

The Construction and Dissemination of Beijing's City Image in Douyin Short Videos from a Multimodal Perspective

Jianuo Shi

Beijing International Studies University, Beijing, China

Abstract

Short videos have become an important medium to spread and promote the city image. Based on the visual narrative theoretical framework of multimodal discourse analysis, this paper selects three representative popular short videos related to Tiktok on the platform and analyzes the visual narrative meaning of short videos of the city image of Beijing from three levels: representation meaning, interaction meaning and composition meaning. The article focuses on various language forms in short videos, including text, images, music, etc., analyzing how these modalities are intertwined. The study found that these videos present the urban image of Beijing, which combines traditional and modern elements, a dream chasing city, and a sense of national pride and belonging. This provides reference for other cities to enhance their urban image in the dissemination of short videos.

Keywords

Multimodal discourse analysis; Short videos; Beijing's urban image

1. City image and short video communication

Lynch, an American scholar, first put forward the concept of "city image". In his book *The Image of the City*, he mentioned that a city can be viewed as a story, and the process of constructing the urban image is the interaction between people and the city. Therefore, how to tell a good city story is an important way to build a city image. Since the 1990s, mass media research has been integrated with the cultural field. Each city has its own unique culture. As a capital, Beijing enjoys unparalleled advantages. It not only boasts a rich historical and cultural heritage but also serves as the national political center, cultural hub, international communication center, and technological innovation center. Short videos, as one of the current mainstream communication channels, have distinct advantages in terms of communication subjects, content dissemination, fragmentary editing and dissemination, and communication speed. On the Douyin platform, short videos related to the city image of Beijing can be divided into two main categories: PGC (Professional Generated Content) and UGC (User Generated Content). In this context, users can edit and publish videos (such as vlog, video blog, plog) to record their daily travel experience and show local life. UGC (User Generated Content) breaks the previous trend where city brand communication was dominated by authoritative officials. Some studies have found that compared with PGC (Professional Generated Content), people are more willing to believe in the content generated and disseminated by UGC. A large number of high-quality UGC videos attract the attention of the audience and add new urban memorable moments, which is very important for the dissemination of urban image. This article takes short videos related to Beijing posted by users on the Douyin platform as a case study, analyzing how ordinary users use multimodal means to construct and disseminate the urban image of Beijing.

2. Multi-modal discourse analysis theory

With the development of society and continuous advancements in media technology, there is a noticeable trend toward diversification in people’s social mode. Social modes of communication that were once predominantly oral or written have evolved into a more comprehensive utilization of multiple senses, including auditory, visual, and tactile. In this process of transformation, human discourse is no longer confined to a singular language mode, increasingly exhibiting multimodal characteristics. Multimodal discourse is composed of various linguistic forms and modes, where different symbols interweave to create language meaning. This includes not only the language itself, but also multiple modes such as images, colors, sounds, music, spatial elements, and bodily movements, collectively forming a complex linguistic symbol set. In the field of systemic functional linguistics, Halliday (1985) argues that language modalities have three meta-functions: conceptual function, interpersonal function, and discourse function. Based on the research of functional linguistics, Kress and Leeuwen proposed that visual modality also includes three parts: representative meaning, interactive meaning, and compositional meaning. This paper employs the multimodal visual analysis framework developed by Kress and Van Leeuwen to analyze three popular short videos on the Douyin platform related to Beijing. By exploring the application of different modalities in short videos and how these modalities cooperate with each other, this study aims to provide some suggestions for other cities to enhance their urban image and construct a distinctive city identity in the realm of short video dissemination.

3. Multimodal discourse analysis of Beijing- related short videos on the Douyin platform

By using the search bar on the Douyin platform and entering the keyword “北京” (Beijing), along with selecting the automatic sorting function under the “最热” (Hottest) section. The author specially selected the videos published between 2022 and 2023, and comprehensively considered multiple factors such as the number of likes, comments, and favorites of the videos, and the most popular Beijing promotional short videos were selected for analysis (see Table 1). The focus of the study is to conduct a detailed analysis of non-verbal modalities in terms of editing techniques, camera transitions, and video soundtracks. By delving into the role of these symbols in generating discourse meaning, constructing content, and expression, the aim is to understand the interaction between different modalities, and then explore how they jointly construct and present the image of Beijing.

(Table 1)

Number	Title	Duration	Author	Release Time	Likes	Comments	Favorites	Forwarding
1	Challenge to Shoot a Series of Urban Promotional Films – China Beijing	38s	Diandi Lahai	2022/7/25	226.6w	10w	5.5w	9.1w
2	Sunset, Life Begins. Today, Can You Let yourself Go First?	1min35s	FangQi kiki	2023/11/21	166.2w	1.8w	6.5w	6w

3	Let's go to see the flag-raising together, this is a romance unique to us Chinese	20s	Time Capsule	2022/9/29	133.9w	7.7w	7.4w	13.5w
---	---	-----	--------------	-----------	--------	------	------	-------

3.1. The combination of ancient and modern

Representative meaning in this context refers to the interactions among characters, locations, and objects in the video, enabling the audience to understand their relationships and then form specific perceptions and feelings based on personal experiences. This aspect mainly focuses on the participants of the video image, the construction of visual processes, and the components of the surrounding environment, exploring how these factors represent the objective world of people, places, and scenes. In the visual narrative framework of the first short video, “挑战拍一组城市宣传大片之中国·北京”(Challenge to Shoot a Series of Urban Promotional Films – China Beijing), the author’s opening statement, “挑战拍一组城市宣传大片” (Challenge to shoot a series of urban promotional films), narrows the psychological distance with the audience. Subsequently, in the brief 42-second video, scenes shift every two seconds, showing ten activities with Beijing characteristics carried out by the author in Beijing. (First: Climbing the Great Wall, Second: Visiting the Hutongs, Third: Drinking Douzhi, Fourth: Playing Chess, Fifth: Drawing Facial Masks, Sixth: Eating Hot Pot, Seventh: Making Clay Rabbits for the Mid-Autumn Festival, Eighth: Eating Sugar Figurine, Ninth: Eating Beijing Roast Duck, Tenth: Exploring the Forbidden City). The video seamlessly incorporates Beijing’s unique historical landmarks, such as a bird’s-eye view of the Great Wall and cross-dressing photos taken against the backdrop of the Forbidden City. Simultaneously, it portrays the lifestyle of Beijing residents, featuring scenes of elderly individuals strolling through the hutongs, carrying birdcages, and locals sitting at the hutong entrance engaged in a game of chess. From the perspective of tourists, the video showcases the leisurely and pleasant entertainment of local life in Beijing.

In a concise format, the video encapsulates the rich cultural experiences of tourism in Beijing, cleverly utilizing a narrative style popular among young audiences to present distinctive activities for visitors in Beijing.

The interactive meaning of the first short video can be analyzed by examining aspects such as the social distance, attitude, and modality between the audience and the depicted world. Regarding social distance, the focus is on the images presented within the framing of the camera, directly impacting the relationship between participants and viewers. [6] At the beginning of the video, the progression of the camera, transitioning from a distant view to a close-up, narrows the distance between participants and viewers. This shift from a distant to a close-up perspective allows the audience to feel as if they are present, experiencing Beijing’s activities alongside the participants. In terms of attitude interaction, the video predominantly employs the shooting perspective to convey its message. The extensive use of a horizontal perspective signifies an equal relationship and a sense of collective participation between viewers and participants. Additionally, the video employs a scene transition editing technique with a cut every 2 seconds, providing visual impact for the audience. This dynamic approach vividly showcases the rich urban landscapes and distinctive activities in Beijing. In terms of modality, the video presents a nostalgic and retro feel through color manipulation. The rich color layers and saturated colors of the picture not only enhance the audience’s artistic experience, trigger the wonderful imagination of Beijing, but also emotionally resonate with the viewers. The entire video emphasizes the characteristics of Beijing as a historic and cultural city and also

presents the city as a charming and enchanting place through an artistic and cultural lens. The significance conveyed by this video is that Beijing is a unique city, showing its profound cultural heritage and humanistic sentiments. This portrayal makes the short video highly captivating when depicting the urban image of Beijing, eliciting a strong emotional response from the audience.

Compositional meaning involves the organic integration of graphic elements and interactive components, forming a meaningful whole, primarily manifested in the dimensions of content value, salience, and framing. The text in the video mainly consists of the names of various scenic spots and activities in Beijing and is arranged at the top of the screen. In this way the information in the center of the image is more prominent. This arrangement allows the audience to quickly focus on key information, enhancing visual salience. Additionally, in this video, there is a harmonious interplay between the music modality and the image modality. At the 20-second mark, the music modality transitions to Peking Opera, while the image modality simultaneously presents scenes of Peking Opera performances in the Forbidden City and a series of cross-dressing photographs of the creators wearing costumes that blend modern and traditional elements. Through the vivid presentation of images, the ingenious combination of Peking Opera and contemporary life is vividly presented, creating a sense of harmonious coexistence across time and space. This interaction between music and images facilitates a deeper understanding of Peking Opera culture, providing the audience with a richer and more diverse experience. Compositional meaning, achieved through thoughtful text layout and the skillful combination of audio and visuals, not only optimizes information delivery in the video but also presents an organic fusion of Peking Opera and modern culture, offering the audience a captivating audio-visual feast across the era.

3.2. City of Dreams

The second short video, “日落起，生活起。今天，能不能先放过自己” (Sunset, Life Begins. Today, Can You Let yourself Go First?) with a duration of 1 minute and 35 seconds, begins with the night scenery of Beijing. The creator vividly portrays the lifestyle of young individuals striving in Beijing by showcasing busy city streets, modern architecture, and emotionally charged narration. This imagery brings the audience closer to the creator, evoking a strong resonance among the hardworking urban community. Subsequently, the creator expresses admiration for Beijing’s sunset, transitioning the video with the narration “It’s okay, even if the sun has to rest, let yourself go today.” The visuals shift to the sunset at the Temple of Heaven, the picturesque scenery of the Summer Palace, and the city’s cold office buildings transforming into colorful trousers under the setting sun. These few minutes of sunset seem to soothe the people struggling in the city, conveying a sense of tranquility and release. In terms of compositional meaning, both the music modality and the image modality in the video mutually reinforce each other. For example, at the 7-second mark, the image modality transitions from a panoramic view of Beijing’s night scene to a focus on the creator, while the music modality shifts from music with operatic elements to the creator’s monologue and alternating light music. This combination of sound and image enhances emotional resonance, guiding the audience through the creator’s unique emotional expression for the city of Beijing. Through clever editing and the use of music, this video showcases the distinctive charm of Beijing as a city of dreams, making the short film more emotionally engaging in conveying the urban image of Beijing.

3.3. National pride and sense of belonging

The third short video “艾特ta一起去看升旗吧 这是属于我们中国人独有的浪漫” (Let’s go to see the flag-raising together, this is a romance unique to us Chinese) with a duration of 20 seconds, uses images and text in a narrative manner to showcase the unique flag-raising ceremony at Beijing’s Tiananmen Square. The video features clear visuals, lively background

music, emotionally rich narration, and vivid color contrasts. The people in the picture, regardless of age, are smiling and waving the national flag in their hands, while the soldiers look solemn and solemn and walk neatly. The creator skillfully employs the representational meaning of images, establishing an emotional foundation and bridging the distance with the viewers. The video particularly highlights the modal aspect in the interactive meaning, using sensory details to elicit emotional resonance between the images and the audience. For example, the national flag appears multiple times in the video, from small flags waved by people to the appearance of the national flag guard in a neat military posture with the national flag, and then to the scene of the raising of the Five-Starred Red flag under the first rays of morning sunlight. The national flag holds sacred significance in the heart of every Chinese person, symbolizing the prosperity of the nation. The rising of the Five-Star Red Flag not only represents the “Chinese Dream” and national pride but also showcases the profound cultural heritage and national image of Beijing as the capital of China. In addition, the narration in the video, stating “The daily flag-raising ceremony is a solemn national ritual, it is a place that people of all ethnic groups across the country yearn for.” coupled with uplifting and inspiring music, elevates the audience’s emotions to the highest point. This greatly arouses the audience’s sense of national pride and belonging, inspiring a longing to be present in the capital, Beijing. Through the interaction of multiple modalities, including scenes, narration, and music, the video successfully engages the audience’s visual and auditory senses, establishing an emotional bond and intimate connection between the city of Beijing and the viewers of the video.

4. Conclusion

As a new carrier for the construction of urban image, short video comprehensively uses various multi-modal symbols to present a diverse and vibrant Beijing to the audience. In short video communication, the expression of emotions is crucial. Through the interactive use of various modalities, short videos successfully evoke emotional resonance in the audience, inspiring their unique feelings towards Beijing. This emotional transmission not only serves as the goal of urban image communication but also becomes a key factor influencing audience attitudes and behaviors. These short videos showcase the charm and appeal of Beijing’s cosmopolitan metropolis by multi-perspective storytelling and carefully designed imagery. They effectively convey the rich connotations of Beijing’s urban image, transforming it from a simple definition of a modern city into a city where tradition harmoniously coexists with modernity, and solemnity intertwines with vitality. More importantly, it shows that Beijing is a city of dreams, and that it can inspire a strong feeling of national pride and a sense of belonging. In the future, with the continuous development of social media and short video platforms, the construction and dissemination of urban image will face more challenges and opportunities. In order to better respond to this trend, we need to dig deeper into the potential of multimodal discourse, expand the means of expression, and create higher-quality short videos for the communication of urban images.

References

- [1] Lynch K. (1960). *The image of the city*. MIT press.
- [2] Liao Peiyi (2018) *Comparison of social media content production methods between UGC and PGC*. News Research Guide (16), 98
- [3] VEASNA S,WU W Y,HUANG CH.. (2013). *The impact of destination source credibility on destination satisfaction: The mediating effects of destination attachment and destination image*. Tourism Management.
- [4] Zhang Zuocheng, Chen Yumin (2011). *Review of Three Theoretical Methods of Multimodal Discourse Analysis*. English Education in China

- [5] KRESS, G.& van Leeuwen T. (2021). *Reading Images:The Grammar of Visual Design (3rd ed.)*. London:Routledge.
- [6] KRESS, G.& van Leeuwen T. (2021). *Reading Images:The Grammar of Visual Design (3rd ed.)*. London:Routledge.
- [7] Chen Jun (2015). *Multimodal Discourse Interaction Interpretation of Urban Tourism Promotional Videos - Taking Huai'an Tourism Promotional Videos as an Example*. *Modern Communication* (09), 51-53
- [8] Jia Yanyun (2021). *Multimodal Discourse Analysis of Urban Image Construction - Taking the "Best Jinan" City Image Promotion Video as an Example*. *Modern Communication* (19), 94-96