

# Discourse Analysis: Media Bias and Linguistic Patterns on News Reports

Ziran Zhou<sup>1,\*†</sup>

<sup>1</sup>Shangde Experimental School, Shanghai, Shanghai, China

\*Corresponding author. Email: <sup>1</sup>zhouziran@shangdejy.com

## ABSTRACT

A well-directed article may affect the objectivity of the event itself and the reader's interpretation. Language is powerful because it can appeal to emotions and ideologies, and therefore readers need to establish a more objective judgement in the various news reports from different news media. Based on the purpose of the study, this paper discusses the media bias and language patterns of different news reports on the theme of "COVID-19". By using corpus methods, comparative analysis and discourse analysis, this study analyses 20 articles published by CNN and CBN, two news media with left and right political stances respectively. The results reveal that (1) The two news media have their own preferences in phrase collocation. (2) Exposure to media bias is closely related to the high-frequency word "vaccine".

**Keywords:** Media bias, Linguistic patterns, News reports, Corpus approach.

## 1. INTRODUCTION

In 2020, a sudden outbreak of coronavirus (COVID-19) spread worldwide, causing enormous damage in various fields. In this context, society descends into a state of instability. People have been urged to keep informed about the daily trends of the epidemic, mainly through the information available on the Internet. As a result, each media outlet tries to create hot spots to attract more readers and promote its own point of view, which means that the type and scope of coverage related to COVID-19 may depend on the political party or political leanings of the media outlet in question [1].

Different news media have their own language preferences when publishing news, or what one might call "media bias", to show a political stance or viewpoint on a particular event. Media bias is usually represented by journalists, representing news stories that reflect only some of the facts and events in support of the journalist's ideological, party or political position. When media bias is combined with language patterns, examining the deeper meanings and discourses behind the words in news articles can reveal the cognitive and social functions of language choice [2]. Generally, the structure of the language, especially the complexity of the language, affects the credibility of information. Targeted news published by different media can also state bias and may affect the objectivity of the events themselves and the interpretation of readers. Therefore,

the purveyors and institutions of news must not only keep themselves free from bias and follow strict principles of honesty, objectivity and comprehensiveness in reporting news, but also make it clear to newspapers and the public that news organisations are adhering to such a line [3].

The basis of the commentary needs to be built on the articles themselves. This paper aims to verify the relationship between media bias and linguistics through a broader analysis. By using methods that analyse certain linguistic patterns - high-frequency words and word collocations - this research will help to establish a more objective judgement across the various news reports from the different news media in today's society, as disinformation can easily influence people's thoughts and judgements.

This paper is divided into six parts. It begins with an abstract and section 1 introduction. Section 2 and 3 are the review of literature and methodologies. Section 4 presents data collection and analysis, which introduces the results. Section 5 is the findings of the study. At last, section 6 concludes the whole article.

## 2. LITERATURE REVIEW

### 2.1. *Critical discourse analysis and corpus linguistic*

Discourse analysis is part of a broader cognitive and social theory of the rules and strategies that form the basis of media discourse generation and understanding. Critical discourse analysis (CDA) mainly serves as two approaches: (1) correlates the repeated patterns in the text with sociolinguistic features in the original context, and vice versa; (2) CDA uses both quantitative techniques, namely frequency and keyword lists-supplemented by more detailed textual analysis, and combines both techniques in order to reveal the profound meaning of the discourse under investigation, not only the superficial meaning [4]. Then, an ideological meaning is added to the textual evidence, which helps reduce the bias of researchers.

On the other hand, with the deepening and development of linguistic research, more and more linguists realize that it is difficult to reveal the characteristics of language structure and the essence of language use by simply studying isolated sentences. Under this circumstance, corpus linguistic analysis, the analysis of linguistic patterns in and across naturally-produced texts by using a programmed computer app provides a way that can dig out the hidden meanings in the vocabulary through inspection [5]. It helps to provide quantitative evidence of the existence of words by identifying repetitive patterns of language use and checking collocations.

Unfortunately, many scholars have raised debate on the two methods. One of the claims suggests that corpus linguistic lacks the ability to produce negative feedback based on the given text and is not able to give a further interpretation [6]. In addition, scholars have questioned that semiotic elements in news media, such as video, pictures, audio files, and hyper-textual links to other blogs, can not be involved within a discourse analysis [7].

Nevertheless, the innovation of corpus linguistics software makes it possible to a more accurate and quick elicitation of given lexemes and phrases based on their sentential contexts [8]. The collocation table will represent the likelihood and habitual juxtaposition of a particular word with another word, which supports the further analysis and explication of the patterns in the context. Also, other factors would not be added as a part of the corpora due to the fact that the text of the news article is the main source.

### 2.2 *Media bias and linguistic patterns in news*

Most of the time, the news is not only a complete or incomplete description of the facts but a new

interpretation of the facts based on the belief of the reporter or certain social patterns and values. According to the findings of Spiteri J, the media continues to play a key role in disseminating health information during the pandemic [9]. In other words, aligning media information with the suggestions advocated by medical experts, without any political leanings, is essential to combat the spread of COVID-19 and future epidemics. Many pieces of research of discourse analysis focused on the semantic structures of news articles, that is, on meaning, since previous studies mostly explored the sentence structure [10].

In addition, different units of analysis can be distinguished in discourse: individual words (lexical items), various structures of the clause, the whole sentences, the sequences of sentences (paragraphs), or the whole discourses. To exemplify, while determining the overall theme or purpose of discourse, it can only start from analyzing the semantic level of the discourse as a whole, but not a single word or sentence. Therefore, a major weakness of the previous study is the rough distinction of the “local” and “global” structure of the discourse.

However, in order to decipher the ideology and possible prejudice of the investigated news organization, the study examined the discourse structure of the two news media involved in different political positions. By using the method of analyzing high-frequency words and vocabulary collocations, it is clarified that both positive and negative representations are related to the ideological model of the news media.

## 3. METHODOLOGY

### 3.1 *The choice of news media and articles*

This study chooses twenty news articles from two news media, which correspond to an opposite political position. More specifically, the representative of the US left-wing media shows ideas such as freedom, equality, fraternity, rights, progress, reform, and internationalism; however, the right-wing standpoint focuses on ideologies such as authority, hierarchy, order, duty, tradition, reaction and nationalism [11]. To support these positions, journalists spend huge ink on applying implications and repeating the keywords which are useful to their news reports. Based on the above information, the choice of such two news media was carefully considered. Cable News Network (CNN) and Christian Broadcasting Network (CBN) are both well-known news media in the United States, so the quality of the articles can be ensured. Furthermore, articles chosen were published between March 2020, and October 2021, mainly focused on political or medical context; in other words, the themes of the articles selected from the two media will not be too deviated.

### 3.2 Corpus approach and discourse analysis

The Corpus Approach is interpreted with four major features: (1) It analyzes the actual patterns of language use in natural texts, which leads to an empirical outcome. (2) It utilizes a large and scaled collection of natural texts as the basis for analysis. (3) It relies on extensive use of computers for analysis. (4) It depends on both quantitative and qualitative analytical techniques [12]. All the characteristics are suitable for establishing a wide-ranged data analysis.

With the use of the application AntConc, comparison and analysis of the linguistic patterns can be conducted within two small corpora, which consist of

20 news reports on the COVID-19 from Cable News Network and Christian Broadcasting Network (10 from each). Firstly, each corpus will establish a “Top 12 Content Lexemes” table with the repeated keywords and the frequency. Thus, the collocation phrases and lines of lexeme with the highest appearance will be presented and analyzed. At last, the key word in context will be interpreted based on the findings of the collocation lines.

## 4. ANALYSIS AND RESULTS

### 4.1 Comparative analysis of the lexical choice

**Table 1.** Top 12 Content Lexemes in the CBN corpus and CNN corpus

The CBN corpus (5410 tokens)			The CNN corpus (7988 tokens)		
Rank	Keyword	Frequency	Rank	Keyword	Frequency
1	vaccine	58	1	vaccine	161
2	vaccinated	35	2	FDA	24
3	religious	19	3	safety	21
4	health	16	4	Pfizer	20
5	vaccination	14	4	health	20
6	risk	12	4	public	20
7	against	10	7	against	18

The articles selected from CBN have an average length of 547 words, while CNN news articles have an average length of 809 words. With all the requirements, 20 randomly selected news articles built two small corpora, containing 13398 tokens in total. From table 1, a comparison between each top 12 content lexemes in CBN corpus and CNN corpus can be interpreted. Although CBN and CNN differ greatly in their political stances, they still share a lot in common in the wording of the text.

The two corpora have many similarities, such as the word “vaccine” has the highest frequency in both corpora. Also, lexemes related to political factors—“freedom” “federal” “government”— show a high frequency of appearance in the context. However, the two corpora still have differences. As expected, CBN holds a relatively conservative tone while choosing the lexemes. From the keywords “risk” and “health”, it can be concluded that the article does not want to express

denial intuitively, that is, oppose remarks such as vaccines, but imply the risks of vaccines through euphemistic expressions. Meanwhile, the CNN corpus selects the words “safety” and cooperate with “effective” at a high rate, which demonstrates the credibility of the vaccine to the readers through repeated emphasis. On the other hand, the word choice “FDA” “Pfizer”, which are all technical terminologies, and other nouns such as “approval” show a relatively positive tone towards the opinion of the vaccine.

## 4.2 Corpus approach and discourse analysis

**Table 2.** Collocation analysis of “vaccine” in CNN corpus

Cluster	Rank	Freq	NormFreq
vaccine development	1	6	0.048
vaccine safety	2	4	0.032
vaccine against	3	3	0.024
vaccine mandates	3	3	0.024
vaccine advisory	6	2	0.016
vaccine efficacy	6	2	0.016
vaccine advisers	8	1	0.008
vaccine approval	8	1	0.008
vaccine clinical	8	1	0.008
vaccine credential	8	1	0.008
vaccine effectiveness	8	1	0.008
	8	1	0.08
vaccine existed	8	1	0.008
vaccine experts	8	1	0.008
vaccine hesitancy	8	1	0.008
vaccine initiatives	8	1	0.008

**Table 3.** Collocation analysis of “vaccine” in CBN corpus

Cluster	Rank	Freq	NormFreq
vaccine mandate	1	3	0.081
vaccine passports	2	3	0.081
vaccine booster	3	1	0.027
vaccine campaign	4	1	0.027
vaccine discouraging	5	1	0.027

vaccine seekers	6	1	0.027
vaccine status	7	1	0.027
vaccine verification	8	1	0.027

From table 2 and table 3, the vaccine-related collocations in both corpora are presented. As shown in the cluster, there are more collocations with the word “vaccine” in the CNN corpus. According to the political stance preference of CNN corpus, the data shows an expected pattern in the lexeme cluster in two corpora, which is in line with the law of media bias. The collocation of lexemes within the CNN corpus mainly focuses on the theme of positive effects and guidance. Although the cluster “vaccine hesitancy ” appears one time, it has no effect on the overall trend. Clusters “vaccine clinical” “vaccine efficacy” “vaccine approval” all represent an official and professional position, trying to persuade people to the opinions that vaccine has more benefits than risks by presenting

creating reliable context and collocation. Nevertheless, it is unexpected that even if CBN is biased towards the right in its political stance, which republicans and rightists continue to raise vaccine skepticism, words that directly express opposition to the vaccination rarely appear. From the eight clusters, except “vaccine discouraging”, all the collocations show a relatively neutral and objective expression. This shows that in the CBN corpus, the frequency of the combination of neutral words and “vaccine” is much higher than that of word collocations with positive or negative orientation.

**4.3 Further interpretation of key word in context**

**Table 4.** Collocation analysis of “vaccine” in CBN corpus

	File	Left Context	Hit	Right Context
1	CBN-2021-9-20.	RFI Communications Director Nathan Berkeley cautions people to sort out their reasons for opposing the	vaccine	and avoid pursuing the religious exemption if motivated by politics. “A lot of people
2	CBN-2021-8-4.	Biden's approval fell 10 points since June from 62 to 52 percent as the government pushes new	vaccine	and mask mandates. On Tuesday, he called for tighter restrictions to combat the highly
3	CBN-2021-5-6.	neither report cited by the author provided evidence to showcase a direct link between the	vaccine	and the deaths in question — a fact the author did not highlight. In fact,
4	CBN-2021-1-6.	leading authority on Covid-19, says he won't be first in line to get this	vaccine	because there haven't been enough studies about possible long-term complications. "If somebody
5	CBN-2021-1-6.	people who are really scared about the virus and if they want to take the	vaccine	because they think they're going to be safer then, by all means, do
6	CBN-2021-9-12.	worked hard to educate, encourage, cajole (and) support individuals to get comfortable with receiving the	vaccine	but we are not passing judgment on any single person who says it's

7	CBN-2021-4-18.	of 62 doses per 100 people, according to Our World in Data, an online research site. The	vaccine	campaign offered hope in places like Nashville, Tennessee, where the Music City Center bustled
8	CBN-2021-8-3.	percent) who attend religious services and are hesitant say a faith-based approach about the	vaccine	could change their mind. "Faith-based approaches have been and will continue to be
9	CBN-2021-5-6.	was taken down in accordance with the company's policy of removing accounts that distribute "	vaccine	discouraging information on the platform." It also allegedly accused the site of publishing "false
10	CBN-2021-9-17.	wave.â€• Several doctors on the panel, however, noted concerns about myocarditis. Dr. Paul Offitt, a	vaccine	expert with the Children's Hospital of Philadelphia, said he would "really have trouble"

Extending from the collocation studies, the word "vaccine" is chosen to be the hit for the KWIC analysis and to prove the trustiness of the objectivity of CBN news articles. Different from the findings in the previous part, with the left and right context, the text with vaccines becomes more biased. For instance, "... had been nervous about getting the vaccine after having a bad reaction to a flu shot" and "... he won't be first in line to get this vaccine because there haven't been enough studies about possible long-term complications" all allude to an implicit negative opinion towards the vaccine. While describing the facts, the author's relatively vague language uses and attitude create uncertainties about the vaccine. Meanwhile, the voice of professionals quoted in the article— "...noted concerns about myocarditis. Dr. Paul Offitt, a vaccine expert with the Children's Hospital of Philadelphia, said he would "really have trouble" supporting booster shots for everyone over 16 years old"— also clearly points out that vaccination is not suggested. In other words, the expression of articles from CBN news may be more pertinent in lexical collocation; however, the attitude of CBN news is still not utterly objective through the context of the whole sentence.

## 5. CONCLUSION

The results show that although the two news media hold different preferences on positive and negative word choices respectively, the lexeme "vaccine" and lexemes related to the vaccine are pivotal in the context from both corpora. This proves that news media choose to favor specific words and phrases while describing a certain fact. The outcome of the CNN corpus, as expected, represents the collocation mostly focuses on affirmative facts and guidances; while the CBN corpus

does not have a strong directivity on word choices, but the neutral statement of facts. Nevertheless, the KWIC presents two major characteristics of the CBN news articles in a more critical way: (1)The use of implicit phrases which implies the downsides of vaccine; (2)Quotes from professionals that help strengthen the points made by the article. Therefore, the conclusion indicates that media bias of news agencies is indeed related to the theories in the field of linguistics.

In summary, although the current study has limitations on the range of data and more detailed analyzing method, the results still appeal to a clear relationship between the language pattern, especially lexeme collocation and the KWIC, and the biases of different media. With the analyzing procedure that symbolizes how the articles two news media's political position and ideologies, the consequences of this research provide methodological significance for future discourse analysis on news articles.

## REFERENCES

- [1] Spiteri, J. (2021) Media bias exposure and the incidence of COVID-19 in the USA. *BMJ Global Health* 2021;6:e006798. DOI: 10.1136/ bmjgh-2021-006798
- [2] Van Dijk, T. A. (1983). *Discourse Analysis: Its Development and Application to the Structure of News*, *Journal of Communication*, Vol. 33, Issue 2, Pages 20–36. DOI: <https://doi.org/10.1111/j.1460-2466.1983.tb02386.x>
- [3] Van Dijk, T. A. (1996). *Discourse studies: a multidisciplinary introduction*. Vol. 2: *Discourse as social interaction*. London: SAGE.

- [4] Shrivastava, K. M. (2007). News agencies from pigeon to internet. Berkshire, UK: New Dawn Press.
- [5] Lynne, F. (2012). Corpus-based discourse analysis. J.P. Gee & M. Handford (Eds.), Routledge Handbook of Discourse Analysis (pp.174-187). Routledge.
- [6] Laura, A. (2013). Corpus linguistic analysis of written language: How to use “I” and more. URL: <https://www.digitalrhetoriccollaborative.org/2013/11/06/using-corpus-linguistic-analysis-of-writing-with-students/>
- [7] Gena R. B. (2010). Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers. Michigan ELT.
- [8] Ooi, V., Tan, P., and Chiang, A. (2007) ‘Analysing personal weblogs in Singapore English: the Wmatrix approach’, in EVariEng (Journal of the Research Unit for Variation, Contacts, and Change in English). Vol. 2: Towards Multimedia in Corpus Studies. Finland: University of Helsinki.
- [9] Sibó, C. (2020). Corpus Linguistics in Critical Discourse Analysis: A Case Study on News Reports of the 2011 Libyan Civil War. *Stream: Culture/Politics/Technology*, 5(1), 21-28. <http://journals.sfu.ca/cpt/index.php/stream/index>
- [10] Wikipedia. (2021, October 7). Left-right political spectrum. [https://en.wikipedia.org/wiki/Left%E2%80%93right\\_political\\_spectrum](https://en.wikipedia.org/wiki/Left%E2%80%93right_political_spectrum)
- [11] Biber, D., Conrad, S., & Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*.
- [12] Shrivastava, K. M. (2007). News agencies from pigeon to internet. Berkshire, UK: New Dawn Press.
- [13] Petro Tolochko & Hajo G. Boomgaarden. (2018). Analysis of Linguistic Complexity in Professional and Citizen Media, *Journalism Studies*, 19:12, 1786-1803, DOI: 10.1080/1461670X.2017.1305285
- [14] Konstantina, L & Ralf, K & Felix N. (2017). Identifying Media Bias by Analyzing Reported Speech. 943-948. DOI: 10.1109/ICDM.2017.119.
- [15] Xiang, Y., & Sarvary, M. (2007). News Consumption and Media Bias. *Marketing Science*, 26(5), 611–628. <http://www.jstor.org/stable/40057083>