



## A Corpus-based Discourse Analysis of Elected News Reports in CNN

Chen Guojing<sup>1</sup>, Li Yanfang<sup>2\*</sup>

<sup>1</sup>Kashi University, Kashi, China

<sup>2\*</sup>Kashi University, Kashi, China

Email: Chen Guojing1 ksdiana@aliyun.com, Li Yanfang2 joanna202210@126.com

---

**Abstract:** Since the outbreak of COVID-19, news reports related to the global pandemic have risen sharply in a short period. In that regard, the perspectives and attitudes of the reports related to the pandemic have been characterized by diversity and universality. This paper will investigate the discourse characteristics of the CNN news reports about COVID-19. As we all know, news reports are influenced by specific political, economic, social, cultural, and other factors. These news reports often imply specific attitudes of the media. Based on the self-built small corpus, this paper combines quantitative and qualitative analysis to analyze the English reports about COVID-19 in CNN media from the perspective of CDA, starting from word list, keyword clusters, and collocation of the corpus to find out the characteristic of these news reports. The study found that CNN news reports mainly focus on the time, place, case, and impact that is related to the pandemic. The focus of CNN news is mainly on mortality, infection, and high incidence. The analysis of collocation and its concordance shows that CNN news reported that the government has taken plans and measures to defeat the outbreak of the pandemic.

**Keywords:** corpus linguistics, discourse, COVID-19

---

### Introduction

After the outbreak of COVID-19 in 2020, major foreign newspapers have started to report the pandemic. These news reports play a key role in the international construction of a national image. It is an important way to analyze the characteristics of the reports related to the pandemic and explore the potential meaning behind it to enhance China's international communication in the world. Therefore, this paper aims to analyze the characteristic of the reports related to the pandemic.

Corpus-based Discourse Analysis (CDA) has become a trend in language research in recent years due to its increasingly prominent complementary advantages<sup>[1-6]</sup>. Other corpus linguistics studies have expanded from lexical research to discourse analysis. Many researchers also widely used corpus technology for text analysis<sup>[7-10]</sup>. Baker et al. believes the combination of corpus and discourse is a methodology in discourse analysis. Their potential theoretical and methodological intersection appears to benefit each other and the combination of the two approaches strengthens the theoretical foundation of the study<sup>[4]</sup>.

Based on the Corpus-based Discourse Analysis, this paper examines the news reports reported by an internationally influential media named Cable News Network (CNN). The goals are to find out whether it follows objectivity, impartiality, and equality when it reports the pandemic and what attitude and perspective should be used to construct China's image?

---

[Received 10 Nov 2022; Accepted 16 Dec 2022; Published (online) 31 Dec 2022]



Attribution 4.0 International (CC BY 4.0)

## **The Application of Corpus in Discourse Analysis**

In recent years, many researchers have widely used word frequencies, concordance, word lists, collocations, or word clusters to improve the credibility of discourse and revealed patterns or meanings implicit in the discourse [2] [5-6][11].

Baker & McEnery used the corpus-based discourse analysis to analyze the keywords and related collocations and found significant differences in the descriptions of refugee reporting by major UK news. Through the analysis of keywords and their related words, a strong conclusion is drawn that different discourses are different reflections of the same reality and the observer's attitude toward things [5].

In China, Gao Chao conducted a comparative analysis of the reports of China Daily and Yahoo by combining the subject headings and associative words. The study found that there is a close relationship between the use of words and the expression of the theme. The domestic and international reports are the same in terms of the expression of the theme, indicating that there is more consensus in international public opinion on the issue of public health [12].

After the outbreak of COVID-19, many scholars have begun to analyze the characteristics of the news reports related to the pandemic from the perspective of framework and critical theory [13-15]. These researches focus on exploring the ideology and attitude towards the emergency behind the language of the news reports.

These previous researches share three basic characteristics. Firstly, there is no prejudice, and its arguments depend on the vocabulary generated by statistical methods. Secondly, they are large-scale comprehensive analyses. Thirdly, they are comparative analyses. These researches reveal that the core principle of corpus linguistics is the use of authentic texts. The texts must be completed and compared across different corpora.

## **Research Purpose**

Since its outbreak in early 2020, COVID-19 has been the focus of media attention in various countries. As the most penetrating and disseminating medium, news reports have the characteristics of objectivity and sociability, and it is an important source for the public to obtain information. However, in the course of news reports, reporters are often prone to express ideological meaning implicitly, which has a subtle influence on readers' cognition and judgment, resulting in readers' prejudice and other adverse effects.

Therefore, the discourse mode of news reports and the implicit meaning hidden behind the discourse are worth exploring. This essay uses the corpus-based discourse analysis to explore the characteristics of the news reports about the pandemic selected from CNN. The specific texts are analyzed from the aspects of word-list, key word cluster and collocation. A multi-perspective analysis aims at exploring the discourse characteristics of CNN's pandemic news reports.

## **Methodology**

This paper combines quantitative and qualitative analysis through a self-built corpus and uses modern statistical analysis software to analyze news reports about the pandemic selected from the Cable News Network, an American broadcasting company that sends television news programs worldwide. The reason for choosing CNN is that it provides a way for Chinese people to understand China's image through its reports. Moreover, CNN also ranked 476th in the list of "2018 World Top 500 Brands". The level of research is relatively rich, involving three levels of vocabulary, grammar, and semantics. At the same time, a large number of examples are listed. The characteristics of CNN's pandemic news reports are summarized.

In order to conduct a comprehensive and objective analysis of the characteristics of English news, this study focuses on the international news broadcast company Cable News Network. In order to ensure the continuity of the corpus used in the research, the author selected a total of 145 English news reports based on keywords related to the pandemic from January 2020 to June 2022 with a total of 51347 words. The corpus includes all the reports related to the pandemic. The average word count in each news report is around 300 words. The news reports that are above 300 words were removed in order to make consistent of the length and style. Finally, the author built a corpus named the CNN International Radio Broadcasting Corpus of English News on the Pandemic.

## Research Instrument

This research combined quantitative and qualitative analysis to provide an objective analysis. AntConc analysis software is used to analyze the characteristics of the corpus. Developed in 2002 by Laurence Anthony, AntConc is a free research tool for monolingual corpus analysis. It offers functions including concordance, concordance plot, file view, cluster/n-gram, collocates, word list, keyword list, and so on. This professional software is suitable for corpus linguistics, translation studies, and foreign language teaching, which is suitable for this study.

## Data Gathering

Small self-built corpora can be based on the researcher's specific research purpose. The purpose is to select an appropriate analysis tool, which can analyze the data offline or online. This article intends to search for the keywords with "epidemic", "pandemic", "novel coronavirus" and "COVID-19" on the CNN website (<http://www.bbc.com>). The time ranges from January 1, 2020, to June 1, 2022. The results were used to build a small pandemic news corpus with 51347 words.

Topic	Word Count	Source	Word Types	Word tokens	Number	Average length of each report
COVID-19	51347	CNN	4942	52061	145	354

Table 1. Descriptive Data of the Corpus

## Research Design

This research is conducted in three steps. Firstly, the author selected CNN's news reports with the topic of COVID-19 and built a small news corpus. Secondly, this research used the software AntConc3.2.4 to analyze the discourse characteristics of word lists, key word clusters and collaboration in CNN news reports. Thirdly, the author analyzed and discussed the results from the CAD perspectives.

## Results and Discussion

In this part, the author aims at analyzing and discussing the results from the CAD perspective. To address the objectives of this study, three aspects of the discourse were conducted. Firstly, the word frequency or word list was examined. Secondly, the keyword list or keyword clusters of the corpus was examined. Finally, word collocation was investigated. Therefore, in this part, the author will illustrate the results from the following aspects.

## Word List

Word frequency is based on the analysis of discourse at the vocabulary level, which is one of the basic applications of corpus analysis technology. News reports can leave a deep impression on the reader by using high-frequency words and mentioning a word or concept many times over and over again, euphemistically expressing the author's tendencies and ideologies. So, starting with high-frequency words and analyzing the lexical characteristics of discourse helps us to grasp the salient characteristics of discourse. Based on the corpus software, the self-built corpus are sorted by high-frequency words, and some words with no practical meaning are removed to obtain the first 30 high-frequency words. The results are shown below in table 2 and table 3.

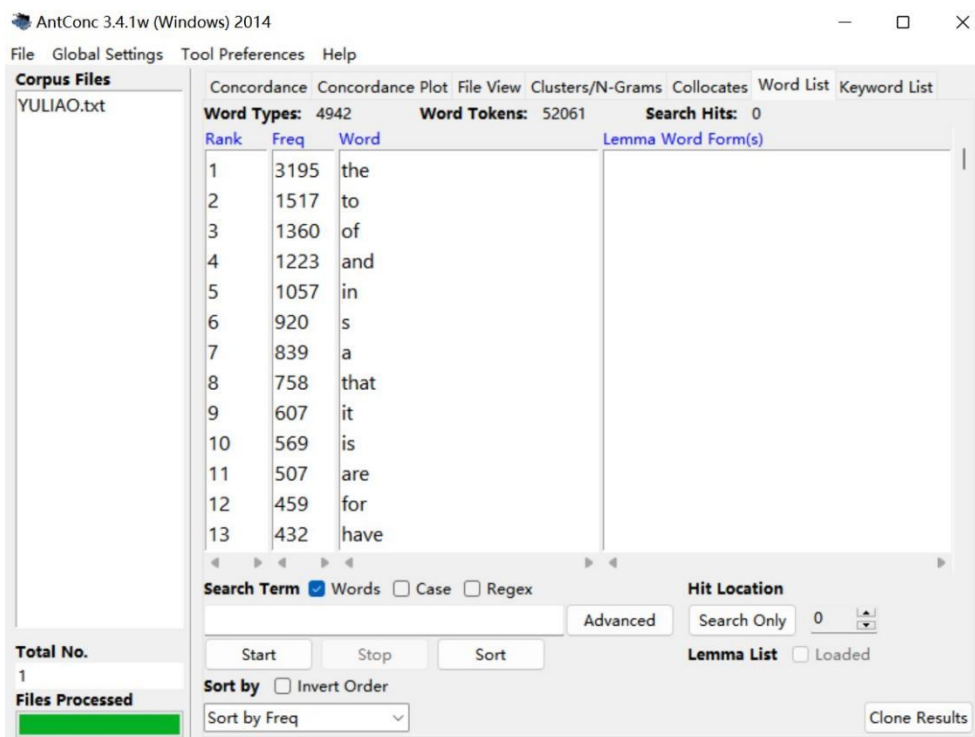


Table 2. Word list out put of AntConc

Rank	Frequency	words
1	469	says/said/say
2	398	coronavirus
3	370	people
4	240	covid
5	218	states/state
6	176	vaccine/vaccines
7	168	country/countries
8	161	health
9	157	disease
10	152	cases
11	142	virus
12	117	officials
13	117	world
14	115	united
15	114	get

16	111	number
17	107	spread
18	105	pandemic
19	104	government
20	98	president
21	68	Americas
22	68	million
23	64	death
24	44	symptoms
25	42	emergency
26	42	infections
27	42	positive
28	39	experts
29	39	medical
30	36	increase

Table 3. Word Frequency

From the table 2 it becomes apparent that the most frequent in the corpus are functional words (also known as function words). Some words category, such as pronouns, determiners, conjunctions, prepositions that we do not talk in this research. This essay only analyze the content words (nouns, verbs, adjective, adverb).

As can be seen from the list of high-frequency words, some words in CNN media reports with the topic of COVID-19 appear significantly more frequently than other words, such as coronavirus, people, virus, health, and epidemic and so on. These words show that American media focus on the outbreak of the COVID-19 and its influence, which is suitable for the requirements of objectivity in journalistic reporting. However, it is worth noticing that in the list of high-frequency words, the words says/said/say and states/state in the CNN appear more frequently, indicating that the government and official shows a high degree of concern about the COVID-19 that happened around the world and has carried continuous reporting.

Besides that, a major feature of news report is the citation. In order to reflect the authenticity, the words of the relevant people are often quoted and commented on, so as to implicitly convey the views of the reporters. Careful analysis shows that these reports focus on the vaccine research and development by the government in the face of the sudden outbreak of the epidemic. The words of the news reports are full of the word of “people”, it means that COVID-19 has severely impacted people’s work, life and school learning. This study will conduct an key cluster and collocation analysis of some representative high-frequency words in the next part.

### Key clusters

Word clusters are studied with the units of multiple words that appear continuously in a text. By observing the word clusters, we can find the words that often appear together with some specific words and the recurring word clusters, so as to deepen the understanding of the vocabulary and the attitude of the reporter, and interpret the intended meaning and information. In this study, the high-frequency word “people” will be searched for the cluster and further verify the results of the previous analysis.

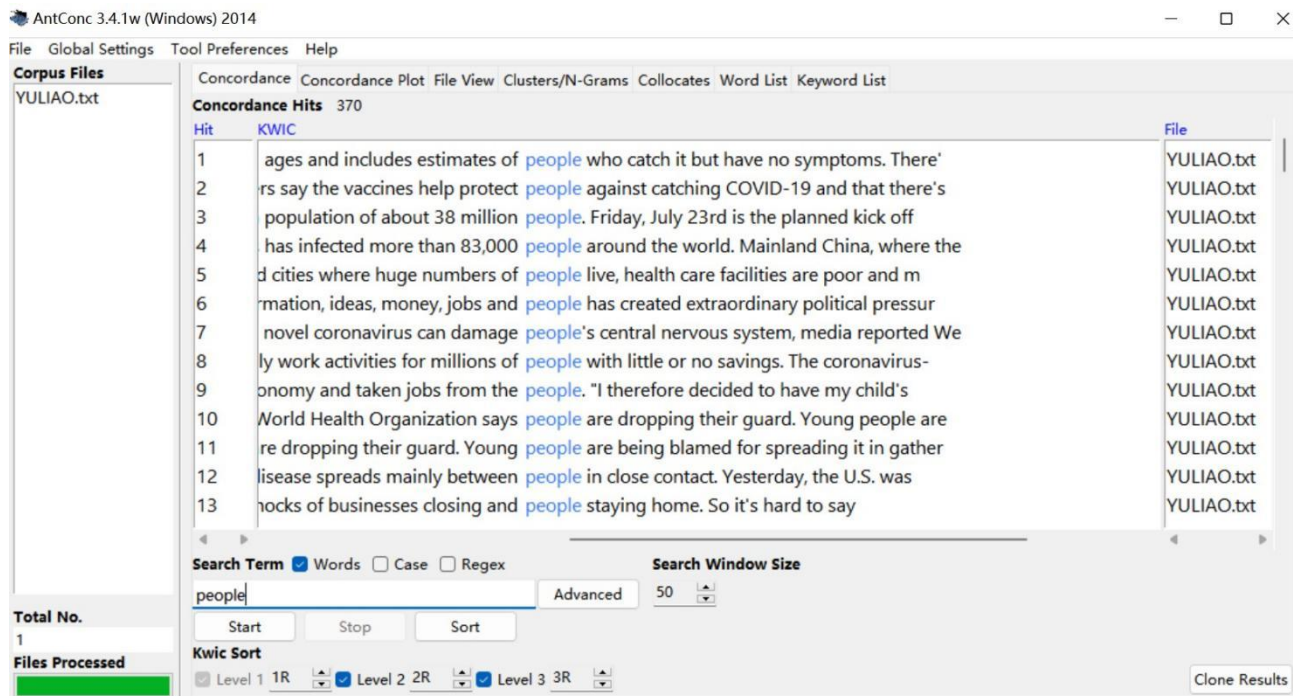


Table 4. The concordance output of “people” through AntConc

---

**Total No. of Cluster Types: 289**

---

**Total No. of Cluster Tokens: 370**

Rank	Freq.	Range	Cluster
1	12	1	people who catch
2	7	1	people who've
3	6	1	people who are
4	5	1	people around the
5	5	1	people have been
6	5	1	people have died
7	4	1	people are being
8	4	1	people in the
9	3	1	people are dropping
10	3	1	people have recovered
11	3	1	people in close
12	3	1	people living in
13	3	1	people not to

---

14	3	1	people who have
15	3	1	people will be
16	3	1	people worldwide and
17	2	1	people against the
18	2	1	people apart from
19	2	1	people are allowed
20	2	1	people are not

Table 5. Key cluster frequency distribution of “people”

The total number of “people” cluster types is 289 and the total number of clusters is 370. The list of keywords clusters about “people” reflect the severity of the pandemic and the serious impact on the people, as well as the reaction of the people to the pandemic. For example, the co-occurrence with these words, such as “people who catch”, “people who’ve”, “people are dropping”, and “people against”. It shows that CNN media is paying more attention to the serious problems and damage caused by the pandemic.

### Collocation

Comparative analysis of collocation words was carried out, focusing mainly on the characteristics of evaluative adjectives and evaluative verbs. The study also conducted an analysis of collocation words in discourse. If collocation is reused to form a fixed phrase, it means that people are thinking about the problem in the way the words are related. For example, people often use “lovely” to describe small babies and use “scary” to describe witches. Evaluative adjectives often reflect people's way of thinking and values. Therefore, the analysis of collocation words can deepen the understanding of words and reveal the ideology contained in them. In this part, “coronavirus”, and “government” will be selected to analyze and discuss the word combinations.

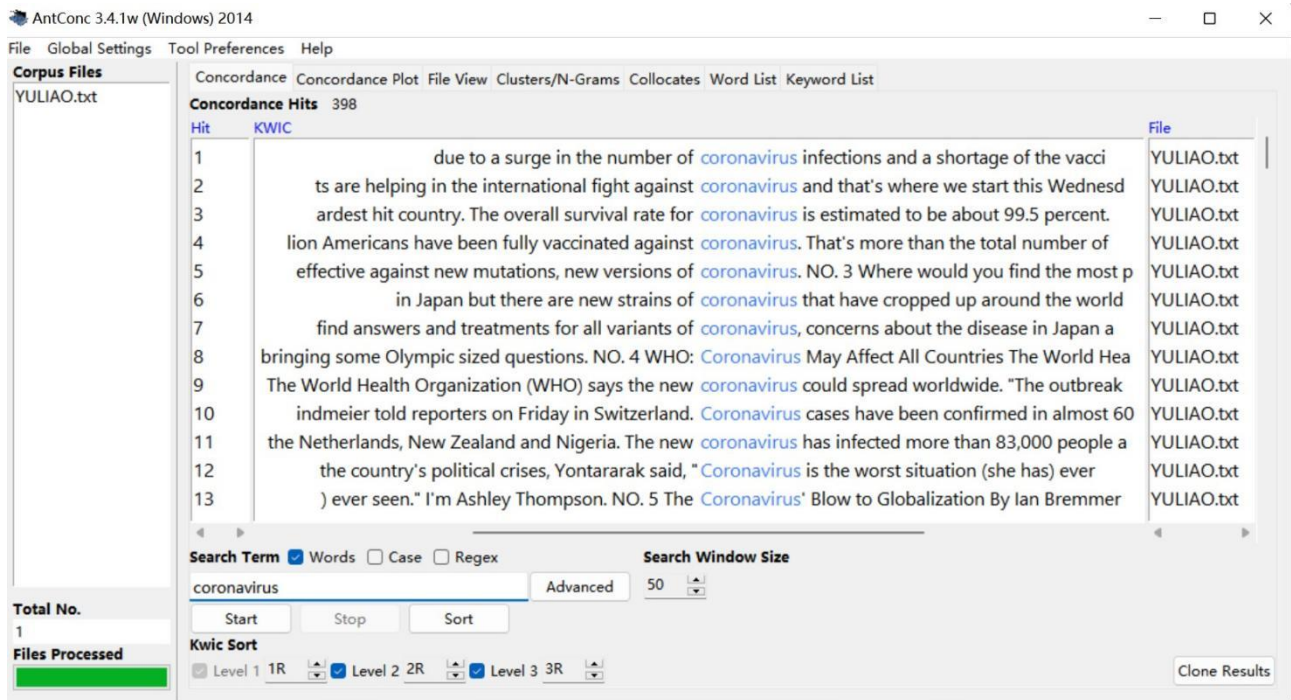


Table 6. The concordance output of “coronavirus” through AntConc

<b>Total No. of Collocate Types: 441</b>					
<b>Total No. of Collocate Tokens: 3390</b>					
<b>Rank</b>	<b>Freq.</b>	<b>Freq.(L)</b>	<b>Freq.(R)</b>	<b>Stat</b>	<b>Collocate</b>
1	331	237	94	3.76038	the
2	168	142	26	4.01422	of
3	92	59	33	2.98785	to
4	90	29	61	3.47738	in
5	68	22	46	2.86254	and
6	65	29	36	3.34113	a
7	62	53	9	5.05475	new
8	60	26	34	3.09269	s
9	51	0	51	5.98947	pandemic
10	49	6	43	3.49372	is
11	47	17	30	3.01983	that
12	46	10	36	4.38493	has
13	44	10	34	5.24279	cases
14	37	10	27	2.99519	it
15	33	25	8	5.33422	spread

Table 7. Collocation frequency distribution of “coronavirus”

Coronavirus is a high-frequency word in the corpus. Through the examination of its collocation vocabulary, it is possible to reveal intuitively the discursive tendencies and reporting positions of the writer. In CNN’s report, the words that collocate with “coronavirus” are adjectives that mainly describe the suddenness of the pandemic and the serious consequences, such as novel, bat-related, new, suspected, global, etc. Another word category is the verb, which expresses the measures that the government has been taken after the outbreak of the pandemic, such as “treat”, “fight”, “tackle”, “against”, “research”, “blames”, “develop”, etc. These expressions show that the government treats the pandemic more objectively and takes positive actions to respond effectively to the pandemic.



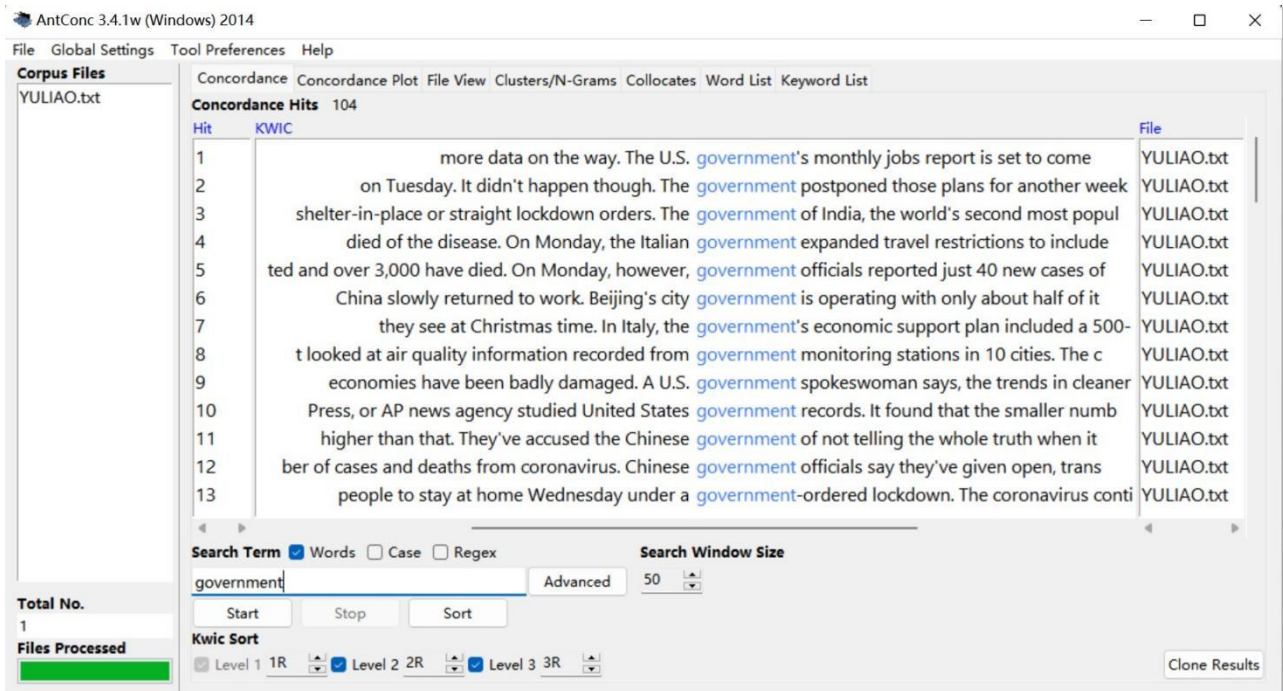


Table 8. The concordance output of “government” through AntConc

---

**Total No. of Collocate Types: 147**

---

**Total No. of Collocate Tokens: 755**

---

Rank	Freq.	Freq.(L)	Freq.(R)	Stat	Collocate
1	112	84	28	4.13323	the
2	51	27	24	4.79441	s
3	30	10	20	3.30736	to
4	25	9	16	4.04528	that
5	22	11	11	3.01752	of
6	17	17	0	5.50035	u
7	15	8	7	2.61816	and
8	14	5	9	3.52928	it
9	13	10	3	2.95539	a
10	11	1	10	3.27462	is
11	11	11	0	7.38251	federal
12	10	0	10	5.03202	says
13	10	3	7	3.44705	for

---

Table 9. Collocation frequency distribution of “government”

As one of the high-frequency words, the modifier of government often reveals the ideological tendencies of news reporters, especially some evaluation words that are more likely to reflect the tendency of the writing and the attitude of the author. In CNN, words that accompany “government” can be divided into two categories: the first is nouns, such as the U.S., Italian, Chinese, Indian, federal, etc. The other category is mainly adjectives, such as authoritarian, formal, official, many, etc. The collocation of “government” terms suggests that CNN is very concerned about the development of COVID-19 and the impact of the pandemic around the world.

## Conclusion

Based on the data obtained from the corpus resources, this study analyzes the relevant news reports about the pandemic in the CNN media from the perspective of CDA. The author analyzes the discourse characteristics from three aspects, word list, keyword clusters, and collocation. Based on the analysis results, three conclusions can be drawn as follows.

Firstly, CNN news reports mainly focus on the time, place, case, and impact that is related to the pandemic. It tends to report the measures that the government has taken after the outbreak of the pandemic in a timely and objective manner.

Secondly, the focus of CNN news is mainly on mortality, infection, and high incidence, and is concerned that the pandemic will affect the global economy, such as industry, tourism, stock markets, oil prices, etc.

Thirdly, in the analysis of collocation, the “government” and “coronavirus” that are associated with the pandemic were selected. The analysis of its concordance shows that CNN news reported that the government has taken plans and measures to defeat the outbreak of the pandemic.

## References

- [1] Stubbs, M. (1996). *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford: Blackwell.
- [2] Partington, A. (2003). *The Linguistics of Political Argumentation: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.
- [3] Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- [4] Baker, P. et al. (2008). *A useful methodological synergy. Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press*. *Discourse and Society*. 19 (3): 273-306.
- [5] Baker, P. & T. McEnery. (2005). *A corpus-based approach to discourses of refugees in UN and newspaper texts*. *Journal of Language and Politics*. 4 (2): 197-226.
- [6] McEnery, T. (2006). *Swearing in English: Bad Language. Purity and Power from 1586 to the Present*. London: Routledge.
- [7] Fairclough, N. G. Cortese & P. Ardizzone. (2007). *Discourse and Contemporary Social Change*. Bern: Peter Lang.
- [8] Qian Shufang. (2010). *Combining Corpus Studies with Critical Discourse Analysis*. *Foreign Language Teaching and Research*. 42(3): 40-44, 83.
- [9] Qian Yufang, Tian Hailong. (2011). *Discourse and Social Change in China: Taking Government Working Reports For Example [J]*. *Foreign Languages and Their Teaching*. (3): 43-46.
- [10] Tang Liping. (2011). *The Potentials of Corpus Linguistics in Strengthening Critical Discourse Analysis*. *Journal of Foreign Languages*. 34(4): 45-51.
- [11] Scott, M. & C. (2006). *Tribble. Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: Benjamins.
- [12] Gao Chao. (2004). *A Corpus-based Analysis of the Reportage of SARS*. *Foreign Language Education*. 25(4):41-44.

- [13] Wang Long. (2020). *A Comparative Study of Frames in The Guardian and China Daily News Reports about Coronavirus within Crisis Context*. Journal of Dongguan University of Technology, 27(4), 7.
- [14] Zhang Congying. (2022). *A corpus-based research on the New York Times' reporting and attitude towards the Coronavirus related to China*. Journal of Ningxia University(Humanities & Social Sciences Edition), 44(2), 9.
- [15] Lian Shaoying & Mi Ruoyu. (2021). *Wall Street Journal's Report on China's Response to the New Coronary Pneumonia Pandemic: A Corpus-based Critical Discourse Analysis*. Journalism Evolution (01), 40-52.