

Text Mining and Cognitive Psychological Analysis of Sports News in China Daily: The Case of the 2024 Olympics

Kaige Song^{1*}

Abstract

Sports news studies have been an increasingly important area. China Daily's sports news in English version serves as a window, through which people in English countries can perceive China and the athletes' images. This paper combines the quantitative method with the qualitative method so that the findings can be more persuasive. This study collects the data from China Daily's news reports ranging from July 26th, 2024 to August 11th, 2024, and builds a corpus to conduct text mining. The high-frequency word cloud is generated to observe and examine the prominent words. Their diachronic distributions are also explored. By the comparison of British National Corpus (BNC), the top 20 keywords are retrieved to address the prominent meanings. This paper also deals with the collocation analysis of "Chinese" and "athletes". Based on the reading of the concordance lines and contexts, the meanings of China and its athletes' images are interpreted. It is found that these pieces of news construct China as a high-tech, progressive, friendly, and forward-thinking image and athletes as a victorious, gracious, and sportsmanlike image. These findings are further interpreted by the qualitatively cognitive psychological analysis. This study has implications for interdisciplinary studies on sports psychology.

Keywords: Sports News, Text Mining, Cognitive Psychology, Corpus.

Introduction

Sports news serves as a bridge, by which readers have access to the latest information about distinguished athletes and great sports events. The sports news on the Olympics is typical and significant. The 2024 Summer Olympics, officially known as the Games of the XXXIII Olympiad, are set to take place in Paris, France. This will mark the third time the city has hosted the Olympics, following 1900 and 1924. This event is not just about showing athletic strength but also about highlighting cultural heritage, technological advancements, and environmental sustainability. Paris has been billed as the most inclusive and sustainable Olympics ever, with initiatives aimed at reducing the environmental impact and ensuring gender equality in events. Frankly, it has not been long since the 2024 Summer Olympics was held, so very few studies on this important event have been conducted. However, some research gaps in news discourse studies give this paper some implications. News discourse has been explored in terms of various theoretical views, such as critical discourse analysis (Bednarek & Caple, 2014), corpus linguistics (Collins & Nerlich, 2015), and psychology (Pan, 2023). One of the research gaps in critical discourse analysis is that this theory itself is too selective, partial, and qualitative (Machin & Mayr, 2023). For corpus linguistics, some corpus tools have been utilized to get quantitative data, but they fail to do well in

the interpretations of data results (Baker, 2012). In addition, it has been proved that psychology has the feasibility to align with corpus linguistics (Ellis, Simpson - Vlach, & Maynard, 2008). However, the relevant studies are not enough, and less attention has been paid to the combination of cognitive psychology and quantitative studies. To address these gaps, this paper seeks to combine text mining and cognitive psychology.

The sheer scale of the 2024 Paris Olympics will generate news at an unprecedented rate, creating opportunities and challenges for data researchers. This is where data mining comes into play, serving as a critical tool for exploring textual features from the big data. Data mining refers to "the process of discovering interesting patterns, models, and other kinds of knowledge in large data sets" (Han, Pei, & Tong, 2022). It is an essential tool for making sense of large volumes of data. There is plenty of data about the 2024 Paris Olympics, which makes it possible to conduct data mining about the Olympics. Data mining offers significant advantages, particularly in handling large and complex datasets. One key benefit is its ability to identify hidden patterns and tendencies that may not be immediately apparent through manual analysis. By uncovering these words and patterns, the relevant organizations can gain valuable insights that inform decision-making processes. Data mining also excels in predictive analytics, where it uses historical data to forecast future trends or outcomes. This predictive capability is

¹ School of Foreign Studies, Jilin University, Changchun, 2699 Qianjin Street, Changchun, Jilin 130012, People's Republic of China.

*Correspondence: kgsong@126.com

invaluable in industries such as healthcare (Harutyunyan et al., 2019), finance (West & Bhattacharya, 2016), and sports (Pino-Ortega et al., 2021), where anticipating future behavior or events can lead to better planning and strategic actions. Moreover, data mining enhances efficiency by automating the process of data analysis, saving time and resources that would otherwise be spent on manual data processing. This automation also enables real-time data analysis, allowing for faster and more accurate decision-making. Additionally, data mining helps in optimizing operations and improving resource allocation, as it provides a clearer understanding of where inefficiencies lie and what changes can drive better outcomes (Jo, 2019).

Cognitive psychology is the branch of psychology that studies mental processes such as perception, memory, problem-solving, language, and decision-making (Goldstein, 2015). It focuses on understanding how people acquire, process, store, and use information. Cognitive psychology emerged as a response to behaviorism, which focused solely on observable behaviors, by emphasizing the importance of internal mental states and processes. Cognitive psychology helps explain how readers comprehend written material, process the meaning of words and sentences, and integrate new information with existing knowledge.

This paper seeks to use data mining to analyze the sports news about the 2024 Paris Olympics. The data sources from China Daily. By the application of the quantitative method of data mining, this study intends to investigate the national image and the athletes' images that China Daily conveys. Except for text analysis, this paper also conducts cognitive psychological analysis. This research has great research significance. First, it makes news media reflect on the way how to use discourse appropriately to construct national and athletes' images. Second, it helps readers read news critically. Third, it enriches the interdisciplinary explorations of sports psychology.

Literature Review

A large and growing number of studies have set foot in sports and psychology. This section attempts to summarize and criticize previous studies in terms of sports and cognitive psychology.

Previous studies on sports have three significant features. Firstly, an increasing tendency of interdisciplinary studies has emerged. Chambers et al. (2015) align sports with engineering technology. They conduct a systematic review of the use of microsensors to evaluate sport-specific movements. It is found that the ability of microsensors can give more detail on athlete demands

and performance. Their study provides a new perspective for sports science, which helps coaches give athletes constructive feedback on correct and incorrect techniques. (Lavandera et al., 2017) introduce sports psychology, which combines sports and psychology. They retrospect the history of sports psychology and point out some landmark scholars and events. Their study does much good to the development of sports psychology. Secondly, sports news deserves more attention. Some studies have explored sports news in two ways. One is text analysis. The other is readers' experimental studies. Harrison (2018) focuses on women in the United States' electronic sports media. With post-feminist and Foucauldian discourse analysis as the theoretical foundation, he finds out that sports media treat women with double standards. The interpretation of Post-feminist and Foucauldian discourse analysis enhances the persuasiveness of his findings. Different from the general text analysis, Graefe et al. (2018) take a fresh view of readers' perceptions. By conducting an online experiment, they examine the credibility and readability of computer-written news about sports and finance. Their participants add up to approximately 90,000 active members, so they get abundant quantitative and qualitative data. The study of sports news holds significant research value as it explores the intersection of media, culture, and society. Sports news shapes public perception of athletes, events, and broader societal issues, such as gender, race, and politics. By analyzing how sports are covered, researchers can uncover biases, narrative framing, and media influence on public opinion. Additionally, sports news provides insight into the commercial and entertainment aspects of sports, while also reflecting societal values and trends. Understanding its role can help enhance media literacy, improve journalistic standards, promote more equitable and accurate sports reporting, and help mold a positive national image. At the same time, sports news still leaves much room to explore. Therefore, sports news should receive more attention. Thirdly, the quantitative method turns out to be rare. The qualitative approach is the main research method in most studies on sports news. One of the representative studies is Iosifidis and Nicoli (2020). Their study addresses Facebook's public announcements about the way it handles disinformation. The qualitative content analysis provides this study with the advantage of closely reading texts. Some studies employ the quantitative method to deal with sports news. Ferrer-Conill (2017) points out that a quantitative turn has been emerging in contemporary journalism. Zhang and Caple (2021) use the quantitative method of corpus-assisted

discourse analysis to analyze news reports about the Chinese tennis player Li Na. The corpus tool AntConc is used to retrieve keywords, concordance, and collocation. It indicates that international media tends to emphasize her individualism while Chinese media highlights her collectivism and responsibility towards the nation. The quantitative data provide empirical support for this finding. High-frequency words, collocations, and keywords in corpus linguistics are very useful for exploring language features (Csomay & Crawford, 2024; McEneaney & Brookes, 2024; O’Keeffe & McCarthy, 2010). Although previous sports studies have made much progress in theories and methods, more studies on sports news should combine quantitative data analysis with readers’ acceptance. Cognitive psychology is interested in the perception of readers or audience. This paper hence provides an overview of cognitive psychology.

Cognitive psychology has long focused on understanding the mental processes that underlie human thought, perception, memory, and language use (Barsalou, 2014). A large body of literature examines how individuals process and interpret information, particularly concerning reading and text comprehension. Traditional cognitive psychology studies have emphasized experimental methods to explore how memory works, how attention is directed, and how language is processed. For instance, researchers have studied how working memory supports reading comprehension by allowing readers to retain and manipulate information as they move through a text (McMaster, Espin, & Van Den Broek, 2014). Similarly, attention has been studied in terms of its role in helping readers focus on relevant details while filtering out irrelevant information (Fernandez - Duque & Johnson, 1999). However, most of this research has been conducted in controlled settings, often with relatively simple texts, and has not always accounted for the complexity of real-world media environments, such as the context in which readers engage with news texts. There is growing recognition that cognitive psychology could benefit from expanding its scope to incorporate more diverse forms of media and texts, particularly in the domain of news discourse. News is a unique type of text that carries specific social and cultural meanings, and how individuals cognitively engage with news involves not only reading comprehension but also the interpretation of social and political contexts. Cognitive psychology can provide insight into individual cognitive strategies, background knowledge, and perceptions, while textual analysis can reveal the structural and linguistic features of news that influence comprehension and interpretation. Integrating

these methods would allow cognitive psychologists to investigate how specific text features interact with readers’ cognitive processes. Furthermore, cognitive psychological research has largely overlooked the domain of news discourse, particularly in examining how journalistic language can shape readers’ perceptions and judgments. Incorporating a more focused study of news discourse would strengthen the field by exploring how cognitive mechanisms interact with the social dimensions of news consumption. This would also offer valuable insights into how individuals make sense of complex social and political issues through news, an increasingly important area of research in an era of information overload and media fragmentation.

Although studies on sports and cognitive psychology have enriched data analysis techniques and theoretical views, two gaps remain to address. On one hand, there is a paucity of complementary validation studies on quantitative and qualitative methods. On the other hand, national and athletes’ images in sports news leave much room to explore.

Research Methodology

This paper combines data mining with readers’ cognition. Data mining has been proved to be effective in examining textual features (Xu, 2024), educational differences (Ocumpaugh et al., 2014), and tourists’ cognition (Zhang, Chen, & Li, 2019). It covers various ways, such as corpus (Liu, Zhang, & May, 2022), sentiment analysis (Yadav & Vishwakarma, 2020), topic modelling (Re Lee et al., 2023), and so on. One of the main advantages of data mining is that it can deal with big data. Those data are almost impossible for humans to read line by line. Moreover, quantitative results are “visible”, so that the findings become clearer and more dependable. Therefore, quite a few disciplines and researchers benefit from data mining. The corpus method is the main way of text mining. To explore the national and athletes’ images in China Daily’s news reports, this study builds a corpus that consists of texts about the 2024 Paris Olympics. Some tools will aid in examining the language features. Word cloud, keywords, and collocations are the main pointers to uncover the national images and athletes’ characteristics. However, textual features need further verification in terms of readers’ mental perception. The alliance between data mining and psychology has been a promising tendency (Mariani, Perez - Vega, & Wirtz, 2022). Therefore, this study also conducts a cognitive psychological analysis for China Daily’s readers to elucidate the findings of data analysis.

In the word cloud, those high-frequency words are very prominent. The top 20 high-frequency content words are as follows: Olympic (689), Paris (672), said (618), Chinese (587), games (497), Olympics (451), gold (434), China (431), world (388), team (366), 2024 (302), medal (281), athletes (278), final (268), time (253), sports (253), year (250), China's (237), women's (209), event (187). The high-frequency words in China Daily's reports of the 2024 Paris Olympics reveal several important aspects of the newspaper's reporting focus and broader narrative strategies. The most frequent term, "Olympic" (689), along with closely related terms like "Olympics" (451) and "games" (497), underscores the centrality of the event itself. This suggests that the overall framing of the coverage is squarely focused on the global importance of the Olympics as a major international sporting event. The near-equal frequency of "Paris" (672) and "Olympic" indicates a strong emphasis on the host city, likely reflecting coverage of the cultural significance of holding the Olympics in Paris. The frequent mention of "Paris" may also point to a broader interest in how the host city is presenting itself on the world stage, aligning with global attention on the venue and its preparations. The prominence of words like "Chinese" (587), "China" (431), and "China's" (237) indicates a focus on China's participation and achievements in the games. This suggests that China Daily is using the Olympics not only as an opportunity to report on the event itself but also as a platform to highlight China's role and success in the global sports arena. The frequent appearance of terms like "gold" (434), "medal" (281), and "athletes" (278) supports this interpretation, as it points to a focus on the competition and success narratives, particularly in terms of medal achievements. This aligns with China's long-standing tradition of using the Olympics to boost national pride and project a strong international image, which is a recurring theme in Chinese media when covering major global events. Additionally, the frequent mention of "team" (366) and "final" (268) suggests an emphasis on collective achievements and key moments in the competition, further reinforcing narratives of national accomplishment and sportsmanship. Interestingly, the relatively high frequency of the term "said" (618) points to the reporting style used by China Daily, likely relying heavily on quotations from athletes, officials, and commentators. This approach suggests that the newspaper is using direct voices to enhance the authority and relatability of its coverage. Quotations from Chinese athletes or officials, as well as international voices, could serve to present a balanced

narrative that incorporates both domestic and global perspectives on the games. The appearance of terms like "world" (388) and "sports" (253) suggests that while China's participation is a significant focus, the broader context of global competition is also emphasized. The use of the word "world" likely reflects China Daily's attempts to frame China's achievements within the global sports community, highlighting both the country's position in international sports and the universal significance of the Olympics as a unifying event. Additionally, the frequent mention of "women's" (209) suggests that there is attention being given to gender representation, possibly aligning with the 2024 Paris Olympics' emphasis on gender equality, as this edition of the Olympics is set to have equal numbers of male and female athletes for the first time. The presence of "2024" (302) and "year" (250) shows that the reporting is not just focused on the immediate moment but also contextualizes the Olympics within a broader timeline, perhaps reflecting on how the games fit into a historical narrative or future sporting ambitions. This temporal dimension may also indicate that China Daily is positioning the Paris Olympics as part of a longer trajectory of China's growing prominence in international sports, looking forward to future successes.

Indeed, high-frequency words can reflect the prominent meanings in the sports news. However, frequency only means the overall occurrences. The diachronic distribution in a period of time cannot be dealt with. This study aligns with (Wang, Liu, & Tu, 2023), which combines data mining with a diachronic investigation. Accordingly, with the help of Voyant tools, this study also makes the diachronic distribution of the top 5 words visualized (see Figure 3). The word frequencies of all the pieces of news are normalized.

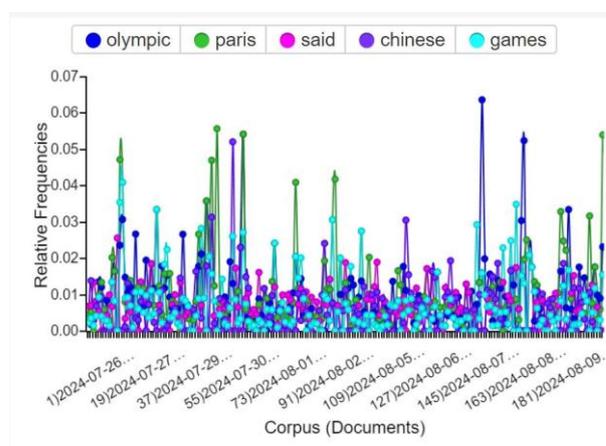


Figure 3: Diachronic Distribution of the Top 5 High-frequency Words.

In [Figure 3](#), “Olympic” is very frequently mentioned from August 6th to August 8th. “Paris” almost is highly frequent almost every day. Relatively speaking, “said” and “Chinese” appear less frequently but also occupy every day. The word “games” is the most frequent on July 26th, the day of the opening ceremony. By reading the contextual concordances of these words in those pieces of news with high-frequency words, some meaningful findings can be reached. The term “Olympic” appears in various contexts related to the athletes’ performances, challenges, and aspirations during the 2024 Paris Olympics. The first occurrence refers to the “Olympic history”, highlighting the narrow margin of defeat in a historical context. It then connects to “Olympic ordeals”, emphasizing the struggles and intense competition faced by Chinese athletes. The phrase “defending Olympic champion” underscores the prestige and pressure of retaining an Olympic title, as seen with Zhang Yufei. Additionally, “Olympic 100m butterfly final” references the specific event in which Zhang competed, underscoring the significance of her performance. Overall, “Olympic” is used to convey the weight of the athletes’ experiences, symbolizing both the honor and pressure of competing at the highest international level. The term “Paris” is consistently used in the context of the city’s role as the host of the 2024 Olympics and its expected economic and tourism-related benefits. “Paris” is linked to the massive influx of international tourists, highlighting how the city will reap significant financial gains from the event. The term also appears when discussing the long-term impact on tourism, as hosting the Olympics will enhance Paris’ global brand and infrastructure, contributing to sustained tourism growth. The overall use of “Paris” emphasizes the city’s transformation into a global tourism hotspot, with the Olympics serving as a catalyst for future economic and cultural benefits. In the July 26th news report, the word “said” is used in the context of quoting key figures in the Chinese Olympic diving team. It introduces direct statements from Zhou Jihong and Chen Yiwen, providing insight into their thoughts and confidence ahead of the competition. Zhou’s statement, “said Zhou Jihong”, conveys her optimism and leadership as a former champion and director, expressing belief in the team’s potential. Chen Yiwen’s quote, “said Chen Yiwen”, highlights her personal feelings of calmness and readiness before the event. In both cases, “said” is used to report their views and attitudes, offering readers a first-hand

perspective on the athletes’ mental preparation and expectations. In the August 6th news report, the term “Chinese” is consistently used in the context of national identity, referring to the athletes from China and their achievements at the Olympics. The phrase “Chinese swimmers” highlights the collective identity of the team as they overcome challenges, including unfair anti-doping scrutiny, and achieve a record-breaking performance. The term reinforces national pride, emphasizing the athletes’ perseverance and success in bringing home medals. Additionally, “Chinese quartet” and “Chinese winners” highlight individual contributions within the team while maintaining a focus on their shared identity as representatives of China. The references to “Chinese” in the context of overcoming false accusations further emphasize resilience, portraying the team as not only victorious in the pool but also in defending their integrity on the global stage. In the August 9th news report, the term “games” refers to the Olympic Games, specifically the 2024 Paris Olympics. It is used to contextualize the achievements of Chinese athletes within the broader framework of this prestigious global sporting event. The phrase “at the Games” signifies the Olympic setting, emphasizing the international competition in which these athletes are participating. Additionally, “the past three Games” refers to previous Olympic cycles, highlighting the return of certain events, such as the men’s canoe double 500 meters, to the Olympic program after a hiatus. In this context, “games” conveys the significance of the Olympics as a recurring event that serves as a platform for athletic excellence and historical milestones, while also underscoring the athletes’ performances in Paris.

Keywords

Except for high-frequency words, keywords are also a crucial pointer to examine lexical features. A keyword may be defined as a word that occurs with unusual frequency in a given text ([Scott, 1997](#)). Keywords are the “aboutness” of a text. That is, it clings to the content or the theme of a text. With BNC as the reference corpus, this study retrieves the top 20 keywords from the news reports on the 24th Olympics in China Daily (see [Table 1](#)). As [Pojanapunya and Watson Todd \(2018\)](#) report, log-likelihood highlights more specialized words serving critically-oriented purposes. The rank is ordered according to the log-likelihood.

Table 1

Top 20 Keywords in China Daily's Sports News Compared with BNC

Rank	Keyword	Frequency	Keyness
1	Olympic	689	6825.23
2	Olympics	451	4889.94
3	China	668	4859.88
4	Paris	672	4811.93
5	Chinese	587	4509.33
6	games	497	3318.28
7	athletes	278	2810.86
8	gold	434	2543.28
9	medal	286	2492.81
10	Zheng	158	2064.12
11	team	390	1564.64
12	sports	253	1528.13
13	doping	129	1510.95
14	Zhang	113	1385.23
15	Wang	119	1097.70
16	medals	121	1039.62
17	final	269	966.71
18	tennis	153	897.04
19	world	422	885.72
20	Tokyo	120	857.38

Some keywords have been examined in high-frequency word analysis. These words include “Olympic”, “Paris”, “Chinese”, and “games”. The top keywords can be roughly divided into two parts. One is related to China, and the other connects with athletes. “gold”, “medal”, “doping”, “medals”, “world”, and “Tokyo” are tied up with China’s image. “athletes”, “Zheng”, “team”, “sports”, “Zhang”, “Wang”, “final”, and “tennis” construct athletes’ images. On the reports of August 7th, “gold medals” is frequently mentioned. The phrase “gold medals” in this context appears within a discussion that reflects a shift in China’s social attitude towards sports achievements, moving away from a singular focus on winning gold. This usage constructs an image of China as a more mature and confident nation that values fair competition and athletes’ efforts over mere medal counts. The emphasis on respect for athletes, regardless of their victories, portrays China as a nation with a balanced and progressive view of success, which transcends narrow definitions based on winning gold medals alone. Additionally, the image of Chinese athletes is shaped as hard-working. They deserve respect whether they win or lose, highlighting

the importance of perseverance and dedication. This broader perspective aligns with China’s national priorities, focusing on improving citizens’ livelihoods and promoting health, which in turn reflects China’s confidence and evolving outlook on both sports and national development. “Zheng” is a keyword with the athlete Zheng Haohao being an important person. On the reports of August 8th, the portrayal of Zheng Haohao constructs an image of a highly talented, yet humble and carefree young athlete. Her calm behavior while performing in the high-pressure setting of the Olympics at such a young age, just 11, suggests maturity beyond her years. The narrative emphasizes her relaxed approach to competition, with Zheng viewing the Olympics as an extension of her everyday skateboarding experience rather than a monumental event. This framing underscores her authenticity and passion for sports, not driven by results but by enjoyment. Descriptions of her smooth performance and interactions with fellow skaters contribute to an image of confidence. Additionally, the focus on her being the youngest Olympian from China highlights her surprising talent, while her attitude towards media attention and competition pressure suggests resilience and a focus on personal growth, constructing her as an inspiring figure. Another keyword “doping” is extremely prominent. The news of July 26th mentions this word many times. “anti-doping” is also frequent in this piece of news. The terms “doping” and “anti-doping” in this piece of news construct a positive image of China as a responsible and cooperative global player, actively committed to fairness, integrity, and international solidarity in sports. By highlighting the China Anti-Doping Agency’s (CHINADA) call for global unity in anti-doping efforts, China is portrayed as a nation that values equity and transparency in the global sports arena, aligning itself with international standards and organizations such as the World Anti-Doping Agency (WADA). On the whole, keywords serve as an efficient way to construct China and its athletes’ images.

Collocations

Next, this study seeks to address China and its athletes’ images by elaborating on the relevant collocations. The corpus tool LancsBox 6.0 is used to visualize the collocation figures. The words “Chinese” and “athletes” tightly connect with China’s national image and the athletes’ images. Figure 4 shows the collocation visualization of the word “Chinese”.

The news goes further in developing the image of women athletes by emphasizing both their physical strength and their appearance. By noting their “elegant makeup, intricately styled hair, and artistically painted nails”, the news illustrates that female athletes are not only powerful competitors but also embrace their femininity. This juxtaposition of athletic strength and traditional beauty challenges stereotypes and pinpoints the complexity of modern female athletes, who can do well in sports and show their individuality and style. This finding is different from [Fink \(2015\)](#)’s result, which points out that female athletes received starkly disparate treatment in the 2012 London Olympic Games. Therefore, gender equality has indeed improved in the 2024 Paris Olympic Games. Through these statements, the news constructs an image of female athletes as symbols of perseverance and grace. It highlights their athleticism while celebrating their femininity, bridging a connection between their sporting achievements and the ongoing fight for gender equality.

To make the findings more dependable, this paper seeks to use cognitive psychological theories to elucidate and triangulate China and its athletes’ images which are reached above. The understanding of news texts is a complex process. In ([Wodak, 1987](#))’s integrative comprehension model, cognitive psychology is an important part of her interdisciplinary framework. China Daily’s news about the 2024 Paris Olympics constructs China and Chinese athletes’ images through several key cognitive and psychological mechanisms, shaping how readers perceive China and its athletes. This paper focuses on how cognitive framing, stereotypes, attribution theory, exposure effects, and self-identity contribute to building China’s image and that of its athletes. Cognitive framing plays a crucial role in how China is depicted in these reports. Through a positive framing of China’s achievements, such as technological advancements in sports broadcasting and the success of Chinese athletes, China Daily emphasizes China’s modernization and global influence. The news headlines often highlight China’s advanced cloud computing, artificial intelligence contributions, and athletic victories, framing China as a pioneer not only in sports but also in innovation. This selective framing encourages readers to interpret China’s involvement in the Olympics as a display of national strength and international cooperation, influencing readers’ cognitive schema and fostering positive attitudes toward China. It is also found that the use of stereotypes influences how Chinese athletes are perceived by readers. Common stereotypes of Chinese athletes emphasize discipline, perseverance, and a strong work ethic. China Daily reinforces these stereotypes by highlighting the hard

work, rigorous training, and mental state of Chinese Olympians. For instance, stories of athletes overcoming challenges or setting new records not only fit into global stereotypes of Asian athletes as highly disciplined but also reinforce a sense of collective pride and unity among Chinese readers. These cognitive shortcuts simplify the complex identities of athletes, making it easier for readers to process information and assign positive attributes to them. Attribution theory further explains how readers ascribe success or failure in sports to certain causes. In China Daily’s news reports, the success of Chinese athletes is frequently attributed to national systems, advanced training methodologies, and government support, rather than solely to individual effort. For example, when Chinese swimmers win medals, their victories are often framed as the result of a well-organized sports infrastructure and China’s commitment to excellence in global competition. This external attribution places the athletes’ success within the broader context of China’s institutional and technological advancements, enhancing the perception of China as a powerful and well-coordinated country. The mere exposure effect also plays a role in shaping both domestic and international perceptions of China. Continuous media reports of Chinese athletes and their achievements increase familiarity and positive effect on them. The frequent mentions of Chinese athletes in various events, from swimming to weightlifting, along with images of their performances and victories, enhance the readers’ emotional attachment and admiration for these figures. The sheer repetition of positive imagery and stories about Chinese athletes leads to an increased likability of both the athletes and, by extension, China they represent. By and large, from a cognitive psychology perspective, China Daily’s news reports of the 2024 Paris Olympics construct a robust and multi-dimensional image of China and its athletes through a series of cognitive and psychological mechanisms. These cognitive processes work together to shape perceptions of China as a rising global power and its athletes as embodiments of national pride, discipline, and perseverance.

Conclusion

This paper combines the text mining method with cognitive psychological analysis to examine China Daily’s news reports on the 24th Olympics. The data range from July 26th, 2024 to August 11th. After data cleaning and checking, the corpus is built. This paper uses the corpus software Voyant tools to deal with high-frequency words and their diachronic distribution. The word cloud indicates that Olympic (689), Paris (672), said (618),

Chinese (587), games (497), Olympics (451), gold (434), China (431), world (388), team (366), 2024 (302), medal (281), athletes (278), final (268), time (253), sports (253), year (250), China's (237), women's (209), and event (187) are the top 20 frequency content words. These words touch upon the main meanings of the news in the corpus. China and its athletes are the main focus. This study also examines the diachronic distribution of the top 5 words. These words are extremely high-frequency on some dates. To explore the theme of these news reports, this study also retrieves the top 20 keywords and analyzes the theme based on the language context. These keywords are, to some extent, similar to the high-frequency words, but there are still some differences. For example, some athletes' names and "doping" are included. These keywords yield benefits to the construction of China and its athletes' images. This study further addresses the collocation visualization of "Chinese" and "athletes", which are both highly frequent and beneficial to the study of China and its athletes' images. Based on the investigation of the collocation words "tech" and "swimmers" of "Chinese", it finds out that the news constructs a forward-thinking image of China and the victorious, gracious, and sportsmanlike athletes' images. The analysis of the collocation words "male" and "female" of "athletes" offers insight into China being a country of gender

equality. These findings are further explained in terms of cognitive and psychological theories, which can do good for the development of sports psychology.

These findings might be used by several groups, such as news media, readers, and scholars to guide their work, life, and research. News media, especially China Daily, can examine whether the national and athletes' images are consistent with their intentions. Then lexical choices and discourse strategies can be adjusted to their aims. Additionally, readers can make clear which words and collocations contribute to their impressions of Chinese and athletes' images. For scholars, this paper shows them the importance of data mining, especially in the era of big data. However, there are still several limitations in this study. For one thing, the images and videos in the news are left unexplored. For another thing, more research methods such as interviews and questionnaires can be used in cognitive psychological analysis. For future studies, the multimodal discourse analysis on sports news is a potential area. In addition, it is expected to conduct some psychological experiments on real readers of sports news. Future research should also bridge the gap between cognitive psychology and media studies by emphasizing the interaction between text features and reader cognition, particularly in the context of news media.

References

- Baker, P. (2012). Acceptable bias? Using corpus linguistics methods with critical discourse analysis. *Critical Discourse Studies*, 9(3), 247-256. <https://doi.org/10.1080/17405904.2012.688297>
- Barsalou, L. W. (2014). *Cognitive Psychology: An Overview for Cognitive Scientists*. Psychology Press. <https://doi.org/10.4324/9781315807485>
- Bednarek, M., & Caple, H. (2014). Why do news values matter? Towards a new methodological framework for analysing news discourse in critical discourse analysis and beyond. *Discourse & Society*, 25(2), 135-158. <https://doi.org/10.1177/0957926513516041>
- Chambers, R., Gabbett, T. J., Cole, M. H., & Beard, A. (2015). The use of wearable microensors to quantify sport-specific movements. *Sports medicine*, 45, 1065-1081. <https://doi.org/10.1007/s40279-015-0332-9>
- China Daily. (2024). *China Daily*. <http://www.chinadaily.com.cn/>
- Collins, L., & Nerlich, B. (2015). Examining user comments for deliberative democracy: A corpus-driven analysis of the climate change debate online. *Environmental communication*, 9(2), 189-207. <https://doi.org/10.1080/17524032.2014.981560>
- Csomay, E., & Crawford, W. J. (2024). *Doing Corpus Linguistics*. Routledge. <https://doi.org/10.4324/9781003363309>
- Ellis, N. C., Simpson - Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL quarterly*, 42(3), 375-396. <https://doi.org/10.1002/j.1545-7249.2008.tb00137.x>
- Fernandez - Duque, D., & Johnson, M. L. (1999). Attention metaphors: How metaphors guide the cognitive psychology of attention. *Cognitive science*, 23(1), 83-116. https://doi.org/10.1207/s15516709cog2301_4
- Ferrer-Conill, R. (2017). Quantifying journalism? A study on the use of data and gamification to motivate journalists. *Television & New Media*, 18(8), 706-720. <https://doi.org/10.1177/1527476417697271>
- Fink, J. S. (2015). Female athletes, women's sport, and the sport media commercial complex: Have we really "come a long way, baby"? *Sport management review*, 18(3), 331-342. <https://doi.org/10.1016/j.smr.2014.05.001>

- Goldstein, E. B. (2015). *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Cengage Learning Stamford, CT. <https://www.cengage.com/c/cognitive-psychology-connecting-mind-research-and-everyday-experience-5e-goldstein/9781337408271>
- Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, 19(5), 595-610. <https://doi.org/10.1177/1464884916641269>
- Han, J., Pei, J., & Tong, H. (2022). *Data Mining: Concepts and Techniques*. Morgan kaufmann.
- Harrison, G. (2018). *On the Sidelines: Postfeminism, Neoliberalism, and the American Female Sportscaster* [Doctoral dissertation, Arizona State University]. <https://hdl.handle.net/2286/R.I.51590>
- Harutyunyan, H., Khachatryan, H., Kale, D. C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. *Scientific data*, 6(1), 96. <https://doi.org/10.1038/s41597-019-0103-9>
- Iosifidis, P., & Nicoli, N. (2020). The battle to end fake news: A qualitative content analysis of Facebook announcements on how it combats disinformation. *International Communication Gazette*, 82(1), 60-81. <https://doi.org/10.1177/1748048519880729>
- Jo, T. (2019). *Text Mining: Concepts, Implementation, and Big Data Challenge*. Springer Cham. <https://doi.org/10.1007/978-3-319-91815-0>
- Lavandera, R. L., Fernández-Gavira, J., Pérez-Villalba, M., & García-Fernández, J. (2017). The integration of research and practice in sport psychology. *Revista de psicología del deporte*, 26(4), 1-4. <https://www.redalyc.org/pdf/2351/235152047001.pdf>
- Liu, Y., Zhang, L. J., & May, S. (2022). Dominance of Anglo-American cultural representations in university English textbooks in China: A corpus linguistics analysis. *Language, Culture and Curriculum*, 35(1), 83-101. <https://doi.org/10.1080/07908318.2021.1941077>
- Machin, D., & Mayr, A. (2023). *How to Do Critical Discourse Analysis : A Multimodal Introduction*. Sage Publications. <http://digital.casalini.it/9781529783759>
- Mariani, M. M., Perez - Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing*, 39(4), 755-776. <https://doi.org/10.1002/mar.21619>
- McEnery, T., & Brookes, G. (2024). Corpus linguistics and the social sciences. *Corpus Linguistics and Linguistic Theory*. <https://doi.org/10.1515/cllt-2024-0036>
- McMaster, K. L., Espin, C. A., & Van Den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research & Practice*, 29(1), 17-24. <https://doi.org/10.1111/ldr.p.12026>
- O'Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics*. Routledge London. <https://doi.org/10.4324/9780367076399>
- Ocuppaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487-501. <https://doi.org/10.1111/bjet.12156>
- Pan, Z. (2023). Incorporating Network Communication and Sports Psychology: An In-depth Quantitative Analysis of Interdisciplinary Knowledge Graphs and Association Patterns in Chinese Sports News Communication. *Revista de Psicología del Deporte (Journal of Sport Psychology)*, 32(4), 234-244. <http://www.rpd-online.com/index.php/rpd/article/view/836>
- Pino-Ortega, J., Rojas-Valverde, D., Gómez-Carmona, C. D., & Rico-González, M. (2021). Training design, performance analysis, and talent identification—A systematic review about the most relevant variables through the principal component analysis in Soccer, Basketball, and Rugby. *International Journal of Environmental Research and Public Health*, 18(5), 2642. <https://doi.org/10.3390/ijerph18052642>
- Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133-167. <https://doi.org/10.1515/cllt-2015-0030>
- Re Lee, K., Hyun Kim, J., Jang, J., Yoon, J., Nan, D., Kim, Y., et al. (2023). News big data analysis of international start-up innovation discourses through topic modelling and network analysis: comparing East Asia and North America. *Asian Journal of Technology Innovation*, 31(3), 581-603. <https://doi.org/10.1080/19761597.2022.2134154>
- Scott, M. (1997). PC analysis of key words—and key key words. *System*, 25(2), 233-245. [https://doi.org/10.1016/S0346-251X\(97\)00011-0](https://doi.org/10.1016/S0346-251X(97)00011-0)
- Wang, G., Liu, Y., & Tu, S. (2023). Discursive use of stability in New York Times' coverage of China: a sentiment analysis approach. *Humanities and Social Sciences Communications*, 10(1), 1-10. <https://doi.org/10.1057/s41599-023-02165-0>

- Wang, X., Kaur, S., Mat Isa, N. A. N., & Kaur, S. (2024). Image restoration strategies in pandemic crisis communication: a comparative analysis of Chinese and American COVID-19 political speeches. *Humanities and Social Sciences Communications*, 11(1), 1-13. <https://doi.org/10.1057/s41599-024-03825-5>
- West, J., & Bhattacharya, M. (2016). Intelligent Financial Fraud Detection: A Comprehensive Review. *Computers & Security*, 57, 47-66. <https://doi.org/10.1016/j.cose.2015.09.005>
- Wodak, R. (1987). 'And where is the Lebanon?' A socio-psycholinguistic investigation of comprehension and intelligibility of news. *Text-Interdisciplinary Journal for the Study of Discourse*, 7(4), 377-410. <https://doi.org/10.1515/text.1.1987.7.4.377>
- Xu, Q. (2024). Identifying disputants' attitudinal variations in family mediations: A data mining approach. *Discourse Studies*, 14614456241237378. <https://doi.org/10.1177/14614456241237378>
- Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335-4385. <https://doi.org/10.1007/s10462-019-09794-5>
- Zhang, K., Chen, Y., & Li, C. (2019). Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management*, 75, 595-608. <https://doi.org/10.1016/j.tourman.2019.07.002>
- Zhang, L., & Caple, H. (2021). The newsworthiness of Li Na—A critical comparative analysis of Chinese and international news media. *Language & Communication*, 77, 70-80. <https://doi.org/10.1016/j.langcom.2021.01.002>