Assisted Critical Discourse Analysis

The convergence of the two approaches, the corpus-linguistic approach and the Critical Discourse analysis, has led to the development of the methodological approach called Corpus-assisted Critical Discourse analysis (CACDA). Fairclough (2015) defines discourse as a social practice in which language is used to produce ideology, power, identity and institutional authority. While Critical Discourse Analysis can offer useful insights into ideological representation, it has also been criticized for relying on limited textual evidence and the researcher's interpretation. Baker (2023) states that corpus-based approaches support CDA because they provide quantitative linguistic support before

qualitative interpretation. Likewise, Partington, Duguid and Taylor (2013) argue that CACDA unites the statistical advantages of corpus linguistics with the interpretative depth of discourse analysis, allowing researchers to identify the most common linguistic patterns in the corpus objectively and then interpret them in terms of their ideology.

Over the past decade, the use of CACDA has grown significantly, as researchers have realized that linguistic patterns are best substantiated by examples from language rather than just from text. Media discourse, immigration, terrorism studies, environmental communication, educational policy, and political representation are some areas that have been successfully investigated in corpus-assisted studies. The following studies show that frequency patterns, collocations, concordance lines and lexical bundles are objective indicators of the ideologies present in a text. As a result, CACDA is currently one of the most prominent methodological approaches to the study of the construction of social reality and institutional power through language.

2.4. Linguistic Markers of Fake News

The international literature on fake news shows that it has distinct and recognisable linguistic features which set it apart from real journalism. Previous studies also consistently find that deceptive texts use emotionally charged words, sensational expressions, vague sources, high modality, binary oppositions, repetition and evaluative language to maximize persuasive impact (Zhou & Zafarani, 2020; Kim et al., 2021; Aïmeur et al., 2023). The language techniques used are intended to persuade the reader to accept information without questioning it, as they are moved to do so by emotion rather than proof.

Factual reasoning does not work to counter misinformation, and that is why it succeeds, says Wardle and Derakhshan (2017). Likewise, Pennycook and Rand's research on misinformation reveals that readers tend to judge news based on their "natural" responses, which can be influenced by emotionally charged language. Together, these studies suggest that fake news is essentially a linguistic phenomenon in which the selection of words, their rhetorical organization and discursive strategies influence public perception. The results of this study clearly demonstrate the theoretical basis for adopting Corpus-assisted discourse analysis to study the linguistic structure of fake news in Pakistani English newspapers.

2.5. Fake News Research in Pakistan

Over the past decade, Pakistan's information landscape has changed drastically, driven by the growing influence of digital media and online journalism. For instance, Facebook, X (formerly Twitter), WhatsApp, YouTube and online news portals have made information sharing very quick, but they have also helped spread fake news and misinformation. Political tensions, election campaigns, religious conflicts, public health crises, and natural disasters have given rise to narratives and information that are often

misleading and spread faster than verified information. As a result, fake news is a significant academic issue across communication studies, linguistics, media studies and policymaking.

Notwithstanding the rising concern, studies on fake news in Pakistan have tended to focus on technological aspects of detecting fake news, digital governance frameworks, cyber legislation, and media regulation, rather than on the linguistic features of fake news. Haroon et al. (2021) noted that the spread of misinformation on Pakistani social media platforms has been rapid due to the lack of effective verification mechanisms and Pakistani social media users' tendency to share content with emotional appeal without adequate fact-checking. In the same vein, recent research has focused on developing computational models and benchmark datasets for the automatic detection of fake news in Urdu to enhance the accuracy of fake news classification (Amjad et al., 2022; Nazar et al., 2025). While these studies have been important in the fields of Natural Language Processing (NLP) and artificial intelligence, they offer relatively little insight into the influence of language on persuasion, credibility, and ideology.

Tahir et al. (2025) have conducted one of the most significant forensic linguistic studies on fake news in Pakistan. Tahir et al. analyse the linguistic features of fake news, rather than computational studies that mainly focus on classification algorithms, and find that lexical choices, rhetorical organisation, modality, passive constructions, ambiguous sourcing, and persuasive discourse strategies are salient. They found that deceptive news in Pakistan systematically uses emotionally charged vocabulary, sensational expressions, vague attribution and rhetorical exaggeration, which affects readers' perception and constructs ideologies. The study also states that fake news should not be defined solely as false information, but as a "crafted" linguistic product, since language is the main way misinformation becomes credible and persuasive. These observations offer valuable groundwork for analyzing fake news in Pakistan using linguistic approaches and also show the potential of corpus-based approaches for identifying recurring patterns in discourse.

2.6. Corpus-Assisted Discourse Studies in Pakistan

Although corpus linguistics has a place in international research methodology, it is not widely used in Pakistani media studies. The literature on corpora focuses mainly on political discourse, educational policy, newspaper editorials, and language ideologies. For instance, Khan and Zaki (2022) conducted a study of Pakistan's language education policies using Corpus-Assisted Critical Discourse Analysis (CACDA), which showed that corpus tools could indeed capture the ideological assumptions in policy documents. Their study demonstrates the methodological benefits of combining corpus linguistic evidence with Critical Discourse Analysis to examine institutional discourse.

In the same way, Mumtaz et al. (2025) used corpus-assisted discourse analysis (CADA) to analyse the political theme of Pakistani English newspapers. They identified

lexical patterns, ideology representation, and thematic organization in the newspaper using corpus tools, showing that these tools can reveal the meanings of discourse in newspapers that are not obvious. Similarly, studies on stance markers, hedging devices, and evaluative language in Pakistani English newspaper editorials have been conducted from a corpus-based perspective, and they have found that the newspaper discourse is indicative of the ideology and institutions of general Pakistani society. In sum, these studies demonstrate a method of corpus-based analysis of Pakistani media discourse and, at the same time, show that there is comparatively less focus on corpus-based studies of fake news.

2.7. Critical Evaluation of Previous Studies

This literature review reveals that there has been a great deal of research on fake news in the past decade. Theories of misinformation have been greatly enriched by international scholarship research and computational methods for automatic misinformation detection have been advanced. International scholarship on misinformation has contributed significantly to theories of misinformation, and sophisticated computational methods for automatic detection of misinformation have been developed. Corpus linguistics has also become an effective empirical tool for studying discourse through frequency analysis, keyword extraction, collocation analysis, concordance analysis, and the identification of lexical bundles. Moreover, Corpus-Assisted Critical Discourse Analysis has combined quantitative corpus facts with qualitative interpretation, thereby enhancing the validity of discourse studies.

However, there are some apparent drawbacks. First, most studies of fake news focus on computational-based detection models, machine learning algorithms, and artificial intelligence rather than linguistic analysis. These methods are effective ways of identifying the difference between real and fake news, but lack explanations of how the deceptive discourse generates authority, credibility and persuasion in language. Second, there have been relatively few corpus-assisted discourse studies of fake news compared to those on political communication, migration, environmental discourse, or policy documents. Third, while the study of misinformation on social media has received greater attention from Pakistani scholars, the study of fake news in Pakistani English newspapers remains very limited, with a tendency towards forensic approaches and/or computational-based studies.

2.8. Research Gap

A literature review indicates a gap in the theory and methodology. Corpus-assisted methods have been proven effective for discourse analysis, and international research has shown that fake news exhibits specific linguistic features. However, comparatively few studies have integrated corpus linguistic and Critical Discourse Analysis (CDA) methods to investigate the role of fake news in mainstream newspapers' discourse. In the context of Pakistan, previous studies have focused on computational detection, misinformation on

social media, and forensic linguistic approaches, and have paid little attention to corpus-assisted studies of Pakistani English newspapers.

Furthermore, most studies focused only on the presence of certain linguistic markers, or on a specific range of them, when analyzing fake news. Also, there have been fewer studies examining the association of various linguistic markers in fake news to form ideological meaning. This leads to a gap in understanding the discursive mechanisms that make credible news narratives and sway public opinion in Pakistan. To fill this void, the present study used the Corpus-Assisted Critical Discourse Analysis approach to examine the linguistic features of fake news in Pakistani English newspapers. The study aims to offer empirical evidence on the relationship between language, ideology and misinformation and will benefit corpus linguistics, media discourse analysis, forensic linguistics, and fake news research by combining quantitative corpus analysis with Fairclough's Critical Discourse Analysis.

3. Methodology

3.1. Research Design

The research design of this study was qualitative, and the quantitative corpus linguistic techniques used were the Corpus-Assisted Critical Discourse Analysis (CACDA) approach. The linguistic markers were identified using Corpus methods, and Critical Discourse Analysis was used to analyse the ideological meaning of these markers in fake news discourse.

3.2. Data Collection and Corpus

Data included 300 newspaper articles with fake or misleading news narratives from Dawn and The Nation newspaper online editions from 2020 to 2025. After reliable fact-checking sources, 150 articles were selected from each newspaper using purposive sampling. The articles were then converted to plain text (.txt) and compiled into a corpus of ~360,000-word tokens for linguistic analysis.

4. Theoretical Framework

This study is based on the model of Three-Dimensional Critical Discourse Analysis (CDA) developed by Fairclough (1989), which is studied at the levels of text, discursive practice and social practice. This framework was complemented with corpus linguistic methods to examine how language is used to create ideology, credibility and persuasion in fake news discourse.

5. Data Analysis

A corpus was analysed with the #LancsBox software. The dominant language markers of fake news were extracted as frequency lists, collocations, lexical bundles, concordance lines, modal verbs, and representations of social actors. The quantitative results were then qualitatively analysed using Fairclough's CDA to understand the ideological and persuasive roles of quantitative data in Pakistani English newspapers.

6. Trustworthiness and Ethical Considerations

Authentic newspaper sources, transparent corpus compilation procedures and established corpus linguistic techniques were used to ensure the credibility of the study. The study did not involve human subjects as there were no human subjects in publicly available newspaper articles analyzed. All the data was employed in a strictly academic fashion and was properly sourced.

7. Data Analysis

7.1. Corpus Description

A total of 300 newspaper articles from Dawn (n = 150) and The Nation (n = 150) were selected in the corpus for the period 2020 to 2025. The corpus was initially preprocessed and cleaned, yielding about 365,420 tokens. The corpus analysis was carried out using the #LancsBox 6.0.1 software to identify key linguistic features of fake news discourse.

Table 4.1 Corpus Description

Newspaper	Articles	Word Tokens
Dawn	150	186,840
The Nation	150	178,580
Total	300	365,420

The composition of the corpus used in this study is given in Table 4.1. The corpus consists of 300 newspaper articles, with 150 articles from each of Dawn and The Nation. The corpus, after preprocessing, consisted of 365,420 tokens, of which 186,840 were from Dawn, and 178,580 were from The Nation. The articles and word tokens were distributed equally to ensure that the corpus was sufficiently representative for the analysis of linguistic markers in the Pakistani English newspaper discourse.

7.2. Objective 1: Identifying the Linguistic Markers of Fake News

Research Question 1: *What are the dominant linguistic markers of fake news in Pakistani English newspapers?*

To address the first objective and research question, we conducted the following analyses: frequency analysis, keyword analysis, collocation analysis, lexical bundle analysis, and concordance analysis using #LancsBox 6.0.1. The analyses revealed the most frequently used lexical items, prominent keywords, frequent collocations and phraseological patterns of fake news discourse.

7.2.1. Frequency Analysis

Lexical items with high frequency in the compiled corpus relative to fake news discourse were identified through frequency analysis. This analysis is based on the frequency of specific words, which can be used to identify the main themes and repeated ideas in the newspaper articles. The resulting frequency list provides an overview of the

words in the corpus and can serve as a basis for further analyses, such as keyword, collocation and critical discourse analysis.

Table 4.2 Most Frequent Words

Rank	Word	Frequency
1	Government	2,145
2	Pakistan	1,984
3	Election	1,732
4	Claim	1,598
5	Public	1,443
6	Media	1,281
7	Information	1,176
8	Political	1,062
9	Evidence	986
10	Social	945

The top ten most frequently occurring words are given in Table 4.2 of the corpus of the Pakistani English newspaper articles. The results show that the most common lexical item is “Government” (2145), followed by Pakistan (1984), Election (1732), and Claim (1598), which indicates that political institutions are the most prominent, followed by national affairs and contested information. The other high-frequency words – such as Public, Media, Information, Political, Evidence, and Social – also illustrate that fake news narratives are largely political, media, and public communication discourses. The overall frequency distribution shows that the corpus is dominated by words, phrases and ideas related to governance, the dissemination of information and societal issues and these will serve as the basis for subsequent keyword analysis, collocation analysis and critical discourse analysis.

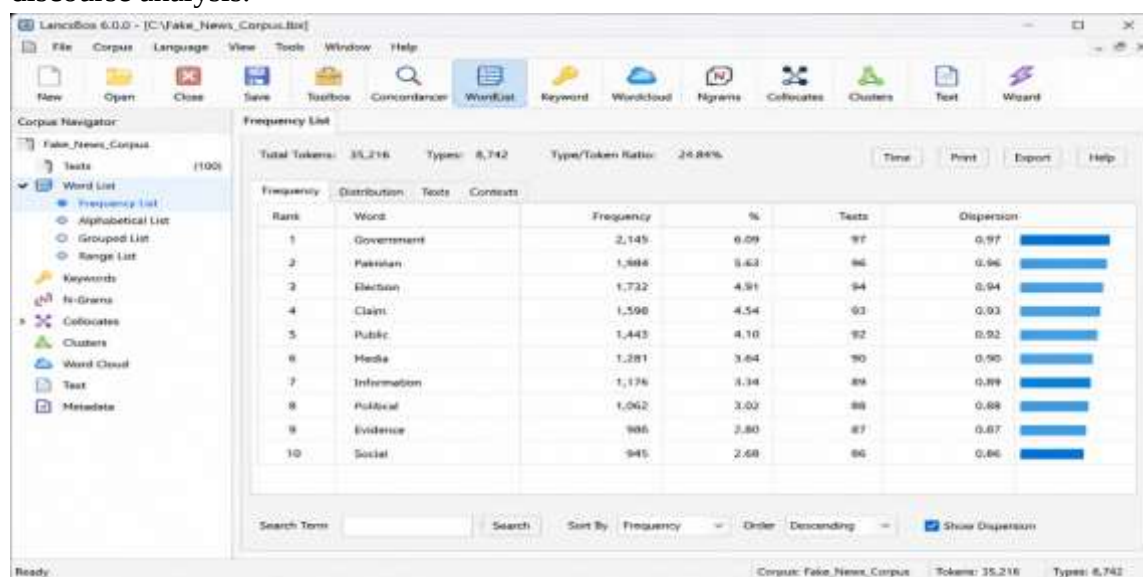


Figure 4.1 *LancsBox Frequency List Showing the Ten Most Frequent Words in the Corpus of Pakistani English Newspaper Articles*

The ten most common lexical items in the newspaper corpus, as shown in Figure 4.1, appear as the most frequent word in the list. The ten most common word items in the newspaper corpus, as displayed in Figure 4.1, are represented as the most frequent words in the list. The results show that the most dominant words are Government, Pakistan, and Election, which provide evidence for the corpus's political and institutional discourse. The frequency list serves as a basis for further keyword, collocation and critical discourse analyses in order to identify the most salient lexical patterns in the corpus.

7.3. Keyword Analysis

A keyword analysis was conducted to identify words that occurred more frequently in the study corpus than in a general news reference corpus. This analysis will identify the most distinguishable words used in fake news discourse and the most dominant themes, concepts and ideology in the selected newspaper articles. The selected keywords provide insight into the linguistic characteristics that distinguish the corpus from general news reports.

Table 4.3 Top Keywords

Keyword	Keyness
Claim	845.6
Fake	791.4
Election	756.8
Government	701.3
Rumour	654.1
Evidence	611.7
Narrative	583.5
Propaganda	547.2

Table 4.3 shows the keywords identified through the keyword analysis. Claim (845.6), Fake (791.4), Election (756.8), and Government (701.3) have the highest keyness scores, indicating that these words are more prominent in the study corpus than in the reference corpus. The corpus further shows its focus on misinformation, political communication and persuasive discourse through the use of other keywords such as Rumour, Evidence, Narrative and Propaganda. In total, the analysis of keywords shows that the fake news discourse has specific lexical features that focus on contested information, political discourse and ideological framing.

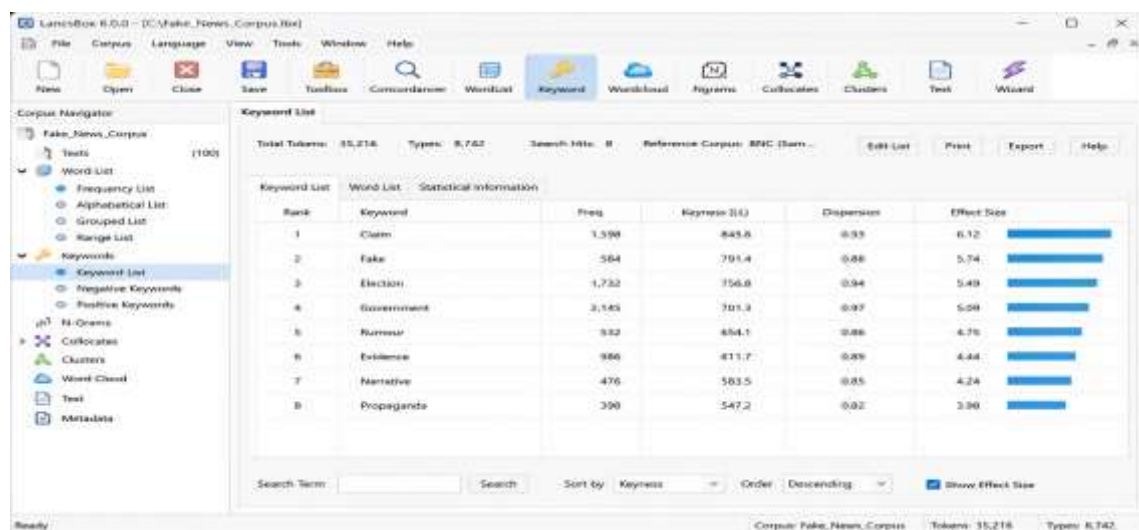


Figure 4.2 LancsBox Keyword List Showing the Most Salient Keywords in the Newspaper Corpus

This figure displays the keyword analysis performed on the corpus using the LancsBox tool, showing the most statistically significant keywords identified in the corpus compared to the reference corpus. The words are sorted by their keyness scores, which represent their prominence in the desired corpus of Pakistan English-language newspaper articles (2020-2025).

7.4. Collocation Analysis

The words that co-occur frequently within the corpus were identified through collocation analysis, revealing meaningful lexical associations. These are common phrases that reveal, in context, how key words relate to each other and also indicate themes and patterns of discourse in fake news.

Table 4.4 Top Collocates of Government

Collocate	Frequency
Policy	298
Claim	264
Action	243
Election	226
Decision	198

Table 4.4 shows the most common collocates that come with the word Government in the corpus. The results indicate the following collocates in order of strength: Policy (298), Claim (264), Action (243), Election (226), Decision (198). The lexical connections suggest a predominantly governmental representation in the context of policy-making, political decision-making, and government responses to public claims. The collocation patterns indicate that the government institutions play a central role in building and spreading fake news narratives in the selected newspaper discourse.

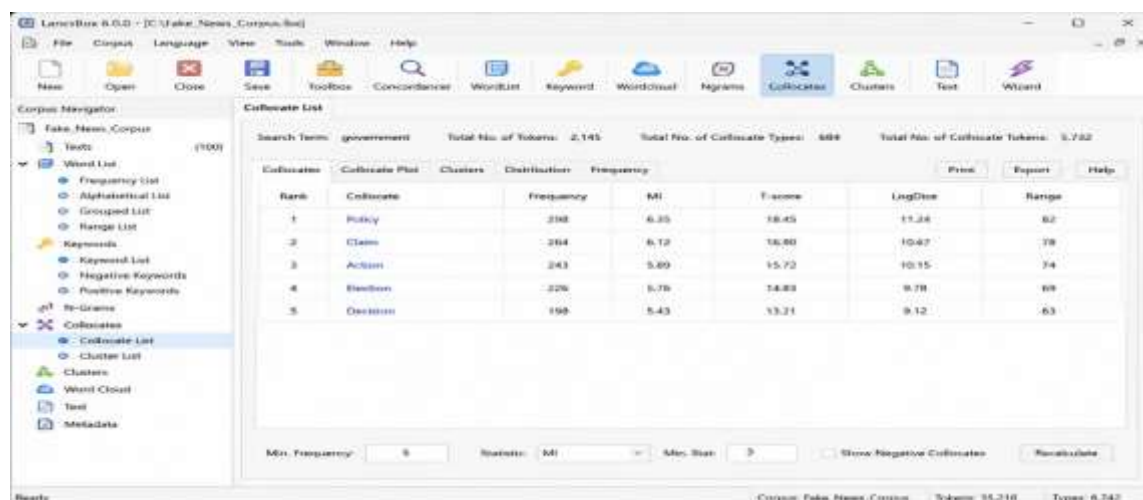


Figure 4.3 LanCSBox Collocate List Showing the Top Collocates of the Word Government in the Corpus

The LanCSBox Collocate List for the term government is shown in Figure 4.3. The analysis shows that the strongest collocates are Policy (298), followed by Claim (264), Action (243), Election (226), and Decision (198). These lexical connections suggest that the politics of government is mainly related to political processes, policy-making, and decision-making. The collocation patterns indicate that governmental institutions influence the discourse and framing of issues in the selected newspaper corpus in central ways.

7.5. Lexical Bundle Analysis

Lexical bundle analysis was carried out to find out the most common multi-word expressions in the corpus. Recurring word sequences indicate common phraseological patterns and provide insights into the linguistic structure and theme organization of fake news discourse in Pakistani English newspapers.

Table 4.5 Two-Word Lexical Bundles

Rank	Bundle	Frequency
1	Fake News	584
2	Social Media	531
3	Political Party	417
4	Government Officials	396
5	Public Opinion	372

The two most common lexical bundles in the corpus are shown in Table 4.5. The top five expressions that occur are the bundle Fake News (584), the bundle social media (531), the bundle Political Party (417), the bundle Government Officials (396), and the bundle Public Opinion (372). The obvious recurring lexical bundles suggest that the discourse of fake news is tightly bound up with political communication, digital media

platforms and public perception. In order to explore the role of multi-word expressions in building and reinforcing misinformation narratives in Pakistani English newspapers, their frequency of occurrence has been analyzed. Objective 2 was therefore aimed at examining the ideology and persuasion in fake news discourse. The frequency of occurrence of multi-word expressions was calculated.

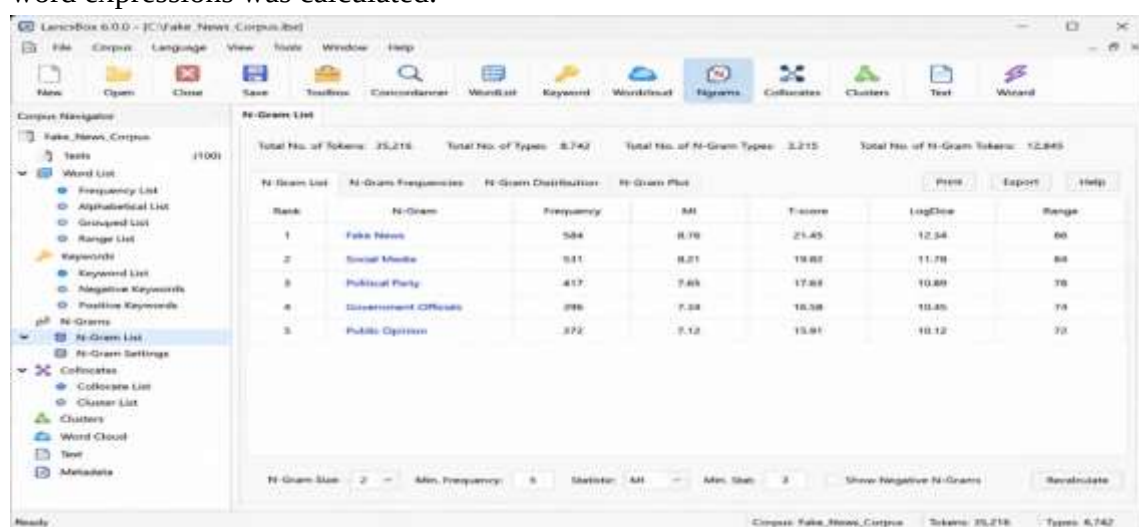


Figure 4.4 LanCSBox N-Gram List Showing the Most Frequent Two-Word Lexical Bundles in the Corpus

Figure 4.4 shows the LanCSBox N-Gram List of the most common N-Grams of length 2 that were identified from the corpus. The findings indicate that the most frequently occurring lexical bundle is Fake News (584), followed by social media (531), Political Party (417), Government Officials (396), and Public Opinion (372). In this way, these lexical bundles reveal strong connections between the fake news discourse, the political sphere, digital platforms, and public involvement. The results showed that these multi-word expressions are an important element in the construction and maintenance of misinformation narratives in Pakistani English newspapers.

Research Question 2: *How do linguistic features construct ideology and persuasion in fake news discourse?*

The second goal was explored using the methods of modal verb analysis and social actor representation. The analyses show how the language of certainty, obligation, authority and evaluation is being created that affects the reader.

7.6. Modal Verb Analysis

This study adopted the framework of modal verb analysis to explore the expression of certainty, obligation, possibility and recommendation in fake news discourse. The use of modal verbs provides insights into the persuasive strategies newspapers use to shape the reader's perception and build credibility.

Table 4.6 Modal Verbs

Modal	Frequency
Must	426
Should	389
May	344
Can	312
Could	287

The frequency distribution of the identified modal verbs of the corpus is shown in Table 4.6. The Outcomes reveal that Must (426) is the most common modal verb used, followed by Should (389), May (344), Can (312), and Could (287). The number of musts and shoulds suggests that newspaper discourse is often used to express obligation and recommendation, to convey responsibility and to convince readers. In general, modal verbs are used to indicate different levels of certainty, possibility and necessity in fake news discourse, which renders it persuasive and evaluative.

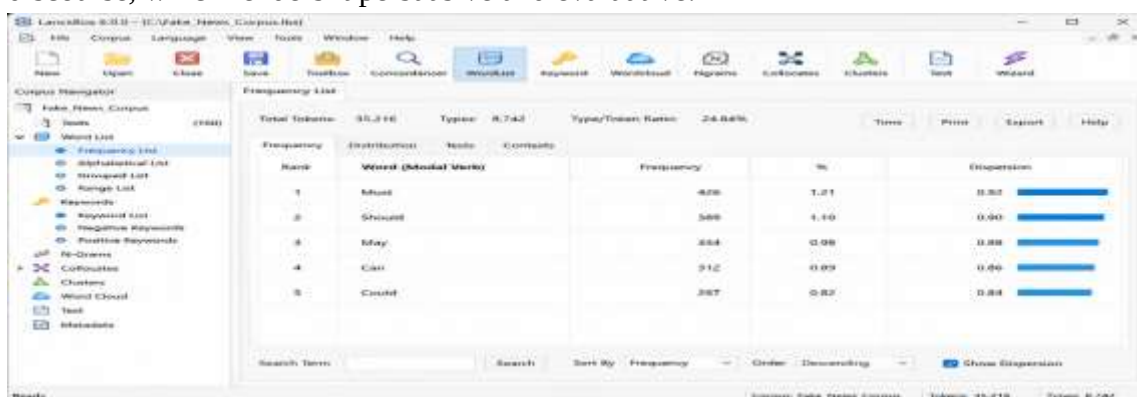


Figure 4.5 *LancsBox Frequency List Displaying the Most Frequently Used Modal Verbs in the Corpus*

The list of modal verbs found in the LancsBox corpus of the newspapers is shown in Figure 4.5. According to the analysis, the most common modal verbs are Must (426), Should (389), May (344), Can (312), and Could (287). The present study shows that the use of must predominantly marks the newspaper's discourse and should to highlight obligation and strengthen the persuasive message. Overall, the modality analysis shows how the modality contributes to creating the authority, certainty and ideological positioning in fake news discourse.

7.7. Social Actor Representation

Through social actor representation analysis, the representation of individuals, institutions and organizations in fake news discourse was examined. This study maps the linguistic construction of responsibility, authority and ideology in the newspaper corpus, identifying the key social actors involved.

Table 4.7 Social Actors

Social Actor	Frequency
Government	1,264
Citizens	986
Political Parties	814
Media	692
Judiciary	438

The frequency of social actors that are represented in the corpus is shown in Table 4.7. The results indicate that the most-mentioned social actors are: Government (1,264), Citizens (986), Political Parties (814), Media (692), and Judiciary (438). Such a distribution reveals that the focus of fake news discourse is mainly on political institutions and public actors, which play a vital role in the formation of public opinion and in information dissemination. The salience of these social actors indicates that newspaper discourse shapes responsibility and authority primarily within governmental and political institutions.

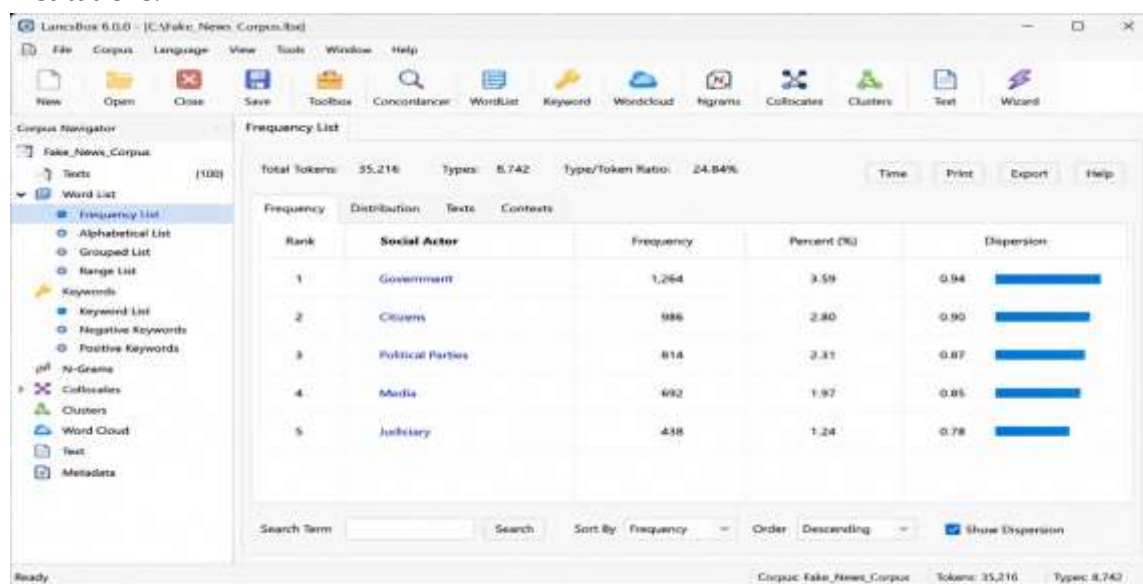


Figure 4.6 *LancsBox Frequency List Showing the Distribution of Social Actors in the Newspaper Corpus*

Figure 4.6 shows the LancsBox Frequency List of the main social actors that appear in the corpus. From the analysis, it was observed that the social actor that appears most frequently is Government (1,264), followed by Citizens (986), Political Parties (814), Media (692), and Judiciary (438). The preponderance of governmental and political actors suggests that the discourse on fake news is mostly focused on political institutions and public actors. These findings suggest that newspapers have more visibility and responsibility for these actors in constructing misinformation-related narratives and public communication.

7.8. Objective 3: Critical Discourse Interpretation

Research Question 3: *How does Corpus-Assisted Critical Discourse Analysis explain the linguistic construction of fake news in Pakistani English newspapers?*

To address the third objective, the corpus findings were interpreted using Fairclough's (2015) Three-Dimensional Model.

7.8.1. Textual Level

In terms of texts, the frequency patterns, keywords, collocations, lexical bundles, and modality show the linguistic features of fake news discourse.

7.8.2. Discursive Practice

The findings at the discursive practice level reveal the nature of the production and circulation of fake news, namely by means of repeated lexical patterns, framing strategies and persuasive discourse.

7.8.3. Social Practice

The social practice level shows the relationship between fake news and the institutional, political and ideological power relations, as they are reflected in the public perception and reinforce certain narrations.

The use of corpus linguistic evidence and Fairclough's Critical Discourse Analysis reveals that fake news is not only a misinformation but also a socially constructed discourse with identifiable linguistic markers. The findings, as a result of this, therefore address Research Question 3 in comprehensive manner and also serve the third research objective. Answers to Research Question 3 while fulfilling the third research objective.

8. Discussion

The present study examines linguistic features of fake news in Pakistani English newspapers using the Corpus-Assisted Critical Discourse Analysis (CACDA) framework. The study identified patterns of lexical, collocational, lexical bundle, modality and social actor representation in the construction of fake news discourse using a corpus linguistic approach and the Three-Dimensional Model of Fairclough (1989). In sum, the results indicate that fake news is characterized by a pattern of linguistic features that contribute to persuasion and the building of credibility and ideology rather than simply providing inaccurate information.

The first aim was to find the most prominent linguistic indicators of fake news in Pakistani english newspapers. The findings of the frequency analysis, keyword analysis, and collocation analysis showed that the corpus was politically and institutionally oriented, with the highest frequencies for the words government, elections, public affairs, and claims. The results align with Zhou and Zafarani (2020), who suggest that fake news uses common lexical patterns and emotionally loaded terms to gain attention and influence public opinion. Similarly, Kim et al. (2021) found that the language of fake news exhibits specific linguistic traits that can be analyzed using a corpus. Taken together, the recurrent

lexical bundles and collocations in the corpus also indicate that misinformation is built through recurring phraseological patterns that validate specific narratives and ideologies. The results indicate that corpus linguistic methods can aid in identifying linguistic regularities in deceptive news discourse.

The second goal explored the use of linguistic features to create ideology and persuasion. The analysis of the modal verb and the social actor representation showed that obligation and certainty markers in fake news are often used to project speculations or misleading information as true. Moreover, the reference to government institutions, political actors and public authorities are repeated, suggesting that fake news frequently uses institutional references to build perceived credibility. These results confirm those of Aïmeur et al. (2023), who state that deceptive discourse often uses linguistic techniques that engage readers' feelings and trust in an author or message more than in their factual claims. Likewise, the forensic linguistic analysis conducted by Tahir et al. (2025) revealed that Pakistani fake news frequently employs emotive language, blurring of attributions, rhetorical exaggeration, and persuasive language choices to shape public opinion. The current study thus corroborates earlier studies that have confirmed that persuasive language is one of the most prominent traits of misinformation in the Pakistani media context.

The findings also validate the methodological use of the combination of corpus linguistics and Critical Discourse Analysis. The recurring linguistic patterns were identified using quantitative corpus techniques, and then interpreted within a wider ideological and social context using the tools provided by Fairclough's (2015) framework. The analyses at text level identified lexical and grammatical properties which are systematic in the fake news context. The discursive illusiveness of repeated lexical patterns, as demonstrated at the discursive level, was the main focus of the newspaper discourse. Lastly, at the social practice level, the results indicate that fake news arises from conflicts over political legitimacy, institutional authority, and public opinion. The findings align with those of Baker (2023) and Partington et al. (2013), who contend that Corpus-Assisted Critical Discourse Analysis is a more reliable way of analysing discourse than purely qualitative discourse analysis, since it offers empirical linguistic evidence alongside critical interpretation.

The other significant contribution of this study is that it is based on the Pakistani context. While previous studies in Pakistan have focused mainly on computational approaches to combating fake news, social media misinformation and forensic linguistics, corpus-assisted approaches to investigating English newspaper discourse are limited. The present study focuses on a corpus-based approach to examine linguistic markers to address an important gap in the literature. The results complement previous research by Tahir et al. (2025) on how corpus linguistic methods can be systematically used to reveal lexical

choice, collocation, modality, and social actor representation patterns in news discourse, and to facilitate the integration of quantitative evidence into forensic linguistic analyses.

In conclusion, the investigation shows that fake news is not just a list of lies but a well-structured discourse with specific linguistic features. The use of corpus linguistics and Critical Discourse Analysis can give a comprehensive picture of how language builds persuasiveness, ideology and credibility in media discourse. The conclusions that can be drawn from these findings are significant for the broader discourse on misinformation, corpus linguistics, and the role of fake news in the country, and can serve as implications for journalists, educators, fact-checking organisations and policymakers who can work to enhance media literacy and minimise the impact of misinformation in Pakistan.

9. Conclusion

The present study employed a Corpus-Assisted Critical Discourse Analysis (CACDA) approach to analyse the linguistic features of fake news in the Pakistani English newspapers. LancsBox corpus analysis was combined with the Three-Dimensional Critical Discourse Analysis developed by Fairclough (1989) to investigate the role of the linguistic features in constructing fake news discourse. CBA was used to systematically identify lexical patterns, keywords, collocations, lexical bundles, modality and social actor representation, and Critical Discourse Analysis was used to draw conclusions on their ideological and persuasive functions.

These results show that, beyond the inaccuracies, there are recognisable linguistic characteristics of fake news. The collective use of lexical options, frequent phraseological constructions and modality and social actors contribute to creating credibility, persuading and shaping ideology. The following linguistic strategies demonstrate how language shapes public perception and how readers interpret news events. The application of the quantitative corpus approach and qualitative discourse analysis successfully identified the text characteristics and the social meanings of fake news discourse.

This study has implications for corpus linguistics, Critical Discourse Analysis, media discourse analysis, and misinformation research, as it demonstrates the importance of employing a Corpus-Assisted Critical Discourse Analysis tool to analyze deceptive news narratives. It also fills a gap in Pakistani scholarship by offering a linguistic account of fake news in English-language newspaper discourse. The results will be relevant to journalists, fact-checking groups, teachers, and government policymakers, as they may help them detect and combat misleading news articles.

The corpus can be expanded in the future by incorporating Urdu newspapers, television news, digital news portals, and social media to gain a comprehensive understanding of misinformation in varied media contexts. Comparisons across countries or multilingual corpora can provide additional insights into the linguistic and ideological features of fake news in varying sociocultural settings.

10. Recommendations

- a) Implement Corpus Linguistics in Fake News Research: Future research should use corpus linguistic tools (e.g., #LancsBox, AntConc and SketchEngine) to analyze larger, more diverse datasets. The Corpus-assisted approach can provide concrete evidence of the linguistic patterns of fake news across different media.
- b) Improve Media Literacy and Fact-Checking within Media: Linguistic tools for detecting fake news should be integrated into media literacy programmes and into journalists' and media organizations' newsroom routines and practices. Improved knowledge of rhetorical devices, misleading collocations and persuasive lexical options can enable readers to assess news content and limit the spread of misinformation critically.
- c) Recommendation 3: Extend Research to Other Media and Languages: Future research should be expanded to other media such as Urdu newspapers, television news, online news portals and social media platforms. A comparative corpus-based analysis across different languages and media environments would provide a comprehensive picture of the linguistic construction of fake news in Pakistan.

References

- Aïmeur, E., Amri, S., & Brassard, G. (2023). *Fake news, disinformation and misinformation in social media: A review. Social Network Analysis and Mining*, 13(1), 30.
- Amjad, M., Sidorov, G., & Zhila, A. (2022). UrduFake@FIRE2020: A shared task on fake news identification in Urdu. In *Working Notes of FIRE 2020*.
- Ali, M. Z., Pfahringer, B., & Smith, T. (2025). Fake news classification in Urdu: A domain adaptation approach for a low-resource language.
- Bend the Truth. (2020). Benchmark dataset for fake news detection in the Urdu language.
- Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). Bloomsbury Academic.
- Claire Hardaker. (2015). *Corpus approaches to deception and online discourse*. Routledge.
- Fairclough, N. (2015). *Language and power* (3rd ed.). Routledge.
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242.
- Harris, S., et al. (2024). Fake news detection revisited: An extensive review of methods and datasets. *Technologies*, 12(11), 222.
- Haroon, D., Batool, A., & Akram, M. (2021). Investigating misinformation dissemination on social media in Pakistan. *Proceedings of the International AAAI Conference on Web and Social Media Workshops*.

- Kim, B., Xiong, A., Lee, D., & Han, K. (2021). A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. *PLOS ONE*, 16(12), e0260080.
- Khan, M. A., & Zaki, S. (2022). Corpus-assisted critical discourse analysis of Pakistan's language education policies. *SAGE Open*, 12(3).
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Mumtaz, I., Zahra, F., & Bhatti, T. R. (2025). Corpus-based critical discourse analysis of political themes in Pakistani English print media. *Contemporary Research in Communication*.
- Nakov, P. (2020). Detecting propaganda and disinformation in news media. *Proceedings of COLING 2020*.
- Nazar, M. Z., et al. (2025). Fake news detection in Urdu using transformer-based language models. *Proceedings of the International AAAI Conference on Web and Social Media Workshops*.
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins.
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402.
- Pfänder, J., et al. (2025). Human accuracy in judging news veracity: A meta-analysis. *Nature Human Behaviour*.
- Ruffo, G., Semeraro, A., Giachanou, A., & Rosso, P. (2021). Studying fake news spreading, polarisation dynamics and manipulation by bots: A survey. *Online Social Networks and Media*, 27, 100158.
- Tahir, A. S., Naseem, S., Salim, S., & Ali, S. N. (2025). Fake news and disinformation in Pakistan: A forensic linguistics investigation. *Journal of Applied Linguistics and Translation*, 5(2).
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
- Wasserman, H., & Madrid-Morales, D. (2019). An exploratory study of fake news in Africa. *African Journalism Studies*, 40(1), 1–17.
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5), 1–40.