Analysis of Phenomenon of Tianshui Malatang's Soaring Popularity from a Multimodal Perspective

Chang Liu[1,a,*]

[1]East China Normal University, Shanghai, 200062, China
a. lccl1023@163.com
*corresponding author

Abstract: Contemporary communication studies show that short videos, as the most important form of information dissemination today, have a more multidimensional and immersive communication effect compared to traditional print media and other channels. Since this information medium inherently contains numerous social symbols with expressive characteristics, this paper will analyze the short videos that led to the viral phenomenon of "Tianshui malatang" based on Kress and Leeuwen's visual grammar and Halliday's theories of language as a social semiotic. This paper analyzes the phenomenon of the "Tianshui malatang" a fusion of regional tourism and local specially food that has gone virtual on Chinese short video platforms, using a multimodal discourse analysis approach. The analysis shows that the "Tianshui malatang" video enhances the sensory interaction between the viewer and the screen in terms of visual effects, and in terms of narrative expression, it leads the viewer to generate positive feedback on the video through the presentation of various elements.

Keywords: multimodal, visual grammar, social semiotic, multimodal interaction, visual communication.

1.   Introduction

From an analytical perspective, the presentation of images and videos is not isolated, and the coherence of information relies on the joint construction of images and text [1]. While the linguistic aspects of language and image have been relatively well-studied, the construction of images needs to be interpreted to understand the complementary relationship between the two. Additionally, the handling techniques of these videos can be analyzed using multimodal discourse analysis tools following the approach of Kress and Leeuwen [2]. They argue that understanding the image as a mode of representation, both in terms of its usage frequency and its role as a social semiotic resource, can help explore the rules of visual representation. This understanding can provide insights into how different filming techniques and visual syntax presentations can be utilized to create more effective and impact video campaigns like these. Therefore, this paper dynamically analyzes two representative videos about Tianshuiwei and Malatang on the Douban short video platform from multiple dimensions. One video is released by China's official media, CCTV News, and the other by an internet celebrity focusing on food. By analyzing and comparing these two videos, it is possible to uncover the communication principles underlying this phenomenon. Through this research, scholars can not only uncover the dissemination patterns and predict the future development of similar phenomena,

but also utilize these findings to implement effective promotional strategies, thereby enhancing the effectiveness of communication.

2. Theoretical background

Multimodal discourse analysis is an interdisciplinary research approach that explores how different semiotic systems collaborate in social communication to construct meaning. As one of the early researchers in this field, Barthes combined linguistics, semiotics, communication studies and visual studies to develop a rich theoretical framework and analytical tools [1].

Kress and van Leeuwen laid the foundation for multimodal discourse analysis by proposing a visual semiotic grammar that examines how images convey meaning through structure and design [2]. Halliday emphasized language as part of social practice, proposing three main functions of language: ideational, interpersonal, and textual [3]. Additionally, O'Toole applied the theory of systemic functional linguistics to analyze meaning-making in visual art [4]. Furthermore, Baldry and Thibault contributed a set of tools and methods for analyzing multimodal discourse, particularly studied for multimedia texts [5]. Van Leeuwen offered a comprehensive introduction to social semiotics, discussing the role and meaning construction of semiotic systems in society [6]. Jewitt compiled the latest research and methods in the field, covering a wide range of applications and theoretical issues [7]. Norris provided a methodological framework for analyzing multimodal interaction, focusing on how different modes collaborate in interaction [8]. Similarly, Forceville explored visual metaphors in advertising, analyzing how images convey complex metaphorical meanings [9]. Kress examined the social semiotic perspective of multimodal communication, analyzing multimodal features in modern communication [10]. Van Leeuwe analyzed the role of sound in virous modes, including speech, music, and other auditory symbols [11]. Lemke explored visual and linguistic semiotics in scientific texts, analyzing how multimodal elements jointly construct meaning [12]. Moreover, Iedema introduced the concept of resemiotization, exploring multimodal practices in discourse analysis [13]. Scollon and Scollon examined the role of language in the material world, analyzing how geographical locations and physical environments influence discourse [14]. Royce proposed a framework for intersemiotic complementarity, exploring how different modes complement each other in multimodal discourse analysis [15]. O'Halloran combined systemic functional linguistics with multimodal discourse analysis to study how language and visual imagery jointly construct ideational meaning [16]. Martinec and Salway proposed a system for image-text relations, examining how images and texts interact in both new and traditional media [17]. Thibault provided methods for multimodal transcription of television advertisements, combining theory and practice to analyze the multimodal features in advertisements [18]. Bateman proposed a systematic approach to analyzing multimodal documents, while Jewitt analyzed the multimodal approach in technology, literacy, and learning, exploring how technology changes and enriches multimodal communication [19-20].

Collectively, these works construct the theoretical and methodological system of multimodal discourse analysis, offering comprehensive resources and tools for understanding and analyzing multimodal texts and interactions. They reveal the complex interactions between different modes, demonstrating the importance and potential of multimodal discourse analysis in contemporary communication.

3. Analysis of video of CCTV

3.1. Data description

The official CCTV account on Douyin released a video on March 14, 2024, featuring Tianshui malatang, which has garnered 606,000 likes and 49,000 comments. The video, which is only 31

seconds long, is composed of the following segments in order of their appearance and duration, and the table 1 shows the distribution of participants and circumstances by time period:

a) The making of the malatang, including stirring and cooking the ingredients (duration: 2 seconds)

b) Interviews with passersby, asking where they are from (duration: 2 seconds)

c) An interview with a passerby in a car, commenting on the taste of the malatang (duration: 2 seconds)

d) Picking up food from the bowl with chopsticks (duration: 4 seconds)

e) A woman opening a car door and inviting tourists in (duration: 1.5 seconds)

f) Three local residents holding a red banner welcoming tourists (duration: 0.5 seconds)

g) Tourists pulling suitcases in a parking lot where cars have Chinese flags on their hoods (duration: 0.5 seconds)

h) A Hongqi car driving on the road (duration: 0.5 seconds)

i) Tourists eating malatang locally (duration: 5 seconds)

j) Local residents speaking into microphones, calling on locals to provide convenience for tourists (duration: 5 seconds)

k) The appearance of a local bus tour line (duration: 1 second)

l) Tourists and the camera saying goodbye to Tianshui (duration: 5 seconds)

Table 1: Distribution of participants and circumstance by time period

Time line	Participants	Circumstance
0:00-0:02	Sesame sauce, chili oil, spoon, chopsticks basin	Restaurant A big pot filled with water vapor
0:03-0:07	Young ladies baggage	Train station Car
0:08-0:10	Red soup Potatos Peanuts Bean-starch noodles	No clear circumstance
0:11-0:20	Car flag wall scroll Billboard lantern table Skewered vegetables	Car park road Food street
0:21-0:31	Bus Microphone Tent A lot of boxes of mineral water	Restaurant Bus stop Train station

3.2. Conceptual representation

When conducting multimodal analysis of videos from the perspective of visual grammar, two distinct approaches are commonly employed: conceptual representation and narrative representation. The conceptual representation in this video mainly lies in the socio-semiotic interactions among the

various components presented in the video. Analyzing the conceptual representation involves two layers: analytical structures and symbolic structures, which are important concepts proposed by Kress and van Leeuwen to understand and analyze how images construct meaning [2]. Analytical structures help viewers understand complex details and systems by demonstrating part-whole relationships, classificatory relationships, and quantitative relationships. Symbolic structures convey abstract concepts and values through symbolic attributes, symbolic processes, and cultural social symbols. From the perspective of analytical structures, the video includes various participants involved in interactions between people, which primarily occur in social activities. Participants related to food appear at the beginning of the video. The video uses a blurred approach, initially only showing the utensils and containers used to make malatang (Figure 1a). Later, the recurrence of food in subsequent segments makes the external representation clearer, allowing viewers to better observe the components of malatang (Figure 1b). Through the perspective of symbolic structures, this presentation of food can enhance sensory interaction between the audience and the screen. Notably, this method highlights that malatang in Gansu is an affordable and high-quality food. In the initial scenes, food is stirred in a stainless steel basin with a diameter of about 20 cm, creating a strong visual impact due to its large capacity. This is followed by a colander about 25 cm deep, filled with vegetables. In the recurring scenes, a large porcelain bowl appears, repeatedly suggesting to viewers that the food portions are substantial. The main color scheme of the food is red, which is visually striking and symbolizes passion, excitement, and energy (Figure 1c). In malatang, red represents spiciness and heat, attracting diners' attention and stimulating their appetite. This presentation is friendly to the broad working-class population in Chinese society, who prefer economical food due to various factors in recent years. Such visual representation aligns with this theme and is a specific method used by official media to promote tourism.

Figure 1: Food scene presentation (a)Blurring scene; (b)Semi-blurred scene; (c)Clear scene

Moreover, the video also reflects this theme in subsequent scenes. Two girls hold numerous vegetable skewers (Figure 2a), and although the participants are not prominently featured or structurally divided, viewers can easily capture this detail. The symbolic structure here highlights the freshness and affordability of the vegetables, indicating that consumers can get more for less. The participants' smiling faces suggest satisfaction with this consumption activity (Figure 2b).
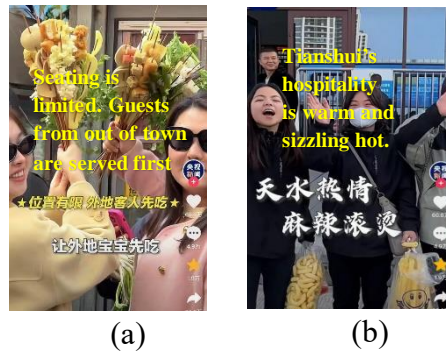
(a)   (b)

Figure 2: Character interaction (a)Interaction during visiting; (b)Interaction after visiting;

Additionally, much of the video's content revolves around transportation venues and tools. From an analytical structure perspective, the scenes are set in high-speed railway stations (Figure 3), closely linked to the rapid development of high-speed rail technology in China in recent years. The official video aims to showcase China's advanced infrastructure construction and suggests that adopting this mode of transportation can reduce travel time. Furthermore, the scene featuring a car club welcoming travelers reflects the locals' enthusiasm and the thoughtful service visitors can expect, thereby attracting more people to visit.

Figure 3: visitor from afar

3.3. Narrative representation

Narrative presentation is a crucial perspective in video analysis that shapes an individual's cognitive narratives through interactions between images, texts, and various participants. It reveals how images reproduce dynamic events and relationships to achieve intended communicative impacts. In this video, the direct representation of malatang uses a technique of initial and subsequent reappearance: the first instance is blurred while the latter is clear. Visually, this blurring technique enhances the viewer's imagination. Although specific details of food are indistinct, the red hue evokes spicy sensations, associated with steam conveying warmth and comfort, which appeals to Chinese preferences for hot dishes. During interviews with nationwide tourists, the camera adopts a side angle, aligned with the interviewees' eye level, creating an observer's perspective that is objective and detached. This angle reduces direct interaction, thus bridging the narrative distance and allowing the viewer to focus more on the content than the emotional dynamics. Interviewees' gaze vectors away from the camera toward specific directions or interviewers, making viewers feel like bystanders, thus fostering a natural and trustworthy dialogue atmosphere. Such settings, typical in official media releases, mitigate accusations of staging. From a vectorial viewpoint, these angles differ from the typical screen-centering approach, offering a more genuine interaction ambiance. Notably, this technique deviates

from formal settings used for hosts or dignitaries, highlighting a more approachable image. Overall, these narrative shifts from food to people and back enhance viewer retention from the initial blurred food scene, creating a lasting visual impact in viewers' minds.

4. Comparison

Under similar conditions, this research analyzes another video about Tianshui Malatang shot by a different Chinese internet celebrity. In presenting conceptual representation, both videos use an overlapping narrative technique that combines images of food with images of transportation facilities. However, in terms of filming techniques, there are significant differences between the two videos, and due to the different lengths of the videos, their content capacity also varies greatly. Regarding the internet celebrity's filming techniques, their video focuses more on presenting and depicting the details of the food. Specifically, the food details in the internet celebrity's video are more meticulously depicted. For example, the video not only shows the overall appearance of the Malatang but also uses close-up shots to display the texture and color of various ingredients. This filming technique allows the audience to intuitively feel the deliciousness and appeal of the food. The video might also include detailed steps of the cooking process, showcasing the handling and cooking techniques of the ingredients, enabling the audience to see the final product and understand the making process.

4.1. Conceptual representation

From the perspective of symbolic structure, the internet celebrity's video records their personal narrative perspective as a consumer, showing their experiences of taking a ride, queuing up, and interesting events happening during the journey in chronological order. This conveys a love for food and attention to life's details, drawing the audience closer to the video content and eliciting resonance and interest. Additionally, the video emphasizes interaction with the audience through humorous and witty commentary and engages with viewers' comments, enhancing audience participation and viewing experiences. While the internet celebrity's video might present transportation facilities more simply, these still serve to imply and emphasize the journey. The video features humorous interactions with passengers on transportation modes such as high-speed trains and local buses. The high-speed train environment appears bright and comfortable, and bus drivers are portrayed as friendly and welcoming, bridging the gap between local residents and tourists (Figure 4). In contrast, the official video does not show the crowded bus environment and focuses on individual car club members providing one-on-one transport, whereas the personal social media video presents a more relatable scenario that an ordinary person might encounter while traveling.

Figure 4: Interaction from an unofficial perspective

4.2. Narrative Representation

From the perspective of narrative representation, there are differences in how the two videos present the bustling business and long queues in terms of vector angles. The official social media video uses a high-angle camera perspective, capturing the scene evenly from the center outward, with participants distributed across various vectors. In contrast, the internet celebrity's videos employ a sequential linear tracking of the crowd with the camera, exhibiting a targeted feature. In other words, the official video might need to selectively present the bustling scene due to length constraints, aiming to convey more information within a limited capacity. On the other hand, the internet celebrity's video, with its more generous content length, allows the audience to immerse more fully in the on-site atmosphere through linear extension (Figure 5).

Figure 5: Scene of queue

5. Discussion & Limitation

5.1. Discussion

This paper analyzes two videos from the perspectives of conceptual presentation and narrative presentation to explore the underlying logic behind the sensational popularity of Tianshui malatang. From the conceptual presentation perspective, the analysis focuses on the visual representation of the food's color, the interaction between the food and the kitchen utensils, and the evocation of the audience's synesthesia. This makes the food visually appear to have three characteristics: high caloric content, large portions, and cost-effectiveness. In a sense, this is closely linked to the overall performance of social and economic activities, as people tend to favor such food. Correspondingly, to promote local economic development, the government combines local specialty cuisine with economic tourism. To highlight the achievements of the Chinese government in infrastructure construction in recent years, such videos often include content related to convenient local transportation facilities. From the perspective of narrative presentation, the feedback from travelers about the food is crucial in determining whether such videos can succeed. This presentation goes beyond the verbal expressions of the travelers, such as direct praise for the food. Instead, it guides the audience to form corresponding positive feedback stimuli in their minds through the interaction of various components. This intuitive presentation is more effective than verbal descriptions. This is a characteristic of multimodal theory in the analysis of social phenomena, as it provides a multidimensional analysis of such issues from various angles. The video content can be summarized by a formula, where N represents the entire video length, D(F) represents the duration of food-related content, D(C) represents the duration of footage showing people enjoying the food, and D(T) represents the duration of content related to local transportation facilities:

$$\text{Analytical Formula: } N=D(F)+D(C)+D(T)$$

5.2. Limitation

It must be said that this paper has some limitations. The analysis is constrained by word limitation, deconstruct the dynamic frames that continuously appear in videos, and then analyze the components of each frame vectorially. As a result, this paper only provides a general analysis of the presentation techniques used in video production. Such relatively vague results can only offer a general direction for future video creators, but more precise processing of videos is needed to uncover the deeper relationships between communication techniques and social phenomena.

6. Conclusion

This paper employs a multimodal analysis from the perspectives of Kress and Leeuwen's visual grammar and Halliday's social semiotics to examine two short videos about "Tianshui Malatang", (one of these videos is from official media while another is from an internet celebrity), each employing distinct conceptual representation and narrative representation to present a common theme. However, delving deeper into the reasons behind their explosive popularity reveals a close connection to contemporary social and economic activities in China. People are drawn to the affordable yet richly flavorful nature of such cuisine, and these videos adeptly showcase the convenience of going Tianshui. Furthermore, endorsements from official channels and celebrities amplify this phenomenon. This paper still has some limitations, without incorporating the relevant theories of communication studies into the interaction of visual text.A more detailed quantitative analysis of the participants presented in the visuals is needed to obtain more precise results,but it is hoped that it can provide a framework for future research into such phenomena, striving for more comprehensive insights.

References

[1] Barthes, R. (1999). Rhetoric of the Image. Visual culture: The reader, 33-40.
[2] Kress, G., & van Leeuwen, T. (1996). Reading Images: The Grammar of Visual Design. Routledge
[3] Halliday, M. A. K. (1978). Language as Social Semiotic: The Social Interpretation of Language and Meaning. Edward Arnold
[4] O'Toole, M. (1994). The Language of Displayed Art. Leicester University Press
[5] Baldry, A., & Thibault, P. J. (2006). Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook. Equinox
[6] van Leeuwen, T. (2005). Introducing Social Semiotics. Routledge
[7] Jewitt, C. (Ed.). (2009). The Routledge Handbook of Multimodal Analysis. Routledge
[8] Norris, S. (2004). Analyzing Multimodal Interaction: A Methodological Framework. Routledge
[9] Forceville, C. (1996). Pictorial Metaphor in Advertising. Routledge
[10] Kress, G. (2010). Multimodality: A Social Semiotic Approach to Contemporary Communication. Routledge
[11] van Leeuwen, T. (1999). Speech, Music, Sound. Macmillan International Higher Education
[12] Lemke, J. L. (1998). Multiplying Meaning: Visual and Verbal Semiotics in Scientific Text. In Reading Science (pp. 87-113). Routledge
[13] Iedema, R. (2003). Multimodality, Resemiotization: Extending the Analysis of Discourse as Multi-Semiotic Practice. Visual Communication, 2(1), 29-57
[14] Scollon, R., & Scollon, S. W. (2003). Discourses in Place: Language in the Material World. Routledge
[15] Royce, T. D. (2007). Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. In New Directions in the Analysis of Multimodal Discourse (pp. 63-109). Lawrence Erlbaum Associates
[16] O'Halloran, K. L. (2008). Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning Using Language and Visual Imagery. Visual Communication, 7(4), 443-475
[17] Martinec, R., & Salway, A. (2005). A System for Image-Text Relations in New (and Old) Media. Visual Communication, 4(3), 337-371
[18] Thibault, P. J. (2000). The Multimodal Transcription of a Television Advertisement: Theory and Practice. In Multimodality and Multimediality in the Distance Learning Age (pp. 311-385). John Benjamins
[19] Bateman, J. A. (2008). Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents. Palgrave Macmillan

[20] Jewitt, C. (2006). Technology, Literacy and Learning: A Multimodal Approach. Routledge