

# Sport or Politics? A Corpus-Based Critical Discourse Analysis of Chinese and American Media Coverage of the Paris 2024 Olympic Games

Yankai Liu

Department of Linguistics and Translation, City University of Hong Kong, Hong Kong SAR, China

Email: [yankailiu2-c@my.cityu.edu.hk](mailto:yankailiu2-c@my.cityu.edu.hk)

**How to cite this paper:** Liu, Y. K. (2024). Sport or Politics? A Corpus-Based Critical Discourse Analysis of Chinese and American Media Coverage of the Paris 2024 Olympic Games. *Open Journal of Social Sciences*, 12, 168-184.

<https://doi.org/10.4236/jss.2024.1210014>

**Received:** September 14, 2024

**Accepted:** October 19, 2024

**Published:** October 22, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

As two leading sporting nations who tied for gold medals in the 2024 Paris Olympic Games, how their media outlets depict such a mega sports event is worth exploring. This study aimed to investigate what and how the Chinese and American media outlets (China Daily and CNN) reported the 2024 Paris Olympic Games. In particular, the question of whether they regarded this quadrennial event as an opportunity to deliver politics-related topics to serve their own interests in every aspect also needed to be addressed. Corpus-based critical discourse analysis, a mixed methodological framework, was applied to conduct this study. Techniques from Antconc, including keywords, collocation, frequency, and lexical distribution, were employed on the two study corpora, which were compiled from the 2024 Paris Olympic Games news reports from China Daily and CNN. Based on the empirical evidence, the author concluded that the political topics in these two media outlets were mainly discussed, and the author demonstrated how they used specific language patterns to depict them. The author hopes to provide an efficient, operative, and comprehensive framework and methodological pattern for analyzing texts in social and political contexts by demonstrating the potential of combining corpus linguistics and critical discourse analysis.

## Keywords

The 2024 Paris Olympic Games, Corpus Linguistics, Critical Discourse Analysis, News Report

## 1. Introduction

The modern Olympic Games, established in 1894, aimed to foster individual

and collective goodwill and contribute to cultural exchanges and world peace. This mega sports event attracts large audiences worldwide to witness the world's foremost sports competition, where athletes compete under the principle of "Faster, Higher, Stronger." However, for most, this experience is only possible through the lens of the media. The media, by selecting what and how to report, wields significant control over the dissemination and representation of the Olympic Games, shaping public perception and understanding of the event. The media's role in highlighting the sports, athletes, performances, and communication and cultural exchanges between countries is crucial. As the father of the modern Olympic Games, Pierre de Coubertin, said, "free from political interference." However, the Olympic Games are never politically neutral, and sports have long been linked with politics. [Guttman \(2002\)](#) points out that the very introduction of the modern Olympic Games in Athens in 1896 was political, and little has changed since then. As the modern Olympics continue to develop, it becomes more political since the nations see it as an opportunity to showcase national power and promote ideologies ([Grix, 2013](#)). One example is that the recent Paris 2024 Summer Olympics were overshadowed by the political upheaval of the war in Ukraine, the Israeli-Palestinian conflict, gender and doping issues, and such politically related topics that naturally caught the attention of media outlets. China and America, as two leading sporting nations who tied for gold medals in the 2024 Paris Olympic Games, whether their media outlets only focused on the sports itself worth in-depth exploration. Both countries have invested heavily in sports programs to uplift the national spirit, showcase power, and increase their voices in international society; their media outlets would likely regard this mega sports event as a window to add political factors into reports to serve their interests.

Therefore, the general question of this study is whether Chinese and American media outlets included political content in the 2024 Paris Olympic Games news reports, what political topics were involved, and how they used language to depict these topics. To this end, the author built two study corpora, comprising news reports from China Daily (the largest English portal in China) and Cable News Network (the first and largest all-news television channel in the United States), to systematically analyze what and how they conveyed political information to the readers. By combining the corpus linguistics (CL) method with critical discourse analysis (CDA) to conduct this study, the author aimed to provide an example that demonstrates the immense potential of such methodological synergy. Moreover, an efficient, replicable, and comprehensive analysis pattern is expected for similar studies investigating media discourses.

This study is structured as follows: Section 2 discusses the theoretical feasibility of combining the methodologies of corpus linguistics (CL) with critical discourse analysis (CDA); Section 3 provides detailed methodological information on this study; Section 4 demonstrates the results and corresponding analysis; Section 5 concludes the study and discusses the implications for the future further study.

## 2. Literature Review

Corpus linguistics (CL) is a methodology that can be defined as a study of natural language on examples of real-life language use via a corpus, an extensive collection of text that is representative of a particular variety of languages and in a machine-readable form (McEnery & Wilson, 2001). CL can be applied to multiple aspects of linguistics, such as lexicography, pedagogy, sociolinguistics, language teaching, dialectology, psycholinguistics, pragmatics, cultural studies, and discourse analysis (McEnery, 2019). The wide application of CL methodology mainly credits to the natural strengths of the corpus-based approach. Biber et al. (1994) concluded that a corpus provides sufficient databases of naturally occurring discourse, allowing empirical assessments of an authentic language usage pattern, and when combined with computational tools, the corpus-based approach enables analyses of a scope unfeasible otherwise. Meanwhile, the corpus-based approach emphasizes quantitative data, including frequency counts and statistical measures, enabling the replication of research and verifying the statistical validity of the analysis's findings (McEnery & Gabrielatos, 2006). Such strengths indicate that the corpus-based approach enables the conclusion of convincing and repetitive linguistics patterns by providing real-life language instead of hypothesizing it, conducting quantitative analysis, and providing solid linguistic evidence.

Emerging from discourse analysis, critical discourse analysis (CDA) focuses on the relationships between power and inequality in language, integrates social-theoretical insights into discourse analysis, and advocates social commitment and interventionism in research (Blommaert & Bulcaen, 2000). CDA sees language as integrated into its sociolinguistic context; thus, it studies how social processes and occurrences are expressed through lexical or grammatical choices (Fairclough, 2013). Since its introduction, CDA has been used to study how language is used in various contexts, such as politics, media, law, and education, and to investigate how discourse is connected to social factors, including power dynamics, ideologies, social identities, etc. In this study, media discourse was specifically investigated because CDA provides a conceptual and analytical framework to examine how the media communicates meaning and constructs ideologies through linguistic selections so that they can be easily challenged (Sriwimon & Zilli, 2017). In fact, CDA has been identified as “a tool for deconstructing the ideologies of the mass media and other elite groups” (Henry & Tator, 2002: p. 72). Such features make CDA suitable for identifying the political content from the two study corpora in this study as CDA “studies the way social-power abuse and inequality are enacted, reproduced, legitimated, and resisted by text and talk in the social and political context.” (Van Dijk, 2015: p. 466)

The synergy of CL and CDA responds to the limitations of both fields, mainly the methodological weaknesses. One main criticism against the corpus-based approach is that it does not consider the text's contextual features. As Widdowson (1998) points out, corpus data is separated from the communicative context in which it was created, leaving the context all behind. In fact, Baker (2023) sees the

negligence of social-cultural contexts and language usage patterns observed from decontextualized examples as one of the most serious drawbacks of using the corpus-based approach. Such weakness is unavoidable because searching for the hidden semantic meaning and concluding in-depth interpretations from a large-scale text is unrealistic. Flowerdew (2005) especially points out that when it comes to pragmatic aspects of the text, which might only be retrieved from the socio-cultural context, the corpus analyst may find it extremely difficult to deal with due to the absence of contextual elements, highlighting the practical challenges of corpus-based analysis.

A repeated criticism against CDA is that language interpretations are politically rather than linguistically driven (Stubbs, 1997), and it is even regarded as an ideological interpretation instead of linguistic analysis (Widdowson, 1995). Therefore, the objectivity of CDA is challenged as too many political factors are involved, and the analysts may only draw conclusions they expect to see. In addition, the small data size in CDA research is also criticized as the analysis results may lack representativeness. Stubbs (1997) expresses his concerns over the limited and selectively chosen texts in the CDA study not being representative enough. Therefore, a small-scale CDA study may be unable to identify frequent linguistic patterns that represent strong discourses and ignore infrequent patterns that minority discourses (Baker et al., 2008).

Stubbs (1997) proposes that CDA should be strengthened by employing a broad corpus to generalize reliable patterns concerning language use. The synergy of CL and CDA focuses on the lexical patterns, which means the analysis can be quantified (Orpin, 2005). Moreover, CL procedures and techniques (like keywords, frequency, dispersion, collocation, clusters, and concordance analysis) can be applied in quantifying discursal phenomena that have already been identified by CDA, i.e., determining their absolute and relative frequencies in the corpus by looking at the various language strategies used to express them (Baker et al., 2008). By providing empirical evidence, the analysis results generated through corpus-based CDA could be systematically tested for reliability. In sum, combining CL with CDA helps exploit their strong points while canceling constraints, thereby making a more robust methodological framework to address related to the discursive representation of social wrongs and to discover new areas of public discourse for a more systematic analysis (Nartey & Mwinlaaru, 2019). For this study, the author aimed to reveal the ideologies hidden under the media discourse, thus further verifying the possibility and potential of such methodological synergy.

### 3. Methodology

#### 3.1. Toolkits and Analysis Steps

The author conducted his study with the help of Antconc (Version 4.3.1), a free-ware corpus toolkit for text analysis developed by Laurence Anthony. It is a multiplatform that provides tools, including the KWIC (Key-Word-In-Context) tool, plot tool, file tool, cluster tool, N-gram tool, collocate tool, wordlist tool, keyword

tool, and word cloud tool, thereby constituting a toolbox for corpus linguists. To answer the question of this study, keyword, collocate, KWIC, wordlist, and plot tools were applied, and such combination was expected to provide a feasible framework for future study.

### **3.1.1. Keyword Tool**

First, the keyword tool was used to examine whether politics-related topics were mentioned in the two study corpora. The keyword tool displays words that occur noticeably more (or less) frequently in the study corpus than the words in the reference corpus, allowing the user to identify characteristic words in the study corpus (Anthony, 2011). Keywords identify recurrent themes or topics in the texts in the target discourse domain (Egbert & Biber, 2019). The keyness of keywords, defined by Scott and Tribble (2006: pp. 55-56), “suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail.” Therefore, through keyword analysis, the author could select words that represent the themes of each corpus, which could be further analyzed to reveal implicit information that hides among a large-scale of texts.

Given that this study only focused on the semantic meaning of language, all grammatical words and irrelevant proper nouns (like names) from the keyword lists were excluded. The author selected the top 100 keywords of each study corpus and divided them into three categories: participants of the Paris 2024 Summer Olympics, sports terms of the Games, and specific terms of the Games. Each category was then divided into a few subcategories for more detailed analysis, and keywords in each categorized table were ranked in order of keyness. The author analyzed each keyword table and selected keywords that could be interpreted from political perspectives for further analysis, leading to findings that will be discussed in detail. Meanwhile, through the keyword analysis step, the author could get a general picture of what information was concluded in these two study corpora.

### **3.1.2. Collocate Tool and KWIC Tool**

Second, the author used the collocate and KWIC tools to understand how these topics were presented in the original contexts. Firth (1957) famously concluded that “you shall know a word by the company it keeps.” The primary purpose of the collocate tool is to demonstrate how the search term is used in the study corpus by listing co-occurring words. Words with a statistical tendency to co-occur in the text could be defined as collocation (Biber, 2009; Stubbs, 2001), and such lexical relations can be characterized as semantic compatibility (Evert, 2008). Therefore, knowing how a word acquires meanings by combining it with other words in different contexts, the author could reveal the language usage strategies of different media outlets and how they convey political content through different language patterns.

In this study, the author mainly focused on the semantic meanings of the selected keywords from the above analysis step. The aim was to investigate how China Daily and CNN used words to “company” the selected keywords. This focus

on the semantic meanings led to the exclusion of all grammatical words, as they did not contribute to understanding the keywords' usage. With a span of five words to the left and five to the right of the node words, all collocated words were ranked by their collocation strength using the log-likelihood statistic. The author then used the KWIC tool to observe the common usage of selected keywords in the original contexts, and corresponding sentences were listed as examples for a detailed demonstration and analysis. Analyzing how the collocation patterns were used in real contexts gave the author a clear view of how the political content in the two study corpora was demonstrated.

### 3.1.3. Wordlist Tool and Plot Tool

Wordlist and plot tools were applied in the final step to calculate the lexical frequency and investigate how the selected keywords were distributed in the two study corpora. This step could provide more detailed information on the linguistic choices of Chinese and American media outlets. Sinclair (1991) noted that those who study a text are likely to know how often each different word form occurs in it. The corpus-based lexical frequency technique adopted in this study provides a clear perspective to compare the amount of political information that audiences of the two study corpora are exposed to when they read news reports of the 2024 Paris Olympic Games. In this study, the author only focused on the horizontal comparison of the frequency of the selected keywords in the two study corpora. As Gries (2010) points out, a higher frequency of certain words in corpora does not necessarily mean that the words could be observed more frequently because the corpus sizes should also be considered. Therefore, converting raw frequency into normalized frequency was needed, and the standard of the norming basis was 10,000 words (demonstrating how often the selected keywords occur per ten thousand words in each corpus), considering the minimal words of the study corpora.

Though lexical frequency analysis indicates the relative importance of the selected terms in the two study corpora, the word may have a high frequency in the corpus not because it is widely used as a whole but because it is widely used in a small number of texts, parts of texts, or within the corpus (Baron et al., 2009). Therefore, how widely distributed the words are within a corpus should be considered to complete the conclusion further. The plot tool shows words in a 'barcode' format, allowing users to locate the position where words appear in the target corpus (Anthony, 2011). The plot's left edge is the corpus's beginning, and the right edge is the end. The search term is shown as a vertical line within the plot, and the vertical line's density illustrates the search term's frequency. In this study, news reports collected for building the two study corpora covered the 2024 Paris Olympics from the beginning till the end (details will be mentioned in the next section), which made it efficient for the author to observe the distribution and frequency change of the selected keyword as the event proceeded. Through this step, the author could draw a more comprehensive and reasonable conclusion on how these two media outlets reported the Games, especially regarding political content.

## 3.2. Study and Reference Corpora

### 3.2.1. Study Corpora

The interest in exploring the Chinese and American media outlets' attention on the 2024 Paris Summer Olympics motivates the author to build two study corpora specifically for this study. The author selected China Daily and CNN as his target media outlets for collecting texts. The written texts collected for building a corpus should be integral and stylistically homogeneous, especially if the corpus is designed for discourse analysis and text linguistics (Atkins et al., 1992). Therefore, all news reports related to the Paris 2024 Summer Olympics from 26 July 2024 (the opening data) to 11 August 2024 (the closing data) on chinadaily.com and edition.cnn.com were collected based on the time order (from the newest to the oldest).

### 3.2.2. Reference Corpus

As mentioned above, a reference corpus was needed to generate the keyword list. The author selected The Open American National Corpus (OANC) as the reference corpus, which is entirely open and unrestricted for any use. OANC is a massive electronic collection of American English containing texts of all genres from 1990 onwards. According to Oostdijk et al. (2013), the reference corpus should give a balanced representation of the standard language and its variations, allowing the researchers to examine language use in a particular domain or by a specific group. The genre of text materials in both corpora should be consistent to better compare the data from the study corpus to the reference corpus. Slate sub-corpus of OANC contains 4531 articles (4,238,808 words) from the Slate archives, including short articles on topics of politics, arts, business, sports, technology, travel, food, etc. Therefore, the Slate sub-corpus could be used as a balanced reference corpus in this research since it is stylistically consistent with the two study corpora and is large enough to generate a comprehensive keyword list.

## 4. Data Results and Analysis

### 4.1. Newsworthiness of the 2024 Paris Olympics Games

Before analyzing the linguistic features of the two study corpora, the author first examined the newsworthiness of the 2024 Paris Olympics Games in the coverage of the two media outlets. As revealed in **Table 1**, the comprehensive statistical analysis of the two study corpora uncovers some initial findings about the newsworthiness of the 2024 Paris Olympic Games. The number of CNN news reports on the Games only slightly surpasses those by China Daily. However, CNN's average report length significantly exceeds that of China Daily, and the total word count clearly demonstrates CNN's substantial coverage of the 2024 Paris Olympics. The CNN study corpus contains over four times the number of words than the China Daily study corpus. On average, CNN published 9.647 news reports related to the Games per day, compared to China Daily's 6.294.

**Table 1.** General statistics of the Paris 2024 Summer Olympics Corpora.

	China Daily	CNN
Number of Words	45,736	190,274
Number of news reports	107	164
Number of days collected	17	17
Average new reports per day	6.294	9.647
Average report length by words	427	1160

From a statistical perspective, China Daily and CNN paid significant attention to disseminating this mega sports event as multiple Olympics-related reports were published daily. It is surprising, however, that CNN has devoted much more space to reporting this event than China Daily. Considering that China has centralized its resources to achieve success in international sports events and outlet plans to develop itself into a leading sports power, one would expect the 2024 Paris Olympics to receive more significant media attention in China than in America. Though the data needs to be more comprehensive to make a comprehensive conclusion regarding the level of the 2024 Paris Olympics on the two media outlets' newsworthiness scale, the initial finding shows that the readers of CNN usually receive more information about this event than the readers of China Daily.

#### 4.2. Keywords Categorization and Analysis

**Table 2** illustrates the keywords related to the participants of the Paris 2024 Summer Olympics, and those words can be further sub-categorized: countries and places involved, people involved, and organizations involved. In the first sub-category, keywords indicate the participating countries in this Games (like China, USA, Belgium, and Spain) and cities that used to host the Olympic Games (like Beijing, Tokyo, and Rio). Among all those keywords, **Ukraine** could be interpreted from two perspectives: one is that **Ukraine** constantly appeared as a regular participant country, and one is that CNN mentioned the war in **Ukraine**. The second interpretation obviously makes CNN's reports political, and its readers' political stances are also influenced, especially for those sports fans who care less about such topics in their usual days, and such interpretation should be further verified.

Based on the second sub-category, the author can speculate that both media outlets emphasized the outcomes or achievements of the Game as keywords like **champion** and **medalist** ranked top in the keyword list. Noticeably, keywords related to organizations in the China Daily study corpus included **WADA** (World Anti Doping Agency) and **USADA** (United States Anti-Doping Agency), two anti-doping agencies, which could be interpreted from multiple angles: Chinese team went through doping suspicions during the Games; China Daily reported doping issues on other teams. Either interpretation involves political content that may influence readers' stances; thus, further in-depth study was needed to explore how these keywords are used.

**Table 2.** Keywords related to the participants of the Paris 2024 Summer Olympics.

	China Daily	CNN
Countries and places involved	Paris, China, France, Beijing, Tokyo, province, Seine, Rio, Belgium	Paris, France, Tokyo, USA, Seine (river), Spain, Ukraine, Eiffel, Australia, Brazil
People involved	Chinese, athletes, team, champion, women, medalist, athlete, fan, fans, delegation, men, teammate, swimmer, French, youth, player	team, athletes, athlete, women, medalist, French, champion, gymnast, swimmer, swimmers, sprinter, individual, gymnasts, Olympian, fans, teammates, Olympians, coach, crowd, organizers, men
Organizations involved	WADA, USADA, international, agency, global, BYD, platform, BMX	IOC

**Table 3** listed all keywords related to the sports terms of the Paris 2024 Summer Olympics, and those words were further sub-categorized under six themes: general sports terms, gymnastics terms, swimming terms, ball sports terms, other sports terms, and outcomes and achievements. Combining the results of the Games, the author could deduce that China Daily devoted more space to reporting swimming and table tennis, and gymnastics-related reports took more coverage on CNN, as China and America made remarkable achievements on these games, respectively. In addition, keywords from the outcomes and achievement subcategory could further prove the abovementioned assumption that both media outlets emphasized the results of the Games.

**Table 3.** Keywords related to the sports terms of the Paris 2024 Summer Olympics.

	China Daily	CNN
General sports terms	Olympic, Olympics, final, sports, sport, doubles, event, podium, competition, spirit, mixed, match, training, performance, (world) record, sporting, medley, events, relay, challenges, competing, title, finish, clinched, seconds, athletics, synchronized, semifinals, second, arena	Olympic, Olympics, final, sport, competes, meter, competition, match, meters, compete, event, sports, race, relay, finish, breaking, competing, podium, (world) record, game, finished, semifinal, arena, quarterfinal, tournament, new, summer, flag, performance, seconds, moment, percent, ahead, round, competitions, reacts, performs
Gymnastics terms	/	gymnastics, vault, floor, beam
Swimming terms	swimming, freestyle, singles, diving, aquatics, breaststroke, backstroke	swimming, freestyle, swim, water
Ball sports terms	tennis, table (tennis), badminton, rugby	tennis, basketball, netball, volleyball
Other sports terms	archery, skateboarding	triathlon
Outcomes and achievements	gold, medal, medals, bronze, silver, win, won, winning, ceremony, victory, championships, golds	gold, medal, bronze, medals, ceremony, won, win, celebrates, winning, silver, opening, championships, closing

Though keywords from **Table 4** are not related to the Games directly, they

could present a different story of the Games. Keywords in the Controversial issues subcategory describe one persistent troublesome issue, the drug issue, which could be used to accuse an innocent country for political purposes. In this study, China Daily used more keywords focused on this issue than CNN, indicating that it accused other countries or defended such accusations against the Chinese team. The usage of keyword **doping**, especially, should be further investigated since it appeared in both study corpora and could represent semantic meaning in different contexts. As for the technology and publicity subcategory, China Daily clearly seized the opportunity to promote its culture through multiple platforms, which is expected considering China has dedicated to boosting its international image and increasing its soft power.

**Table 4.** Keywords related to the specific terms of the Paris 2024 Summer Olympics.

	China Daily	CNN
Controversial issues	doping, anti, contamination, trenbolone	doping
technology and publicity	AI, story, Weibo, book, noodles	bra

In sum, the author could make his initial assumption of what these two media outlets chose to report for the 2024 Paris Olympic Games through keyword analysis. On the one hand, sports-related content, without doubt, was given the most space as both media outlets paid significant attention to the outcomes and achievements of the Games. On the other hand, political issues like the **Ukraine** war and the **doping** controversy were also included in the two study corpora. Keywords could be interpreted from a political perspective, indicating that such mega sports events were not politics-free; therefore, how these two media outlets used keywords to describe the political aspects of the Games is worth exploring.

### 4.3. Collocation Analysis

As mentioned above, two politic-related keywords, **Ukraine** and **doping**, could be interpreted from multiple perspectives; therefore, an in-depth study was needed to demonstrate how they were used and reveal implicit information. **Table 5** illustrates all words collocated with **Ukraine** in the CNN study corpus and ranked based on their collocation strength. From the list, the author could verify his second interpretation that CNN mentioned **Ukraine** for political purposes to a large extent. Collocation words like **war**, **invasion**, **support**, and **Russia** strongly indicated that CNN added political factors to convey its political stances and ideology in its news reports.

**Table 5.** Collocation list of **Ukraine** in the CNN study corpus.

CNN	
Ukraine	war, amid, ongoing, announced, committee, international, invasion, support, Russia, wars, Oksana, Livach, Cuba, besting, February, Ministry

Examples in **Table 6** gave a clear picture of how CNN used the **Ukraine** keyword in a real context, including accusing and condemning Russia for starting the war, causing **Ukraine** athletes to be absent from the 2024 Paris Olympics, and banning those athletes who supported the **Ukraine** war. It is no surprise that CNN demonstrated its pro-**Ukraine** political stance in its reports considering the diplomatic relationship between **Ukraine** and America; however, such content with obvious political bias may also influence the political stances of its readers, and against the “free from political interference” principle of Coubertin.

**Table 6.** Examples of **war**, **invasion**, **support**, and **Russia** that collocate with **Ukraine** in the CNN study corpus.

---

...injustice-Ukraine at war. My parents in Ukraine, Mykolaiv under bombing. I can't go to Olympics...

...evidence of 13 athletes' support for the war. Ukraine expected that a special commission reviewed...

...banned from the 2024 Summer Games for the invasion of Ukraine-though a small number of Russians...

...sports competitions or events where they support this war against Ukraine, they support the killing...

...“Principles of Participation” regarding Russia's invasion of Ukraine. The participation of Russian...

...prevent athletes who supported Russia's war against Ukraine from participating in the 2024 Olympics...

---

Words collocated with **doping** in both study corpora are illustrated in **Table 7**, respectively. One prominent finding from the list is that CNN seemed to devote significant space to connecting the keyword **doping** with China, which could be observed since **China**, **CHINADA** (China's Anti-Doping Agency), and **Chinese** are top collocation words. Examples from **Table 8** further verified the author's deduction as CNN accused the Chinese team of doping, especially targeting the swimming team. CNN created an image of the Chinese team violating the principles of fair games and indicating its achievements should be questioned and under censorship, which may mislead its readers and cause groundless defamation. Interestingly, China was the only country on the collocation list of the CNN study corpus, and such accusations by America are expected since the Chinese swimming team broke the record kept by the American team for dozens of years. By repeatedly connecting **doping** with **controversy** and **scandal**, calling for so-called **transparency**, and pointing its finger at China, CNN politicized the doping issue, which may influence the judgment of its readers.

Word **swimmers** in the collocation list of the China Daily study corpus indicated that the Chinese swimming team may be under doping suspicions, and examples from **Table 9** could further prove this standpoint. China Daily took a defensive position in its reports and illustrated the fact that Chinese swimmers were maltreated.

**Table 7.** Collocation list of **doping** in the China Daily and CNN study corpora.

Study corpus	
China Daily	anti, agency, violations, tests, cover, cases, false, world, swimmers, allegations, rigorous, narratives, urged, regulations, measures, prior, code, scrutiny, Switzerland, violation, rule
CNN	anti, agency, watchdog, doping, China, CHINADA, hotel, restaurant, controversy, WADA, Chinese, accepted, world, shortly, assessment, smear, scandal, cleared, sports, global, issue, suppress, casts, allegations, transparency, shadow, outrage, result, source, alleged

**Table 8.** Examples of **China**, **CHINADA**, and **Chinese** that collocate with **doping** in the CNN study corpus.

...event during the Paris 2024 Olympic Games on July 31. Related article China doping controversy casts a...

...swimmers had been cleared by the China Anti-Doping Agency (CHINADA) shortly before the Tokyo...

...US and its anti-doping agency. In a statement CHINADA said the latest news report “distorted the fact (s)...

...Fallout from Chinese doping scandal, Chinese swimmers are under an intense microscope in Paris, fairly...

...in the spotlight over alleged doping scandal The Chinese swimming team is at the center of a controversy...

**Table 9.** Examples of **swimmers** that collocate with **doping** in the China Daily study corpus.

...frequent than normal doping tests for Chinese swimmers before and during the Paris Games. Regarding...

...US anti-doping authorities targeting Chinese swimmers, who have been cleared of any wrongdoing in a...

...start affected by unfair anti-doping scrutiny, Chinese swimmers have capped their Olympic campaign...

To defend the Chinese team, China Daily responded to the “double-standard” acts of America by claiming that Chinese swimmers had undergone excessive drug **tests**, revealing the **cover-up** doping violations acts, and identifying the doping controversy as **false** accusations (see examples in **Table 10**). China Daily also took an offensive position by guaranteeing that Chinese swimmers had undergone **rigorous** drug tests, claiming that the Chinese team strictly followed international **regulations**, and it even **urged** America to stop groundless accusations and follow international rules (see examples in **Table 11**).

To sum up, both study corpora were not free from political content, and even worse, languages were politicized and used against other participating countries. Keyword **doping**, in particular, presented different semantic meanings and

conveyed different ideologies in these two study corpora. Regarding the **doping** topic, the author summarized that CNN mainly used language to accuse the Chinese team of violating the rules of anti-doping regulations while making no response to the doping and cover-up accusations posed against the American team. China Daily took a more comprehensive strategy of proving its innocence, going against accusations from America, and questioning whether the American team strictly followed the anti-doping regulations. Such unbalanced usage of the keyword **doping**, therefore, should be further studied to investigate the importance of this theme in the study corpora.

**Table 10.** Examples of **tests**, **cover**, and **false** that collocate with **doping** in the China Daily study corpus.

---

...a higher-than-normal number of doping tests prior to, and during, the Games, as well as fielding questions...

...Chinese swimmers have been subjected to an average of 21 anti-doping tests each, nearly four times the...

...United States Anti-Doping Agency's cover-up of anti-doping rule violations after the World Anti-Doping...

...USADA's anti-doping work. The move is, in effect, a cover-up of anti-doping rule violations under the...

...Chinese swimmers and some of their foreign opponents that stemmed from false doping accusations by...

...such as The New York Times, has been making false doping accusations targeting two Chinese swimmers...

---

**Table 11.** Examples of **rigorous**, **regulations**, and **urged** that collocate with **doping** in the China Daily study corpus.

---

...Pan's record-breaking performance on Wednesday came after having completed a rigorous doping test...

...need for a more transparent, equitable, and scientifically rigorous approach to anti-doping measures. It...

...domestic anti-doping efforts stayed in line with international regulations and the global anti-doping code...

...all Chinese athletes respect and honor anti-doping regulations, stick to strict doping control routines, and...

...25. The China Anti-Doping Agency urged the United States on Tuesday to stop creating false narratives,...

...and appropriate, and urged the United States anti-doping authorities to keep up with international rules...

---

#### 4.4. Normalized Frequency and Lexical Distribution

Though keyword **doping** ranked top in the collocation list of both study corpora,

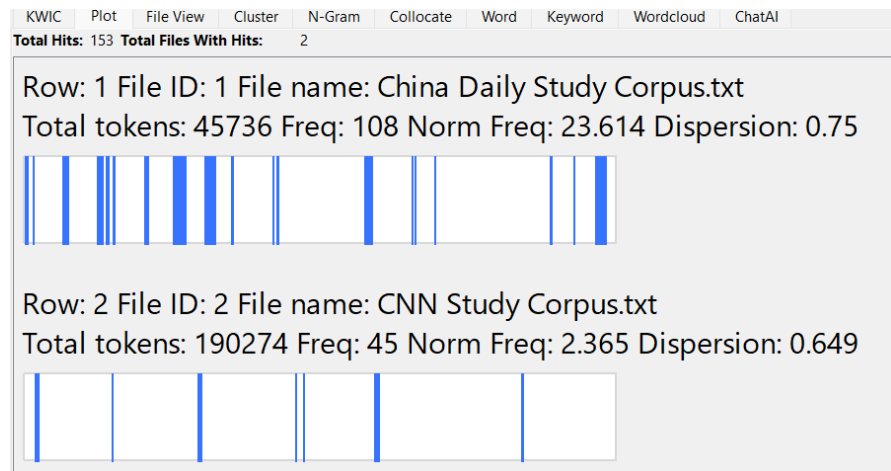
the normalized frequency, which accounts for the different sizes of the two corpora, is crucial for horizontal comparison. The statistical frequency information on keyword **doping** can be observed in **Table 12**, thus demonstrating the relative importance of the **doping** theme in the coverage of the 2024 Paris Olympics across China Daily and CNN. The most obvious observation is that keyword **doping** occurs much more frequently in the China Daily study corpus—almost ten times higher than in the CNN study corpus. China Daily emphasized this topic more than CNN, which is unsurprising since explanations to clean the defamation of **doping** violations and accusations against the American team could take more coverage of the reports. In other words, CNN downplayed the importance of the **doping** theme, which could be interpreted as avoiding making justifications for violating the anti-drug regulations and covering up such acts. The strategy adopted by CNN also ensured that its readers could only get one side of the story, that the Chinese team won the games “unfairly,” without knowing that the same accusations also fell on the American team. Therefore, how the two media outlets use the keyword **doping** is undoubtedly not only a linguistic consideration but also political purposes.

**Table 12.** Normalized frequency of **doping** in the China Daily and CNN study corpora.

Study corpus	Total number in the corpus	Normalized frequency (per 10,000 words)
China Daily	108	23.614
CNN	45	2.365

How the keyword **doping** was distributed in the China Daily and CNN study corpora was demonstrated in **Figure 1**. Notably, **doping** was distributed evenly at every stage of the Games at the two study corpora. However, it was more widely distributed in the China Daily study corpus, suggesting that **doping** was a significant concern for China Daily throughout the event. The dense concentration of **doping** at the right edge of the China Daily plot also suggests that **doping** issues were reported frequently, even at the early stage of the event. This could imply that the Chinese team was accused of **doping** even before the start of the games. As the Games proceeded, the occurrence of **doping** was more frequent in the China Daily study corpus than in CNN, as indicated by the density of plots. It is possible that as the Chinese team continued to achieve success in the games, more accusations against it arose, and it took more effort and coverage to respond.

Through the lexical frequency and distribution analysis, the author concluded that China Daily devoted more space to **doping** issues than CNN at every stage of the Games, especially at its early and late stages. Correspondingly, readers of China Daily are more likely to be exposed to reports related to **doping** issues, which enhances their awareness of doping topics and underscores the media’s role in shaping public perception of **doping** issues.



**Figure 1.** Distribution of keyword **doping** in the China Daily and CNN study corpora.

## 5. Conclusion

The primary goal of this study is to conduct a corpus-based critical discourse analysis of the coverage of the 2024 Paris Olympic Games in two study corpora compiled from the relevant news reports from China Daily and CNN. The analysis results answered the questions of what political topics were discussed in the 2024 Paris Olympic Games news reports and how they used language patterns to present such topics. In general, both media outlets chose to add political factors into their reports of the 2024 Paris Olympic Games, and **doping** was a commonly discussed issue. Differently, CNN tended to use such a topic as a weapon to accuse other participant countries, especially China. China Daily took a more defensive stand on this issue and devoted more space to responding to **doping** accusations. CNN also included the **Ukraine** war in its reports and demonstrated its pro-**Ukraine** political stance by depicting **Ukraine** as a victim and condemning Russia for starting the war.

Such a conclusion was made by employing keyword, collocation, lexical frequency, and distribution techniques, and the mixed methodology and analysis procedures involved in this study provide a feasible pattern to conduct similar studies. The potential of integrating corpus linguistics with critical discourse analysis enlightens the author that such methodological framework could be applied to future large-scale projects, therefore providing an efficient analysis method to draw a more comprehensive conclusion, thereby ensuring the validity and reliability of the study's findings.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Anthony, L. (2011). *AntConc (Windows, Macintosh OS X, and Linux)*.  
 Atkins, S., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic*

- Computing*, 7, 1-16. <https://doi.org/10.1093/llc/7.1.1>
- Baker, P. (2023). *Using Corpora in Discourse Analysis*. Bloomsbury Publishing.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (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 & Society*, 19, 273-306. <https://doi.org/10.1177/0957926508088962>
- Baron, A., Rayson, P., & Archer, D. (2009). Word Frequency and Key Word Statistics in Historical Corpus Linguistics. *Anglistik: International Journal of English Studies*, 20, 41-67.
- Biber, D. (2009). A Corpus-Driven Approach to Formulaic Language in English: Multi-Word Patterns in Speech and Writing. *International Journal of Corpus Linguistics*, 14, 275-311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-Based Approaches to Issues in Applied Linguistics. *Applied Linguistics*, 15, 169-189. <https://doi.org/10.1093/applin/15.2.169>
- Blommaert, J., & Bulcaen, C. (2000). Critical Discourse Analysis. *Annual Review of Anthropology*, 29, 447-466. <https://doi.org/10.1146/annurev.anthro.29.1.447>
- Egbert, J., & Biber, D. (2019). Incorporating Text Dispersion into Keyword Analyses. *Corpora*, 14, 77-104. <https://doi.org/10.3366/cor.2019.0162>
- Evert, S. (2008). Corpora and Collocations. In A. Lüdeling, & M. Kytö (Eds.), *Corpus Linguistics* (pp. 1212-1248). Mouton de Gruyter. <https://doi.org/10.1515/9783110213881.2.1212>
- Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. Routledge.
- Firth, J. (1957). A Synopsis of Linguistic Theory, 1930-1955. In Philological Society (Ed.), *Studies in Linguistic Analysis* (pp. 10-32). Blackwell.
- Flowerdew, L. (2005). An Integration of Corpus-Based and Genre-Based Approaches to Text Analysis in EAP/ESP: Countering Criticisms against Corpus-Based Methodologies. *English for Specific Purposes*, 24, 321-332. <https://doi.org/10.1016/j.esp.2004.09.002>
- Gries, S. T. (2010). Useful Statistics for Corpus Linguistics. In A. S. Pérez, & M. A. Sánchez (Eds.), *A Mosaic of Corpus Linguistics: Selected Approaches* (pp. 269-291). Peter Lang.
- Grix, J. (2013). Sport Politics and the Olympics. *Political Studies Review*, 11, 15-25. <https://doi.org/10.1111/1478-9302.12001>
- Guttmann, A. (2002). *The Olympics: A History of the Modern Games*. University of Illinois Press.
- Henry, F., & Tator, C. (2002). *Discourses of Domination: Racial Bias in the Canadian English-Language Press*. University of Toronto Press. <https://doi.org/10.3138/9781442673946>
- McEnery, A., & Wilson, A. (2001). *Corpus Linguistics* (2nd ed.). Edinburgh University Press.
- McEnery, T. (2019). *Corpus Linguistics*. Edinburgh University Press.
- McEnery, T., & Gabrielatos, C. (2006). English Corpus Linguistics. In B. Aarts, & A. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 33-71). Blackwell Publishing. <https://doi.org/10.1002/9780470753002.ch3>
- Nartey, M., & Mwinlaaru, I. N. (2019). Towards a Decade of Synergising Corpus Linguistics and Critical Discourse Analysis: A Meta-Analysis. *Corpora*, 14, 203-235. <https://doi.org/10.3366/cor.2019.0169>
- Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In P. Spyns, & J.

- Odijk (Eds.), *Essential Speech and Language Technology for Dutch* (pp. 219-247). Springer. [https://doi.org/10.1007/978-3-642-30910-6\\_13](https://doi.org/10.1007/978-3-642-30910-6_13)
- Orpin, D. (2005). Corpus Linguistics and Critical Discourse Analysis: Examining the Ideology of Sleaze. *International Journal of Corpus Linguistics*, 10, 37-61. <https://doi.org/10.1075/ijcl.10.1.03orp>
- Scott, M., & Tribble, C. (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.22>
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sriwimon, L., & Zilli, P. J. (2017). Applying Critical Discourse Analysis as a Conceptual Framework for Investigating Gender Stereotypes in Political Media Discourse. *Kasetsart Journal of Social Sciences*, 38, 136-142. <https://doi.org/10.1016/j.kjss.2016.04.004>
- Stubbs, M. (1997). Whorf's Children: Critical Comments on Critical Discourse Analysis (CDA). *British Studies in Applied Linguistics*, 12, 100-116.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. John Wiley & Sons.
- Van Dijk, T. A. (2015). Critical Discourse Analysis. In D. Tannen, H.E. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (pp. 466-485). Wiley-Blackwell. <https://doi.org/10.1002/9781118584194.ch22>
- Widdowson, H. G. (1995). Discourse Analysis: A Critical View. *Language and Literature: International Journal of Stylistics*, 4, 157-172. <https://doi.org/10.1177/096394709500400301>
- Widdowson, H. G. (1998). Context, Community, and Authentic Language. *TESOL Quarterly*, 32, 705-716. <https://doi.org/10.2307/3588001>