HiLL PUBLISHING

Translation and Foreign Language Learning, 2026, 2(2), 384-389
https://www.hillpublisher.com/Journals/tfll/
ISSN Online: 3069-0315
ISSN Print: 3070-3077

A Multimodal Discourse Analysis of Beijing City Image Promotional Video—A Case Study of *This is Beijing*

Ruiqi Li

Jiangsu Normal University, Xuzhou 221116, Jiangsu, China.

How to cite this paper: Ruiqi Li. (2026) A Multimodal Discourse Analysis of Beijing City Image Promotional Video—A Case Study of *This is Beijing*. *Translation and Foreign Language Learning*, 2(2), 384-389. DOI: 10.26855/tfll.2026.02.029

Received: December 20, 2025
Accepted: January 6, 2026
Published: February 28, 2026

***Corresponding author:** Ruiqi Li, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China.

© 2026 by the author(s).
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) license, which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and is not modified or adapted.
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract

Under the background of globalization, city promotional videos serve as an important vehicles for showcasing the charm of cities and carrying out cross-cultural exchanges. Promotional videos usually combine visual, verbal, and auditory resources to construct a city's image through coordinated multi-semiotic strategies. Based on relevant theories such as Halliday and Zhang Delu, this study selects *This is Beijing* as a research case to explore how the video can build a city image with the help of various modal symbols. Through qualitative analysis of images, subtitles and audio, the study found that the video shapes Beijing as an international metropolis with profound historical background, scientific and technological development vitality, and open and inclusive characteristics. Local cultural symbols, bilingual narration, visual composition, and sound design work together to improve the external publicity and communication effect of the city. This research enriches the empirical research materials of multimodal discourse analysis, and also provides reference for other cities in China to optimize the production of external promotional videos.

Keywords

City image promotional videos; *This is Beijing*; Multimodal discourse analysis; International communication

1. Introduction

The image of a city directly influences public perception and constitutes a key factor in urban competitiveness. In the era of digital media, promotional videos rely on rich audiovisual performance to showcase the unique style of the city and become the core tool to improve the international popularity.

The image of the city is not merely a display of landscape but the result of various symbols interacting to create a sense of identity and belonging. *This is Beijing* is a typical example, it does not limit itself to showing landmark buildings but combines traditional cultural symbols, such as the Forbidden City, with technological innovation and daily life scenes of citizens to reveal Beijing that balances ancient characteristics with modern vitality.

While existing research focuses on brand communication strategies, few studies have analyzed the inherent logic of multimodal discourse in such videos. Scholars such as Halliday (1978), O'Halloran (2011), and Kress and Van Leeuwen (2020) have provided the theoretical foundation for this field. Bateman (2022) emphasizes empirical research, while Bateman and Tseng (2023) demonstrate how multimodal analysis reveals narrative strategies in audiovisual texts. This article discusses how various symbols jointly shape Beijing's international image, and improve the efficiency of overseas exchanges.

2. Literature Review

2.1 Studies on Multimodal Discourse Analysis

Multimodal discourse analysis is rooted in social semiotics and systemic functional linguistics. Halliday (1978) defined language as a social semiotic system, laying the theoretical foundation for communication research. Jewitt (2009) emphasized the relationship between symbolic resources and social practices. O'Halloran (2011) advocated integrating visual, verbal, and auditory resources into a unified analytical system. Bateman (2014) and Aiello and Parry (2019) further argued that visual and verbal modes are interconnected systems constructing social identities and cultural meanings.

Kress and Van Leeuwen (2020) proposed a foundational visual grammar, suggesting that images possess three metafunctions: representational meaning, which depicts objects and actions; interactive meaning, which establishes relationships between the image and the viewer; and compositional meaning, which organizes elements within the visual space. The theoretical framework has a profound impact on subsequent relevant research.

2.2 Studies on City Image Communication

City image is a symbolic construction product formed by integrating cultural, historical and public elements. The complete and three-dimensional city image includes not only physical elements such as architecture and infrastructure, but also intangible contents such as culture, lifestyle and collective historical memory. From the perspective of multimodal critical discourse analysis, Machin (2013) argued that visual and linguistic resources reproduce social meanings and ideological positions. Recent research on Chinese city branding also shows that promotional videos construct city identities through historical narratives, visual symbols and modern values (Wang & Feng, 2023). City promotional video is a multimodal text that highly integrates various symbols, and is a communication bridge for cities to connect with global audiences. Carefully selected pictures, narratives and sound effects can highlight the characteristics of the city and reshape the audience's inherent stereotype of the city.

However, compared with commercial advertising and tourism publicity research, the results of using multimodal theory to analyse domestic city promotional videos are still scarce, especially there is a lack of in-depth research on how various symbolic elements can link to shape the overseas image of Chinese cities. Based on it, this article combines the multimodal discourse analysis with the research of city image communication, and explores the complete logic of the video's construction of Beijing's city image with *This is Beijing* as a case.

3. Theoretical Framework and Research Questions

3.1 Theoretical Framework

This study adopts Zhang Delu's multimodal discourse analysis framework, which integrates systemic functional linguistics and social semiotics, and is divided into four analytical levels: culture, context, content and expression. While frameworks like Kress and Van Leeuwen's visual grammar focus heavily on visual composition, Zhang's model is uniquely suited for analyzing Chinese audiovisual texts because it explicitly integrates cultural and contextual dimensions, allowing for a deeper exploration of how ideological and social factors inform the choice of semiotic resources.

More specifically, the cultural level focusses on the ideology and cultural connotation embedded in multimodal text, and analyses how various symbols carry group identity and local cultural traditions; the context level explores the relationship between the text and the communication scene, and analyses how the audience positioning, communication goal, social environment and other factors influence the meaning generation; the content level disassembles various symbols such as text, pictures, sound effects, etc., and analyses how different modalities coordinate to convey information and construct a complete expression; the external presentation form of the expression layer focusses on symbols, covering visual composition, text expression, audio effects, and various shooting production techniques.

3.2 Research Questions

Based on the above theoretical analysis framework, this paper raises three research questions:

(1) What kind of cultural image of Beijing is built in the promotional video *This is Beijing* with the help of various multimodal symbols?

(2) How can the visual, verbal, and auditory resources jointly present Beijing's historical heritage, modern development and international characteristics?

(3) What references can the multimodal communication methods used in this video bring to other Chinese cities to build an international city image?

4. Research Object and Methodology

The research object is the city promotional video *This is Beijing* launched by People's Daily on New Year's Day in 2025. The video is aimed at the global audience and shows Beijing's history and culture, urban development and international appearance in an all-round way. Unlike traditional landmark-focused promotional videos, it takes story narration as the main line, integrating bilingual subtitles, voiceovers, images, soundtracks and other multiple symbols, and organically connects historical monuments, scientific and technological innovation, cultural activities and citizens' daily life, taking into account tradition and modernity, which is very suitable for multimodal research from the perspective of cross-cultural communication.

Based on Zhang Delu's multimodal discourse analysis framework, this research adopts a qualitative approach to analyse how various symbolic elements work together to shape the city image from the four dimensions of culture, context, content and expression.

5. Multimodal Discourse Analysis of This is Beijing

5.1 Culture Level: Constructing Identity

Zhang Delu (2009) proposed that the cultural hierarchy is the basic dimension of multimodal discourse analysis, carrying the value orientation, ideological connotation and cultural heritage behind the communication. The promotional video *This is Beijing* combines traditional cultural symbols, historical narratives and modern values to build a unique city cultural identity of Beijing.

5.1.1 Historical Symbols and Cultural Continuity

The most intuitive way for the promotional video to shape the city image is to use historical and cultural symbols to pave the narrative. The opening part of the video presents the high-definition scenes of the Forbidden City, the Great Wall and the Temple of Heaven. These landmarks serve not merely as tourist attractions, but as embodiments of the millennia-long history of Chinese civilization.

According to the visual grammar, images convey meaning by portraying specific characters, scenes and objects. The mediaeval architecture of this video occupies the core of the picture, with a regular layout, magnificent volume and traditional decoration, which intuitively conveys the strong and calm cultural temperament. For overseas audiences who are not familiar with Chinese history, this set of visual language is easy to understand, and it can be perceived at a glance that this is a city with deep historical accumulation. The Forbidden City is not only a building, but also symbolizes the history, artistic achievements and cultural inheritance of the dynasty; the Great Wall is not just a wall, but represents the tenacious national spirit and identity.

In short, the video shapes Beijing with deep historical roots through the screening and arrangement of various cultural symbols, leaving a deep first impression on the audience.

5.1.2 Integration of Tradition and Modernity

Although the video uses a lot of historical and cultural elements, it does not shape Beijing as an ancient city that stagnates in the past but takes the initiative to integrate traditional culture and modern development to build a narrative vein that connects history and innovation.

The camera is constantly switching, for example, sometimes it presents a four-way transportation network, science and technology park, international convention and exhibition centre, and sometimes it cuts back to the ancient wall and the royal palace. The ancient and modern pictures are intertwined, opening up the connection between the cultural heritage and the future of the city. For example, the Forbidden City carries a thousand-year history, and the Zhongguancun Science City represents the vitality of scientific and technological innovation. Under the comparison of the two groups of pictures, the image of Beijing is both a cultural ancient capital and an innovative highland.

Historical, scientific, and technological symbols are interwoven throughout the video, illustrating that cultural inheritance and modernization are not mutually exclusive but can progress in tandem.

5.1.3 Cultural Identity and International Recognition

Another outstanding feature of the urban cultural construction of the video is the translation and interpretation of local cultural symbols, giving overseas audiences a connotation that can understand.

The Great Wall and the Forbidden City have long been globally recognized symbols of Chinese culture. The use of such pictures can lower the threshold of cross-cultural understanding and allow audiences from different cultural backgrounds to quickly establish their understanding of Beijing. At the same time, the video also adds cultural elements with relatively low overseas recognition, including Beijing's central axis, traditional opera, various museum venues, etc., broadening the audience's cognitive boundaries of Beijing, breaking the stereotype of a single moment, and presenting a rich and three-dimensional urban cultural landscape.

The promotional video combines the world-renowned iconic symbols with characteristic local cultural details to create a unique city image for *This is Beijing*: it not only retains distinctive local cultural characteristics, but also can successfully complete cross-cultural communication, which is easy for overseas audiences to understand and accept.

5.2 Context Level: Constructing an Inclusive and International City Image

The context dimension mainly analyses how the communication scene, audience and social relations influence the meaning construction of the text. *This is Beijing* is created for international audiences. The video focuss on dispelling cultural barriers and establishing emotional connections with global audiences.

5.2.1 Audience-Oriented Communication Strategies

The video does not list the resources owned by Beijing in one direction, but focuses on the real life of local residents, transforming the publicity mode from "urban landscape display" into "people-centered communication narrative".

A large number of ordinary citizens, foreign tourists, cultural activities and daily scenes in the market appear in the video, so that the audience has a sense of immersion in the city. For example, the pictures of foreign tourists visiting monuments, experiencing traditional culture, and interacting with local residents make Beijing no longer a collection of cold political and historical landmarks, but an open, inclusive, liveable and friendly city facing the world.

5.2.2 From Symbolic City to Livable City

Another communication strategy in the context level is to transform Beijing from a symbolic historical city into a livable urban space. Traditional city promotional videos focus on landmark display. *This is Beijing* broadens the scope of shooting, covering public transportation, urban public spaces, cultural activities, ecological environment, daily life and other contents, changing the audience's perception: Beijing is no longer only a historical monument, but also a city full of fireworks. Instead of abstract national symbols, the video showcases authentic scenes of people's work, social interactions, and daily lives. This narrative strategy enhances the realism of the city's image, portraying Beijing as a people-oriented, globalized international metropolis.

5.2.3 International Communication and Cultural Translation

Aimed at a global audience, *This is Beijing* employs specific cultural translation and communication strategies. Through the use of bilingual subtitles, global symbols, and popular cultural commentary, the video transform the connotation of local culture into information that is accessible to international audiences. O'Halloran (2011) proposed that multimodal communication relies on multimodal cooperation to achieve efficient meaning transmission. In the video, the visuals provide an intuitive visual experience, while the narration complements them with the corresponding cultural background. Relying on this set of multimodal translation methods, Beijing's local cultural symbols are successfully adapted for overseas dissemination.

5.3 Content Level: Interaction of Textual, Visual, and Auditory Modes

Drawing on Zhang Delu's (2009) framework for multimodal discourse analysis, the content dimension focuses on how various symbolic resources can collaboratively construct the meaning. Within *This is Beijing*, textual, visual and auditory modes do not operate independently, but form a integrated system where various modes mutually reinforce the narrative.

5.3.1 Textual Mode: Providing Cultural Information and Guiding Interpretation

The text information in the video is primarily conveyed through bilingual subtitles, place name annotation and narration. Compared to the visuals, the text clearly explain the profound cultural connotations, thereby helping overseas

audiences better understand the video's content.

The most prominent feature is the use of bilingual subtitles. The video contains 34 subtitles, which can be categorized by function into three themes: historical symbols, cultural institutions, global engagement and innovation. The subtitles serve as more than just a simple translation, but also as a bridge connecting Chinese culture with international audiences. Without such textual commentary, it would be difficult for overseas audiences to understand specific historical allusions. Effectively, the subtitles transform local cultural elements into information accessible to global audiences.

Table 1. Classification of Textual Resources and Their Communicative Functions in *This is Beijing*

Category	Examples	Communicative Function
Historical symbols	The Forbidden City; The Great Wall; The Temple of Heaven; Beijing Central Axis	Construct historical identity
Cultural institutions	National Museum of China; National Art Museum of China; National Center for the Performing Arts	Represent cultural vitality
Global engagemen and innovation	The Belt and Road Forum; China International Fair for Trade in Services; Zhongguancun Science City	Construct an open and innovative image

5.3.2 Visual Mode: Symbolic Representation and Spatial Narrative

This is Beijing uses visual elements as its core medium to shape the audience's impression of the city by showcasing landmarks such as the Forbidden City and the Temple of Heaven. Extensive use of wide-angle and symmetrical shots conveys a sense of solemnity and highlights its millennial history. In contrast, when transitioning to modern urban scenes, the camera rhythm noticeably accelerates. Through dynamic shots and rapid editing, the video presents skyscrapers, highways and scientific parks, highlighting the city's vitality, efficiency and innovation. Together, these two distinct visual styles trace the full trajectory of Beijing's development.

5.3.3 Auditory Mode: Emotional Construction and Cultural Atmosphere

Recent studies have increasingly recognized the significance of auditory resources in city image construction. Lang (2024), for instance, incorporated auditory embeddings into a multimodal framework to analyze Hangzhou's promotional videos. In this study, the audio serves as more than a background soundtrack; it changes synchronously with the visual scene. Historical and cultural segments utilize traditional folk music to evoke a solemn atmosphere, whereas sections highlighting scientific and technological innovation feature upbeat to underscore the vitality of the city. Furthermore, the narration employs the Beijing dialect rather than standard Mandarin, lending a down-to-earth, friendly, and approachable tone. This instantly dispels the psychological distance from the capital, shaping an image of a vibrant and accessible city. Coupled with ambient sounds such as street traffic and birdsong in parks, these auditory elements enhance realism and deepen the audience's immersive experience. Ultimately, the audio is not arbitrarily assembled but is carefully orchestrated with the visuals to highlight Beijing's unique blend of tradition and modernity, as well as its profound yet approachable urban character, thereby reinforcing the audience's overall impression.

5.4 Expression Level: Integration and Realization of Multimodal Resources

The level of expression focuses on the concrete presentation and coordination mechanism of various symbols. *This is Beijing* relies on editing, subtitle design and camera work to realize the seamless integration of language, visuals and audio.

The visuals in the video are highly synchronized with the narration: as historical landmarks appears on screen, the voiceover and subtitles simultaneously provide the relevant cultural context. This ensures immediate comprehension for overseas audience and maintains an optimal pace of information delivery. Furthermore, the subtitle design is exquisite, and the English translation is concise and fluent, conveying comprehensive information without compromising the visual aesthetics of the frames. The editing is mature and seamless, facilitating natural transitions between ancient and modern scenes. By establishing clear chronological threads, these seemingly disparate shots are woven into a cohesive narrative of Beijing's development.

6. Conclusion

Drawing upon multimodal discourse analysis, this paper investigates the strategies employed in the promotional video *This is Beijing* to construct the city image as an international metropolis. Integrating Zhang Delu's analytical framework with classical multimodal theory, the study conducts a systematic examination across four dimensions of culture, context, content and expression.

The findings indicate that the video constructs the city image by leveraging historical and cultural symbols, narratives of modern development, bilingual subtitles, visual composition and auditory elements. Collectively, these modalities present a modern Beijing characterized by profound cultural heritage, technological vitality, inclusivity, and global connectivity. Theoretically, this study validates the applicability of multimodal discourse analysis to city image communication, demonstrating that such an image is not the product of a single symbol, but rather the result of synergistic multimodal resources. Practically, the analysis suggests that effective international communication for premier cities needs to balance the cultural authenticity with the receptivity of global audiences.

This study has certain limitations. The research sample consists of a single promotional video, and the analysis is primarily qualitative. Future research could incorporate audience reception questionnaires and corpus analysis, or select multiple videos for comparative study, to more comprehensively explore multimodal strategies in urban international communication.

Funding

This paper is a phased achievement of the 2026 Jiangsu Province Postgraduate Research and Practice Innovation Program project "A Study on the Effects of Prompt Engineering on Large Language Models' English Translations of *Liaozhai Zhiyi* (26CXJH5081)."

References

Aiello, G., & Parry, K. (2019). Visual communication: Understanding images in media culture. *Communication and Critical/Cultural Studies*, *18*(1), 1-18.

Bateman, J. A. (2022). Growing theory for practice: Empirical multimodality beyond the case study. *Multimodal Communication*, *11*(1), 63-74.

Bateman, J. A., & Tseng, C. I. (2023). Multimodal discourse analysis as a method for revealing narrative strategies in news videos. *Multimodal Communication*, *12*(3), 261-285.

Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.

Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G., & Van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge.

Lang, L. (2024). The Asian Games and city branding in China: Multimodal critical discourse analysis of Hangzhou's promotional videos on YouTube. *Place Branding and Public Diplomacy*, *20*(4), 504-516.

Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies*, *10*(4), 347-355.

O'Halloran, K. L. (2011). Multimodal discourse analysis. In *The Bloomsbury handbook of discourse analysis* (pp. 249-282). Bloomsbury.

Wang, Y., & Feng, D. (2023). History, modernity, and city branding in China: A multimodal critical discourse analysis of Xi'an's promotional videos on social media. *Social Semiotics*, *33*(2), 402-425.

Zhang, D. L. (2009). Multimodal discourse analysis: Exploring the comprehensive theoretical framework. *Foreign Languages in China*, *6*(1), 24-30.