

Text Mining Analysis of Malaysia's National Image in *China Daily*

ZHANG JIE *

*Faculty of Modern Languages and Communication
Universiti Putra Malaysia, Malaysia
gs66398@student.upm.edu.my*

MOHD AZIDAN ABDUL JABAR

*Faculty of Modern Languages and Communication
Universiti Putra Malaysia, Malaysia*

FARHANA MUSLIM MOHD JALIS

*Faculty of Modern Languages and Communication
Universiti Putra Malaysia, Malaysia*

ABSTRACT

This study takes October. 4. 2013, as a temporal dividing point, conducting a comparative analysis of news closely related to Malaysia in China Daily, over the 10 years before and after this date. The qualitative research software Leximancer 5.0 is employed for text mining analysis. The findings reveal that the themes in the news transitioned from relatively dispersed, generalised coverage to a more focused, in-depth exploration of major events and China-Malaysia cooperation. The primary social actors demonstrated minimal change over the 20 years but became increasingly focused. Sentiment analysis shows that reporting on Malaysia has predominantly been positive, with the proportion of positive sentiment showing an upward trend. However, some negative factors were also observed, such as a shift in the portrayal of the 6th Prime Minister of Malaysia, from positive to negative. Notably, although the MH370 incident was reported with a predominantly negative perspective, it did not persist for long. Positive sentiments towards all associated social actors, except Indonesia, have continued to grow, including China, Singapore, Australia, and ASEAN. This study portrays Malaysia as a promising partner in economic growth and development, an important player in regional and global relations, and a stable and enduring strategic ally. By employing an innovative combination of text mining techniques and discourse analysis, the research not only provides empirical evidence of the evolving discourse surrounding Malaysia but also offers methodological insights for applying computational tools in media studies.

Keywords: national image construction; text mining analysis; qualitative research software; media discourse analysis; diachronic study

INTRODUCTION

As one of the core concepts in this study, national image is not a new topic. It has evolved alongside international interactions since the establishment of nation-states and emerged as an academic subject in the 1990s (Papadopoulos & Heslop, 2014). Over the past three decades, research on national image has continued to deepen and expand, becoming both a significant practical issue and a focal point in academic discourse (Chavez, 2023). As early as the mid-twentieth century, Boulding (1956) defined national image as a combination of a country's self-perception and others' perceptions of it in the international system. Thus, Rusciano (1997) differentiated between self-perception and others-perception. National image is defined as the perception held by foreign audiences, while national identity refers to the self-perception of a nation's citizens, representing a form of collective identity grounded in individual-level perceptions (Rusciano, 2003). The

externally constructed national image serves as a mirror, reflecting both the international community's evaluations of a nation and the effectiveness of its information dissemination in external communication. If a country focuses solely on what it intends to convey (self-construction) while neglecting how its message is understood by others (other-construction), communication failures are likely to occur. Therefore, studying externally constructed national image provides a more direct insight into the effectiveness of communication efforts (Chang, 2017).

Fowler (1991) challenges readers, media students, and journalists involved in news production to recognise how language constructs rather than merely reflects the world. Studying the representation of news discourse could help researchers understand how news discourse represents and constructs reality (Van Dijk, 2013). Newspaper language reflects, and in return constructs, the prevailing values of a society in a particular historical moment (Fowler, 2013). Therefore, the current study, focusing on the construction of Malaysia's national image in news reports from the Chinese mainstream media *China Daily*^①, defines national image as a comprehensive perception and evaluation formed by foreign audiences through information intermediaries, particularly mass media, and adopts this concept throughout the analysis. A key reason for selecting *China Daily* as the data source is its status as China's leading English-language newspaper (G. Wang, 2018), serving as a primary channel for international audiences to understand China's perspectives on global affairs. As a state-affiliated media outlet, its coverage reflects China's official stance on diplomatic relations (Xu & Wang, 2022).

Changes in both domestic and international environments, including political system reforms, economic transformations, cultural dissemination, and adjustments in international relations, have profound impacts on national image. In this sense, national image is inherently dynamic. The advancement of media and the globalisation of information exchange have further accelerated and complicated the construction and dissemination of national image. The stability of a national image is relative, while its dynamism is absolute (Wu & Zhang, 2016, p.61). Given this, studying national image requires incorporating the historical context and social environment of different periods, adopting a dynamic and diachronic perspective to uncover its developmental trajectory and underlying logic.

Despite increasing scholarly attention to national image, research on Malaysia's media representation remains scarce. Existing studies have primarily focused on developed countries or major global powers, leaving Southeast Asian developing countries underexplored. Moreover, most national image studies adopt a synchronic perspective, making it difficult to capture long-term discursive shifts. While text mining has been applied in discourse studies, its potential for large-scale, diachronic analysis of national image remains underutilised. Given China's growing influence and the role of *China Daily* as a major English-language media, it is crucial to examine how Malaysia's national image has evolved in this context.

Consequently, the research objectives of this study are to examine the thematic evolution, representation of social actors, and sentiment trends in *China Daily*'s coverage of Malaysia. Accordingly, the following research questions will be addressed:

- RO1. How have the themes in *China Daily*'s coverage of Malaysia evolved over two decades?
- RQ2. What social actors were present in *China Daily*'s reporting on Malaysia in each of the two time periods?
- RQ3. How has *China Daily*'s sentiment towards Malaysia and its key social actors changed over the past 20 years?

LITERATURE REVIEW

In recent years, research on national image has shown a discursive shift (Chen, 2018). Linguists have explored the construction of national image through various social media channels. National image is often disseminated through influential individuals, who may be either political figures or ordinary citizens. Popova *et al.* (2023) analysed the image of the world-renowned political figure, Russian President Vladimir Putin, in constructing Russia's national image. The result coincided with those of Lyudmila *et al.* (2021). Restad (2020) evaluated American national image construction through their studies on U.S. President Donald John Trump. Besides influential political figures, ordinary individuals also contribute to constructing the national image through their works. Z. T. Chen and Whyke (2022) and Whyke *et al.* (2022) examined Li Ziqi, a globally recognised mega influencer, and her YouTube vlogs, which narrated tales of rural Chinese life, crafting an appealing image of China. In addition, the influence of mega-events, especially on the construction of the national image, should not be underestimated. Siyu and Feng (2023) conducted studies on the 2022 Beijing Winter Olympics, while Lemus-Delgado (2023) examined the 2021 Tokyo Olympics, using sports image as a reflection of its national image.

The above studies are all synchronic studies targeting specific populations and events. However, synchronic studies only collect data at a single point in time, making it difficult to observe and analyse changes or developments over time or to determine whether certain changes are temporary or long-term (Karasti *et al.*, 2010). Moreover, such studies cannot provide the historical context of the data (Tai, 2023). Diachronic studies help reveal long-term trends in media narratives and public perception, offering deeper and more comprehensive insights.

Discourse analysts often face challenges in processing large, diverse news datasets. Text mining, with its capacity to handle unstructured data efficiently, supports such research by uncovering hidden patterns and relationships. It automates the extraction of novel insights from massive text corpora. Applied to news archives spanning decades, text mining reveals thematic developments and discourse shifts (Khan *et al.*, 2023). Consequently, it is increasingly used to refine and objectify national image analysis. Tjellander (2022) applied a mixed-method approach to examine Sweden's media image in *The Guardian* (2010--2020) through text mining. The study identified immigration, the Nobel Prize, climate, and foreign policy as key concerns. Notably, it demonstrates how diachronic analysis of large corpora can advance public diplomacy research by linking digital humanities with international communication. Prieto (2021) examined national stereotypes of Spain from the perspective of others by analysing Facebook posts containing the words "Spain" and "Spanish" from June to December 2016. The corpus consisted of online discussions in Austria, Belgium, Italy, France, Bulgaria, and Poland. The research findings reveal that, despite being in the same time period and focusing on the same subject, different sources or actors with varying backgrounds exhibit distinctly divergent characteristics in their external construction of national image. Lin and Miao (2023) conducted a study on how six mainstream U.S. media constructed China's national image through the concepts and themes, as well as the representation of China as a social actor in the coverage. Leximancer was used to analyse four decades of news data, focusing on themes and the portrayal of China as a social actor. A key contribution of the study was the development of a text mining-based social actor framework, highlighting the integration of computational tools into discourse analysis.

To reduce subjectivity in qualitative research, often shaped by the "researcher-as-instrument" issue (Wa-Mbaleka, 2020), qualitative software has become essential. It supports

systematic data handling, standardised coding and visualisation, improving consistency and rigour. Such tools also enhance transparency and reproducibility (O’Kane *et al.*, 2021), making analysis more traceable. Additionally, software streamlines data workflows, enabling researchers to concentrate on theory-building and interpretation, thereby broadening methodological scope. With these tools, discourse analysis can be conducted with greater depth and reliability.

In conclusion, while national image research has grown, existing studies remain limited. Much of the existing work is synchronically focused, lacking the ability to trace discursive shifts over time. Although some diachronic efforts exist, they often span short periods or rely on qualitative methods without strong empirical grounding. Despite increased use of text mining, its potential in discourse studies remains underexplored.

To address these research gaps, this study not only adopts a diachronic text mining approach but also incorporates a multi-dimensional theoretical framework to guide the analysis. First, the study is grounded in the discourse construction theory advanced by Fowler (1991) and Van Dijk (2013), which posits that news reporting is not a passive reflection of reality but an active linguistic practice that constructs national image. Through selective reporting, media outlets play a pivotal role in constructing a country’s perceived status and significance in the international arena (Van Dijk, 2013). Second, this study draws on the conceptual distinction between “national image” and “national identity” proposed by Rusciano (1997), with a particular focus on other-constructed national image. Furthermore, the study applies Van Leeuwen’s (2008) theory of the representation of social actors. Combined with the “Name-like Concepts” function in Leximancer software, this framework enables a dual-layered analysis from both discursive and computational perspectives.

RESEARCH METHODOLOGY

DATA SELECTION AND COLLECTION

This study adopts a diachronic perspective to examine the changes in *China Daily*’s news coverage of Malaysia over two decades, divided into two periods: October 4, 2003, to October 4, 2013, and October 5, 2013, to October 4, 2023. The date of October 4, 2013, is chosen as the key turning point for two main reasons: first, Chinese President Xi Jinping engaged in discussions with Prime Minister Najib at the Malaysian Prime Minister’s Office, where the two leaders reached a significant consensus and agreed to upgrade their bilateral relations to a comprehensive strategic partnership^②, furthermore, following Xi Jinping’s proposal of the “21st Century Maritime Silk Road” on October 3, 2013^③. Malaysia has actively embraced and participated in the initiative, fostering a strong and cooperative relationship with China across diplomatic, economic, and other domains (Liu & Lim, 2020).

The data collection process involves extracting relevant news articles from the *China Daily* website using a Python script. Before scraping the data, the Advanced Search function was configured to include only news where both the title and full text contained the terms “Malaysia” or “Malaysian”. Blogs and photo-related news were excluded, and the sources selected for inclusion were *China Daily*, *Web News*, *Business Weekly*, *HK Edition*, *Global Edition*, and *BBS*. This process initially yielded 1,340 articles. After manual verification, duplicates, invalid web pages, and video content were removed. Ultimately, a total of 1,286 news articles closely related to Malaysia were retained for analysis, as shown in Figure 1.

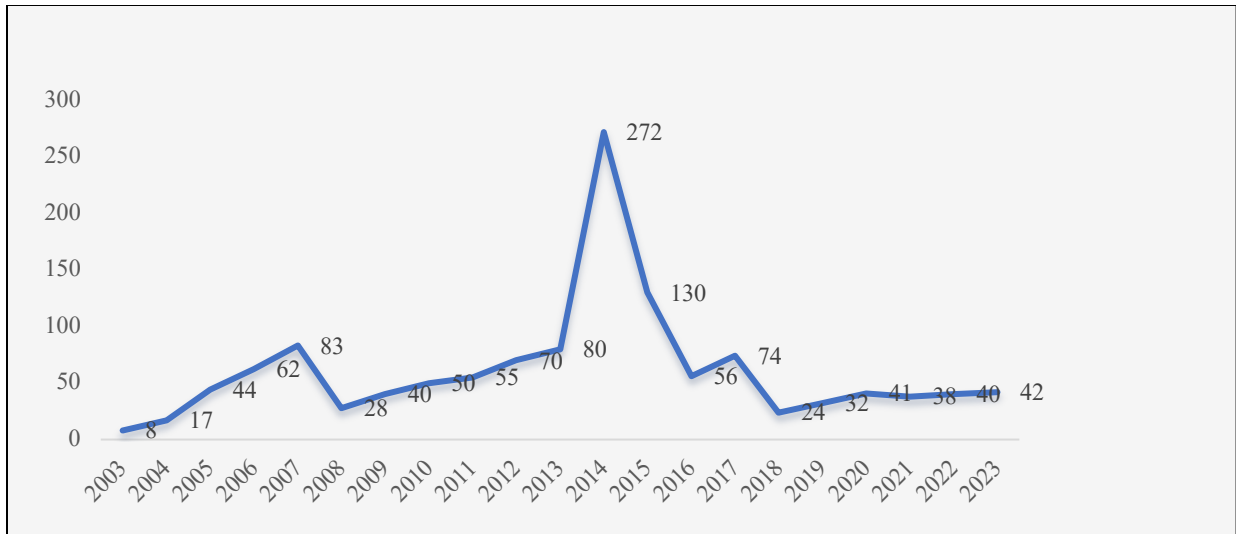


FIGURE 1. The coverage of *China Daily* on Malaysia from 4.10.2013 to 4.10. 2023

DATA ANALYSIS PROCEDURE

The study utilises Leximancer 5.0 to perform a thematic and sentiment analysis of *China Daily*'s coverage related to Malaysia. The qualitative research tool is extensively applied in disciplines such as social sciences, market research, or media studies for analysing textual data and generating concept maps (Goh & Wilk, 2022). It enables researchers to grasp the thematic organisation of large-scale text corpora. As an unsupervised text analysis software, it eliminates the need for predefined themes or categories.

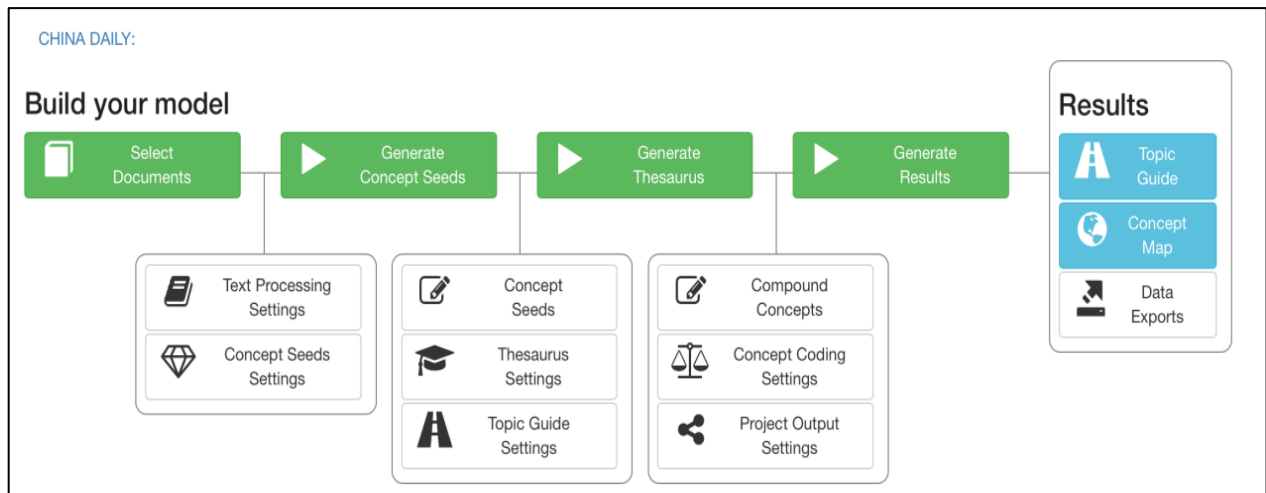


FIGURE 2. The Operation Process of Leximancer 5.0

Two models were built for the analysis. The first model, named *CD1*, includes 512 news articles published between October. 4. 2003 and October. 4. 2013. The second model, named *CD2*, comprises 774 news articles published between October 5, 2014, and October 4, 2023. As a specific operational process illustrated in Figure 2, the initial step involves uploading the cleaned

corpus into the software to construct two research modules. During this phase, text processing configurations are adjusted, including pre-processing tasks such as stop word removal and stemming. Utilising the text data, the system identifies preliminary concept seeds and develops a thesaurus for advanced analysis. Once these processes are completed, the software delivers the analysis output, which includes a Topic Guide and a Concept Map for further exploration.

RESULTS AND DISCUSSION

THEMATIC ANALYSIS OF MALAYSIA'S NATIONAL IMAGE

By employing Leximancer 5.0, configuring the concept threshold at 50% along with a theme size of 35%, the analysis revealed 12 primary conceptual themes in *CD1* and 11 in *CD2*, as illustrated in Figure 3. Each coloured circle signifies a thematic domain, with distinct colours representing unique core themes. Larger and darker circles indicate themes of greater relevance. The words within each thematic area signify key concepts linked to that theme, and their size denotes their relative significance or frequency of appearance. The lines connecting concepts illustrate co-occurrence relationships in the text, with denser connections indicating stronger associations between the concepts.

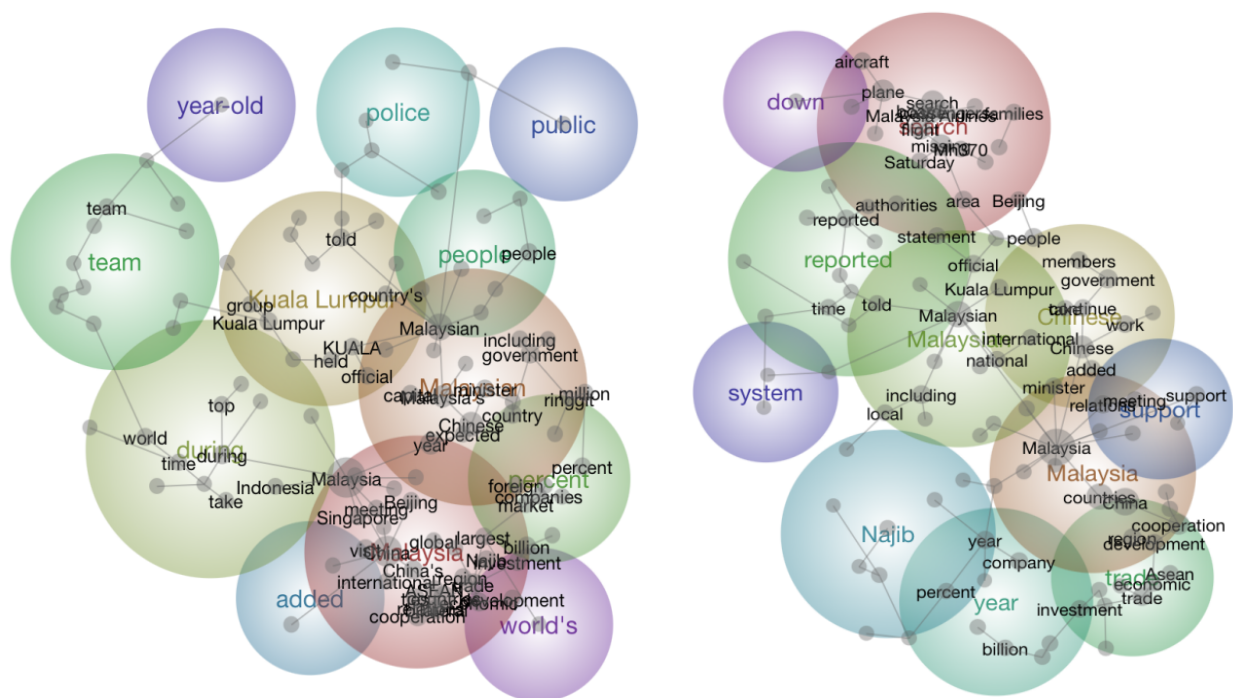


FIGURE 3. The concept map of text mining on Malaysia-related reports in *China Daily*
(Left: Period from October 4, 2003, to October 4, 2013; Right: Period from October 5, 2014, to October 4, 2023)

In *CD1*, despite consisting of only 512 news articles, significantly fewer than the 774 articles in *CD2*, the thematic keywords encompass a wide range of reporting content, reflecting the overall inclusiveness and diversity of the news coverage. The reporting leans more toward regular and comprehensive news coverage, with fewer specialised or targeted topics. The core

thematic keywords related to nation and image include “Malaysia”, “Malaysian” and “Kuala Lumpur”. The keyword “Malaysia” frequently appears in contexts related to diplomacy, economic cooperation, social development, and culture. “Malaysian” is typically used to describe the people of Malaysia, such as in “Malaysian government”, “Malaysian people”, or their associated attributes, emphasising the importance of Malaysian identity and its people in the narrative. “Kuala Lumpur”, as the capital of Malaysia, is often used to symbolise the nation’s political, economic, cultural, and educational centre. The thematic keywords related to society and the public include “People” and “Public”, reflecting a focus on issues concerning social perspective and ordinary citizens. The keywords underscore the attention to public interest and societal impact. “Percent” typically appears in reports involving statistics, such as investment or market share, reflecting the data-driven nature of the news. “Team” is often linked to international sports, collaboration, or rescue efforts, highlighting collective action. “Police” denotes reports on crime, law, and security, underscoring its role in maintaining order. “Added” introduces indirect quotes or supplementary details, improving clarity and continuity. “World’s” conveys a global context, reinforcing Malaysia’s relevance in international affairs.

Compared to the looser thematic links in *CD1*, the clusters in *CD2* are more concentrated, with stronger and even overlapping connections. Besides “Malaysia” and “Malaysian”, the appearance of “Chinese” reflects closer Malaysia-China ties. The shift suggests increased focus on bilateral interaction and cooperation, reinforcing Malaysia’s engagement with China globally. Furthermore, the news highlights a substantial number of themes related to economic cooperation. Keywords like “trade”, “investment”, and “cooperation” link Malaysia and China, demonstrating the prominent economic ties. The blue circle labelled “support” cluster overlaps both countries, signalling substantive progress in diplomacy and economic collaboration. Major events take centre stage in *CD2*. The red cluster, marked by terms “search”, “plane”, and “MH370” shows the crash’s central role in media coverage. Former Prime Minister “Najib” also appeared frequently, with reports focusing on his political and diplomatic engagements, indicating sustained media interest in Malaysia’s leadership. Overall, *CD2* themes are more integrated, highlighting increased attention to Malaysia-China relations, key events, and political figures. It reveals changing media priorities, with a greater focus on bilateral ties, global affairs, and leadership.

An analysis of coverage on Malaysia across *CD1* and *CD2* reveals a shift from broad, general reporting to a stronger focus on major events and bilateral cooperation. In *CD1*, topics were diverse, covering diplomacy, economy, social development, and culture, reflecting a wide-ranging news agenda. Over time, *CD2* shows more concentrated and interconnected themes, with Malaysia-China economic ties and the emergency incident. At the same time, attention to political figures and Malaysia’s overall development was maintained.

It was identified that “Malaysia” and “Malaysian” are core themes consistently shared across both phases. According to the automatic topic concept extraction function in Leximancer 5.0, a comparative analysis was conducted on the concepts associated with “Malaysia” and “Malaysian”, as shown in Table 1.

TABLE 1. Themes and Concepts in CD1 and CD2

CD1		CD2	
Theme	Concepts	Theme	Concepts
Malaysia	Malaysia, China, countries, trade, cooperation, bilateral, ties, relations, visit, China's, economic, ASEAN, billion, investment, Najib, development, largest, region, Beijing, meeting, Singapore, global, international	Malaysia	Malaysia, China, countries, cooperation, relations
Malaysian	Malaysian, Chinese, Malaysia's, year, government, country, minister, million, capital, expected.	Malaysian	Malaysian, including official, Kuala Lumpur, national, local

“Malaysia” and “Malaysian” are central themes shared across both times. It indicates that the selected corpus of this study is fully aligned with the research objective. For “Malaysia”, in the CD1 phase, it was associated with 23 sub-concepts, covering a wide range of topics such as nation, region, diplomacy, economy, cooperation, and development. However, in the CD2 phase, despite an increase in the number of reports, the sub-concepts narrowed significantly to just five, primarily focusing on Malaysia-China relations and cooperation. It indicates a transition from thematic diversity to a more concentrated focus, emphasising Malaysia’s prominent role as a national entity in bilateral cooperation.

Regarding “Malaysian”, in the CD1 phase, its associated keywords included “government”, “Chinese”, “country”, “million”, and “capital” primarily describing Malaysian citizens, government, and national attributes. While some references were made to Malaysia-China relations, the overall focus remained limited, with a tendency to emphasise Malaysian identity as a marker of national and governmental participation in international affairs. In the CD2 phase, the keywords shifted to “official”, “Kuala Lumpur”, “national”, and “local” with the coverage focusing more on governmental roles and specific local affairs. The shift reflects a transition from broad descriptions of identity and attributes to a more specific focus on government officials, local affairs, and major events, indicating increased media attention on the representative roles of Malaysian citizens.

Overall, *China Daily*’s coverage of Malaysia demonstrates a clear transition from general news reporting to a more targeted focus on major events and bilateral cooperation. The evolution reflects not only a growing recognition of the importance of China-Malaysia relations but also a deeper engagement with Malaysia’s role in regional and global affairs. From the perspective of national image construction, it indicates a functional reconfiguration of Malaysia’s image by the media, emphasising its roles as a regional partner, an economic collaborator, and a cooperative actor in addressing global challenges. By highlighting economic cooperation, shared crises, and diplomatic interactions, the Chinese media constructs Malaysia as a stable, trustworthy, and collaborative nation.

SOCIAL ACTORS ANALYSIS OF MALAYSIA’S NATIONAL IMAGE

Name-like Concept function refers to concepts appearing in the text in the form of names, typically including country names, personal names, organisational names, or place names^④. The concepts are used to identify specific social actors, entities, or geographic regions. In news reporting, they are often employed to construct a narrative focus or convey certain identity attributes.

As shown in Table 2, a total of 16 Name-like Concepts were extracted from each phase. Among them, “Malaysia”, “China”, “Malaysian”, “Chinese”, “Kuala Lumpur”, “Najib”, and “Beijing” are highlighted in bold, indicating their significant positions in both *CD1* and *CD2*, with consistent rankings and relevance scores. The consistency demonstrates *China Daily*’s stability and structure of reporting on Malaysia across both phases. By maintaining these recurring themes, the newspaper not only establishes a consistent narrative framework but also provides readers with a coherent perspective on Malaysia-China relations and key political figures.

TABLE 2. Name-like Concept in *CD1* and *CD2*

Name-like Concept	<i>CD1</i>		Name-like Concept	<i>CD2</i>	
	Count	Relevance		Count	Relevance
Malaysia	1094	100	Malaysia	1776	100
China	591	54	China	974	55
Malaysian	544	50	Malaysian	894	50
Chinese	421	38	Chinese	677	38
Malaysia’s	319	29	Kuala Lumpur	281	16
Kuala Lumpur	200	18	Malaysia Airlines	257	14
China’s	146	13	Saturday	207	12
ASEAN	135	12	ASEAN	204	11
Najib	130	12	Najib	202	11
Beijing	97	9	Beijing	178	10
Indonesia	94	9	Mh370	166	9
Singapore	91	8	March	165	9
Japan	71	6	MH	148	8
Malaysians	65	6	Australia	143	8
Saturday	61	6	Indonesia	112	6
Australia	57	5	Singapore	107	6

In *CD1*, the roles of secondary countries were notably prominent. For instance, the mentions of “Indonesia”, “Singapore”, and “Japan” indicate a focus on regional relationships. The focus on regional countries reflects Malaysia’s role as a key Southeast Asian actor in regional cooperation and multilateral diplomacy. Notably, Japan appears only in *CD1*, indicating broader coverage of East Asian multilateralism during that phase. References to China in *CD1*, including “China”, “Chinese”, “Beijing”, and “China’s”, further signal the early significance of Malaysia-China ties. Although the relations were visible in *CD1*, the coverage remained broad and general, offering limited depth and emphasising overall trends rather than detailed analysis of specific events or areas.

In *CD2*, “Malaysia” appeared 1,776 times with a relevance of 100%, while “China” appeared 974 times with a relevance of 55%, nearly doubling their frequencies compared to *CD1*. The sharp increase suggests that Malaysia-China relations became the dominant narrative. Mentions of other countries declined, while event-driven and thematic reporting rose notably. About a quarter of the mentions related to the aviation disaster, including keywords like “Malaysia Airlines”, “MH370”, “March”, and “Saturday”. The prominence of this topic highlights its significance in *CD2*, reflecting the media’s strong focus on international incidents and its deep engagement with globally important issues. The shift shows that, on the one hand, *CD2*’s coverage became more focused on high-impact international topics. On the other hand, the fact that 154 Chinese passengers were aboard the flight (Fong & Ponnann, 2019) drew significant attention from China, elevating the incident to a matter of national concern. This dual focus underscores the media’s role in amplifying the humanitarian and diplomatic dimensions of such a major crisis (Robinson, 2000).

From the perspective of social actors, “Malaysia” and “China” consistently emerged as the core concepts across both periods. Their persistent prominence, along with the significantly increased frequency and connectivity observed in *CD2*, underscores the centrality of the “China-Malaysia community” narrative in reporting. The continuity and intensification suggest a deliberate media strategy to highlight the depth and stability of bilateral relations, thereby portraying Malaysia as a key strategic partner in China’s peripheral diplomacy. Meanwhile, the frequent appearance of “Najib” reflects the media’s attention to individual leadership representation as part of the broader national image, revealing a dynamic interaction between individual and national image construction. Overall, the fact that nearly half of the Name-like Concepts retained their rankings across both phases demonstrates a high degree of thematic and narrative continuity. More importantly, it reflects a long-term discursive logic centred on stability, cooperation, and mutual trust, through which Malaysia’s national image is consistently reinforced in Chinese media discourse.

SENTIMENT ANALYSIS OF MALAYSIA’S NATIONAL IMAGE

In this section, the Sentiment Analysis feature in Leximancer 5.0 will be employed to analyse the sentiment orientation. Through analysis, a deeper understanding can be gained of *China Daily*’s attitude toward different Name-like Concepts in its coverage of Malaysia. This approach not only supplements the qualitative interpretation of the reporting themes but also provides data-driven insights to uncover the implicit biases behind the media narrative. Through selection, seven terms closely related to Malaysia are identified in both *CD1* and *CD2*, as shown in Table 3. In both phases, positive sentiment (Favour) predominates, indicating that *China Daily*’s coverage of Malaysia-related concepts is largely characterised by a positive portrayal. Negative sentiment (Unfavour) accounts for a relatively smaller proportion and is highlighted in bold in the Table 3.

TABLE 3. Sentiment on the Name-like Concept of Malaysia in *China Daily*

	<i>CD1</i>			<i>CD2</i>			
	Sentiment	Co-Count	Likelihood	Sentiment	Co-Count	Likelihood	
Malaysia	Favour	162	36%	Malaysia	Favour	253	49%
	Unfavour	56	29%		Unfavour	99	31%
Malaysian	Favour	61	14%	Malaysian	Favour	100	19%
	Unfavour	34	18%		Unfavour	60	19%
Malaysia’s	Favour	50	11%	Kuala	Favour	30	6%
	Unfavour	18	9%	Lumpur	Unfavour	13	4%
Kuala	Favour	31	7%	Malaysia	Favour	10	2%
Lumpur	Unfavour	13	7%	Airlines	Unfavour	25	8%
ASEAN	Favour	21	5%	Asean	Favour	35	7%
	Unfavour	3	2%		Unfavour	2	1%
Najib	Favour	24	5%	Najib	Favour	20	4%
	Unfavour	7	4%		Unfavour	25	8%
Malaysians	Favour	11	2%	Mh370	Favour	4	1%
	Unfavour	7	4%		Unfavour	23	7%

In *CD1*, all concepts had a higher number of positive sentiment mentions compared to negative ones. Among them, “Malaysia”, “Malaysia’s”, “ASEAN”, and “Najib” demonstrated a clear positive inclination in both Co-Count and Likelihood. For “Kuala Lumpur”, although positive sentiment mentions dominated, the Likelihood ratio remained consistent between positive and negative, indicating that the reporting maintained a balanced tone without showing a strong

bias. However, for the concepts “Malaysian” and “Malaysians”, the Likelihood indicator showed a slightly higher proportion of negative sentiment compared to positive sentiment. It suggests that the media, when covering topics related to Malaysian citizens, tend to focus more on social issues or controversial events, resulting in a notable share of negative sentiment. The topics include events such as “child abuse cases”, “electoral fraud”, “unmarried Malaysian adults”, and “a Malaysian editor reprinted controversial Danish cartoons of the Prophet Mohammad”.

In *CD2*, “Malaysia”, “Kuala Lumpur”, and “ASEAN” demonstrate a strong positive sentiment tendency. From both the Co-Count and Likelihood perspectives, the number and proportion of positive sentiment mentions are significantly higher than those of negative sentiment. “Malaysian” similarly shows a trend where positive sentiment mentions far exceed negative ones, but the Likelihood ratio of positive and negative sentiments remains consistent. However, for the two themes related to the aviation disaster, “Malaysia Airlines” and “MH370”, negative sentiment dominates, with the number of negative mentions surpassing that of positive mentions. For example, reports mention that “Malaysia Airlines was in serious financial trouble” and “the tragic MH370 incident”. Coverage of “Najib” during this phase was predominantly negative, including allegations of graft, 1MDB was massively in debt, and criticism for playing golf with U.S. President Barack Obama during severe storms.

The most notable difference between *CD1* and *CD2* is the significant increase in the sentiment tendency toward “Malaysia”, “Malaysian”, “Kuala Lumpur”, and “ASEAN”. This improvement is primarily attributed to the deepening Malaysia-China relations, Malaysia’s active role in regional cooperation, and the enhanced international prominence of Kuala Lumpur as a global hub. In addition, media sentiment toward “Najib” shifted from largely positive to negative. Early praise reflected recognition of his economic and diplomatic efforts, while later criticism stemmed from the 1MDB scandal and political crises. The shift highlights evolving perceptions of his leadership and broader changes in Malaysia’s domestic and global standing. In *CD2*, although the MH370 incident initially harmed Malaysia’s international image, particularly in 2014 and 2015, coverage later emphasised the country’s positive performance. Key themes included strengthened Malaysia-China cooperation, infrastructure projects under the Belt and Road Initiative (e.g., the East Coast Rail Link), and Malaysia’s growing role in regional collaboration.

In summary, the depiction of *China Daily* on Malaysia’s national image in both stages was predominantly positive, with a noticeable shift toward an increasingly positive focus in *CD2* compared to *CD1*. The portrayal of the leader transitioned from positive to negative, a shift closely tied to key events during the tenure. It not only reflects the media’s reassessment of the leader’s performance and governance capabilities but also highlights the evolving public expectations regarding the leader’s image and government transparency. The attitude toward the MH370 incident was negative but did not persist for long. Over time, media coverage gradually shifted from focusing on the crisis management and negative impacts of the incident to broader themes of international cooperation and diplomatic affairs.

In addition, as the core subject of this study, Malaysia has experienced notable changes in its associated countries between *CD1* and *CD2*. Analysing the countries linked to Malaysia in the news helps to gain deeper insights into Malaysia’s international relationship network and its role in regional and global affairs. As shown in Table 4, “ASEAN”, “Indonesia”, “China”, “Singapore”, and “Australia” are identified as countries closely associated with Malaysia.

TABLE 4. Associated countries on “Malaysia” in CD1 and CD2

Name-like	CD1				Name-like	CD2			
	Count	Sentiment	Co-Count	Relevance		Count	Sentiment	Co-Count	Relevance
ASEAN	93	Favour	21	5%	China	651	Favour	196	38%
	69%	Unfavour	3	2%		67%	Unfavour	30	10%
Indonesia	63	Favour	9	2%	Indonesia	70	Favour	5	1%
	67%	Unfavour	12	6%		63%	Unfavour	11	3%
China	353	Favour	102	23%	Singapore	61	Favour	14	3%
	60%	Unfavour	15	8%		57%	Unfavour	5	2%
Singapore	50	Favour	7	2%	ASEAN	116	Favour	35	7%
	55%	Unfavour	3	2%		57%	Unfavour	2	1%
Australia	24	Favour	11	2%	Australia	71	Favour	13	3%
	42%	Unfavour	6	3%		50%	Unfavour	10	3%
Japan	28	Favour	9	2%					
	39%	Unfavour	3	2%					

In CD1, “ASEAN” as the core framework for regional cooperation, holds a high occurrence rate of 69%, making it an essential platform for Malaysia’s diplomacy. It not only highlights Malaysia’s active role in promoting regional economic integration and maintaining peace and security but also underscores its special status as one of ASEAN’s founding members. The frequent mentions of “ASEAN” in the Chinese media’s coverage of Malaysia effectively affirm Malaysia’s leadership and cooperative spirit within this regional organisation. “China”, as a country with predominantly positive sentiment in the CD1 phase, ranks third among the countries associated with Malaysia, accounting for 60% of mentions. This data reflects the relatively stable and positive nature of Malaysia-China relations during this phase, demonstrating that China had already become an important part of Malaysia’s diplomatic network. Next, “Singapore” and “Japan” rank fourth and sixth, respectively, among countries associated with Malaysia, with shares of 55% and 39%. However, in terms of sentiment, both countries show a balanced distribution of positive and negative emotions in the coverage, without any significant bias. The balance can be attributed to their interactions with Malaysia, which include both economic and trade cooperation as well as certain divergences in areas such as sports events. “Indonesia” and “Australia” were portrayed more negatively than positively, as reflected in the bolded table data. Indonesia accounted for 67% of mentions, ranking second, and Australia 42%, ranking fifth. It suggests a more balanced media approach, acknowledging both cooperation and challenges in bilateral ties.

In CD2, “China” overtook “ASEAN” as the country most closely associated with Malaysia, with 651 mentions, accounting for 67%. The rise indicates stronger bilateral ties, especially under the Belt and Road Initiative. Frequent diplomatic exchanges, trade deals, and economic cooperation fueled this development. Cultural, educational, and festival-related activities further underscored China’s role as a key partner for Malaysia. In contrast, while the number of “ASEAN” mentions slightly increased, its share dropped from 69% in CD1 to 57%. The change reflects a shift in media focus during the CD2 phase, with China’s growing association in Malaysia’s diplomatic network reducing the relative prominence of ASEAN as a multilateral regional framework. Nevertheless, this decline is not significant, as ASEAN remains a crucial platform for Malaysia in advancing regional economic integration and coordinating regional affairs. In trade agreements (e.g., RCEP) and regional peace and security cooperation, ASEAN’s role remains indispensable, continuing to support Malaysia’s leadership in regional cooperation. “Singapore” saw increases in its mention frequency, share, and sentiment tendency in CD2. The two countries deepened their collaboration in areas such as trade, transportation, and infrastructure. In particular, cross-border commuting, tourism recovery, and cooperation during the COVID-19

pandemic became focal points in reports. The expanding areas of cooperation significantly enhanced Singapore's presence in media coverage, underscoring the growing closeness of the bilateral relationship. "Australia" exhibited a slightly negative sentiment tendency in *CD1*, but in *CD2*, positive and negative sentiments were balanced, reflecting a diversification and stabilisation of the bilateral relationship. Australia's role in the MH370 search reflected strong bilateral cooperation in crisis response. Ongoing collaboration in education, trade, and culture remained a media focus, showing that these ties persisted despite fluctuations. The ranking and sentiment toward "Indonesia" in two stages showed little change, consistently placing it as Malaysia's second most associated country. However, negative sentiment outweighed positive sentiment, underscoring the complexity of their relationship and unresolved historical issues. The disputes are mainly concentrated on border trespassing, illegal fishing, worker treatment, and undocumented immigration. Despite these challenges, media coverage of the bilateral relationship also highlighted areas of mutual interest in economic cooperation, regional affairs, and environmental protection. The balanced narrative approach, addressing both cooperation and differences, effectively showcases the dynamic and multi-dimensional nature of Malaysia-Indonesia relations (Yaakub, 2009).

In terms of associated actors, the similarity between *CD1* and *CD2* lies in the fact that the associated actors remain largely consistent. This consistency reflects Malaysia's long-standing and stable partnerships within its international relationship network. Furthermore, the continuity of these associated actors highlights the coherence and structured nature of *China Daily's* coverage of Malaysia's international relations. However, in *CD1*, "Japan" was mentioned as an associated actor, whereas in *CD2*, it was no longer referenced. It indicates that Malaysia's diplomacy in the *CD1* phase exhibited a more diversified character. By contrast, in *CD2*, with the significant enhancement of Malaysia-China relations and China's increasing centrality in Malaysia's diplomatic network, Japan's importance gradually diminished, and related coverage decreased accordingly.

For sentiment tendencies, apart from "Indonesia", other countries showed a gradual increase in positive sentiment across both phases. It indicates that Malaysia's interactions and cooperative relationships with its key associated countries deepened over time, and the media coverage of these countries leaned toward a more positive narrative. The rise in positive sentiment not only reflects the acknowledgement of Malaysia's achievements in international cooperation but also signifies the increasing harmony and stability of Malaysia's relations with its associated countries in both regional and international contexts.

In summary, the evolution of themes, the representation of social actors, and the shifts in sentiment collectively demonstrate *China Daily's* strategic adjustments in constructing Malaysia's national image across different periods. To a large extent, these developments reflect the deepening bilateral relations between China and Malaysia. However, such trends should not be understood solely within the framework of bilateral ties; they must also be interpreted within broader political, diplomatic, and Media narrative contexts. First, following the launch of the BRI, Malaysia, one of the earliest and most active participants, has seen its strategic significance rise in China's external communication. Growing cooperation in infrastructure, energy, and finance has contributed to increased visibility and more positive portrayals of Malaysia (Wang & Liang, 2024). Second, changes in the regional geopolitical landscape, such as the advancement of the U.S. Indo-Pacific Strategy and shifts in China-ASEAN relations, have further elevated Malaysia's role in China's regional diplomatic agenda (Sinaga, 2020). Third, transformations within Malaysia's domestic political landscape have also influenced the pathways through which its national image is

constructed in Chinese media. Lastly, as a state-affiliated international communication platform, *China Daily* often exhibits a tendency to align with national strategies and emphasise cooperative outcomes. Accordingly, the media's construction of Malaysia's national image is, to some extent, constructed by China's foreign policy priorities and ideological framing.

CONCLUSION

This study employs Python to extract news reports from *China Daily* that are closely related to Malaysia, categorising them based on time periods: October 4, 2003, to October 4, 2013 (*CD1*) and October 5, 2014, to October 4, 2023 (*CD2*). A comparative analysis was conducted on the two corpora. Utilising the qualitative research software Leximancer 5.0, the researchers performed unsupervised text mining to systematically extract key conceptual themes, semantic relationships, and sentiment tendencies. The approach aims to comprehensively analyse the evolving trends in coverage across different historical periods.

In conclusion, Malaysia is portrayed in *China Daily* as a promising partner in economic growth and development. By extracting the Concept Map in *China Daily*'s news on Malaysia during the two phases, it is evident that in the *CD1* phase, the topics were relatively dispersed, encompassing a wide range of fields such as foreign policy, economic cooperation, social development, and government governance. However, in the *CD2*, the content gradually displayed greater thematic concentration and interconnectedness, indicating a more focused agenda in news reporting. The above conclusion is further supported by the automatic extraction of semantic relationships centred on the two core concepts, "Malaysia" and "Malaysian". The results reveal that from *CD1* to *CD2*, the coverage gradually transitioned from general news coverage to a more targeted focus on specific themes. The shift is evident not only in the increased concentration of coverage on major events and bilateral cooperation topics but also in the narrative style, which evolved from broad, general descriptions to more specific and directional storytelling.

Additionally, Malaysia is depicted as an important player in regional and global relations. By extracting Name-like Concepts representing prominent social actors in the news, it was found that nearly half of these concepts remained consistent within 20 years. Regardless of the broader news framework in *CD1* or the more concentrated and in-depth scope in *CD2*, the rankings of these core concepts remained unchanged. It indicates that the reporting during both phases consistently revolved around core themes, reflecting continuity in topics and narratives. Such stability underscores Malaysia's enduring role as an important regional partner and a key diplomatic player in Chinese media.

Finally, Malaysia is characterised as a stable and enduring strategic ally. Through sentiment analysis of Name-like Concepts related to Malaysia in the news, the study found that the proportion of positive sentiment in *CD2* increased further, indicating that the media's focus on Malaysia gradually shifted towards a more favourable direction. In contrast, the sentiment towards the image of Malaysian leaders showed a significant shift from positive evaluations in *CD1* to more negative descriptions in *CD2*. The change not only reflects the reassessment of the leader's performance but also highlights evolving public expectations regarding leadership, government transparency, and accountability. Notably, while the aviation incident was primarily presented in a negative light, the sentiment did not persist for long. Over time, media coverage shifted to broader topics such as international cooperation and diplomatic affairs. The study also analysed the sentiment trends toward countries associated with Malaysia. The findings reveal that the

countries associated with Malaysia in both phases remained largely consistent, reflecting Malaysia's long-standing stable partnerships in international relations and the continuity in *China Daily*. However, "Japan", mentioned in *CD1*, was absent in *CD2*, indicating a shift in Malaysia's diplomatic focus, with resources and attention being redirected accordingly. Meanwhile, positive sentiments towards all associated countries except "Indonesia" continued to grow, emphasising the deepening collaboration between Malaysia and its key partners.

The innovations of this study are primarily reflected in two aspects. First, it expands the application of text mining technology in discourse studies by integrating Leximancer 5.0, making a novel attempt to incorporate software into discourse research methodologies. Second, it focuses on the construction of Malaysia's national image, addressing the regional bias evident in previous national image studies. In addition, this study offers practical implications for Malaysia in optimising its strategies for national image construction in international communication. On the one hand, in light of the thematic concentration, positive sentiment, and cooperation-oriented narratives observed in Chinese mainstream media coverage, Malaysia could further enhance communication and collaboration with Chinese media by proactively providing authoritative and positive information resources, thereby constructing the image of a "reliable regional partner". On the other hand, Malaysia should strengthen its strategic communication capacity by aligning with media-reported thematic hotspots, such as the BRI, regional cooperation, and disaster response, so as to develop more targeted international communication content. Moreover, with regard to sensitive topics such as the portrayal of political leaders and political events, Malaysian media and relevant institutions need to establish more forward-looking discourse management mechanisms to mitigate potential negative impacts, thereby enabling the "sustainable construction" of national image in the international arena.

However, this study also has certain limitations. First, it does not delve deeply into the use of specific discursive strategies, lacking detailed analysis at the micro-level. Second, the scope of the research corpus is limited to *China Daily*, which restricts the inclusion of more diverse media sources or international perspectives. Future research could integrate Critical Discourse Analysis to explore the use of language in greater depth. Additionally, a cross-national perspective could be adopted to conduct comparative analyses of national image construction by others, offering a more comprehensive understanding of the diversity and complexity of national image construction.

NOTE

1. <http://www.chinadaily.com.cn>
2. https://www.gov.cn/guowuyuan/2013-10/04/content_2584735.htm
3. <https://www.yidaiyilu.gov.cn/p/86671.html>
4. <https://www.leximancer.com/>

REFERENCES

- Abas, N., Abdul Aziz, R., Turiman, S., & Mat Daud, N. S. (2023). A corpus-assisted critical discourse analysis of COVID-19 vaccination-related news discourse in the Malaysian mainstream media. *3L: Language, Linguistics, Literature*, 29(3), 148-165. <https://doi.org/10.17576/3L-2023-2903-11>
- Boulding, K. E. (1956). *The image: Knowledge in life and society* (Vol. 47). University of Michigan press. <https://doi.org/10.3998/mpub.6607>
- Chang, S. (2017). The multifaceted China: A study on the “other-constructed” national image of China--An example of the 2015 “China-U.S. relations” topic in *The New York Times*. *Journalism University*, 3, 138-145.
- Chavez, L. R. (2023). *Covering immigration: Popular images and the politics of the nation*. Univ of California Press. <https://doi.org/10.2307/jj.8501066>
- Chen, L. (2018). The discursive turn in the study of China’s image. *Foreign Language Research*, 3, 33-37.
- Chen, Z. T., & Whyke, T. W. (2022). Manufacturing and commodifying “Chineseness”: micro-celebrity Li Ziqi’s construction of an idyllic rural China and her media representation on YouTube. In *China’s International Communication and Relationship Building* (pp. 150-164). Routledge. <https://doi.org/10.4324/9781003254157-12>
- Fong, Y. L., & Ponnann, R. (2019). Framing the bilateral relations between Malaysia and China: The news coverage of flight MH370 SEARCH. *Journal of Media and Communication Research*, 11(1), 37-56.
- Fowler, R. (1991). *Language in the News*[M]. London: Routledge.
- Fowler, R. (2013). *Language in the News: Discourse and Ideology in the Press*. Routledge. <https://doi.org/10.4324/9781315002057>
- Goh, E., & Wilk, V. (2022). Showcasing Leximancer in tourism and hospitality research: a review of Leximancer-based research published in tourism and hospitality journals during 2014-2020. *Tourism Recreation Research*, 1-14. <https://doi.org/10.1080/02508281.2022.2129284>
- Karasti, H., Baker, K. S., & Millerand, F. (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work*, 19, 377-415. <https://doi.org/10.1007/s10606-010-9113-z>
- Khan, Z. A., Xia, Y., Ali, S., Khan, J. A., Askar, S. S., Abouhawash, M., & El-Rashidy, N. (2023). Identifying hot topic trends in streaming text data using sequential evolution model based on distributed representations. *IEEE Access*, 11, 98787-98830. <https://doi.org/10.1109/ACCESS.2023.3312764>
- Lemus-Delgado, D. (2023). Opening ceremonies and national identity: Beijing 2008 and Tokyo 2020. *National Identities*, 25(5), 463-482. <https://doi.org/10.1080/14608944.2023.2242792>
- Lin, Y. T., & Miao, X. W. (2023). Discourse construction of China’s national image in mainstream American media: A diachronic study based on text mining. *Journal of PLA University of Foreign Languages*, 46(3), 87-95.
- Liu, H., & Lim, G. (2020). The political economy of a rising China in Southeast Asia: Malaysia’s response to the Belt and Road Initiative. In *China’s New Global Strategy* (pp. 158-173). Routledge. <https://doi.org/10.4324/9780429317002-11>
- Lyudmila, S., Anastasia, L., & Ilya, S. (2021). Media representation of the image of the Russian political leader in western online media (on the material daily news and der spiegel). *International Journal of Media and Information Literacy*, 6(2), 396-405. <https://doi.org/10.13187/ijmil.2021.2.396>
- O’Kane, P., Smith, A., & Lerman, M. P. (2021). Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *Organizational Research Methods*, 24(1), 104-139. <https://doi.org/10.1177/1094428119865016>
- Papadopoulos, N., & Heslop, L. A. (2014). *Product-country images: Impact and role in international marketing*. Routledge. <https://doi.org/10.4324/9781315863931>
- Popova, E. O., Barabash, V. V., Burdovskaya, E. Y., Lavrentieva, M. Y., & Melnichenko, A. A. (2023). Russian Image Formation Through the Image of the President of the Country in the English-Language Media in 2020-2023 (Based on the Materials of the British and American Press). *Kurdish Studies*, 11(2), 4599-4617.
- Prieto, R. S. (2021). The Image of Spain in the Eyes of Austrian, Flemish, French, Italian, Polish and Bulgarian Facebook Users: Text-Linguistic Opinion-Mining for Detecting and Analysing National Stereotypes. In *National Stereotyping, Identity Politics, European Crises* (pp. 221-243). Brill. https://doi.org/10.1163/9789004436107_013
- Restad, H. E. (2020). What makes America great? Donald Trump, national identity, and U.S. foreign policy. *Global Affairs*, 6(1), 21-36. <https://doi.org/10.1080/23340460.2020.1734955>
- Robinson, P. (2000). The policy-media interaction model: Measuring media power during humanitarian crisis. *Journal of peace research*, 37(5), 613-633. <https://doi.org/10.1177/0022343300037005006>

- Rusciano, F. L. (1997). First-and third-world newspapers on world opinion: Imagined communities in the cold war and post-cold war eras. *Political Communication*, 14(2), 171-190. <https://doi.org/10.1080/105846097199425>
- Rusciano, F. L. (2003). The Construction of National Identity -- A 23-Nation Study. *Political Research Quarterly*, 56(3), 361–366. <https://doi.org/10.1177/106591290305600310>
- Sinaga, L. C. (2020). Xi Jinping, “China dream”, and Chinese military diplomacy to ASEAN. *Journal of ASEAN Studies*, 8(2), 173-190. <https://doi.org/10.21512/jas.v8i2.6483>
- Siyu, L. I. U., & Feng, J. I. (2023). International sports communication and national image construction by mainstream media--An example of CGTN’s coverage agenda of the Beijing Winter Olympics. *Journal of Physical Education/Tiyu Xuekan*, 30(1).
- Tai, K. W. (2023). Funds of knowledge for synchronous online language teaching: a translanguaging view on an ESL teacher’s pedagogical practices. *International Review of Applied Linguistics in Language Teaching*, 1-48. <https://doi.org/10.1515/iral-2023-0096>
- Tjellander, A. (2022). *Mining the imagery: A text mining and news media content analysis of the Swedish country image in the Guardian, 2010-2020* [Master’s thesis]. Uppsala University.
- Van Dijk, T. A. (2013). *News as discourse*. Routledge. <https://doi.org/10.4324/9780203062784>
- Van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford university press. <https://doi.org/10.1093/acprof:oso/9780195323306.001.0001>
- Wa-Mbaleka, S. (2020). The researcher as an instrument. In *Computer Supported Qualitative Research: New Trends on Qualitative Research (WCQR2019)* 4 (pp. 33-41). Springer International Publishing. https://doi.org/10.1007/978-3-030-31787-4_3
- Wang, D., & Liang, S. (2024). Discourse construction of Chinese modernization from the perspective of Malaysian media. *Humanities and Social Sciences Communications*, 11(1), 1-12. <https://doi.org/10.1057/s41599-024-03152-9>
- Wang, G. (2018). A corpus-assisted critical discourse analysis of news reporting on China’s air pollution in the official Chinese English-language press. *Discourse & Communication*, 12(6), 645-662. <https://doi.org/10.1177/1750481318771431>
- Whyke, T. W., Chen, Z. T., & Lopez-Mugica, J. (2022). An analysis of cultural dissemination and national image construction in Chinese influencer Li Ziqi’s vlogs and its impact on international viewer perceptions on YouTube. *The Journal of Chinese Sociology*, 9(1), 1-19. <https://doi.org/10.1186/s40711-022-00173-2>
- Wu, X., & Zhang, K. (2016). National Image: Concept, Characteristics, and Research Paths Revisited. *Modern Communication: Journal of Communication University of China*, 1, 57-62.
- Xu, W. W., & Wang, R. (2022). Nationalizing truth: digital practices and influences of state-affiliated Media in a Time of global pandemic and geopolitical decoupling. *International Journal of Communication*, 16, 29-30.
- Yaakub, A. N. (2009). *Malaysia and Indonesia: A Study of Foreign Policies with Special Reference to Bilateral Relations*. University of Western Australia.