

Exploring Communication Strategies for Intangible Cultural Heritage(ICH) in Short Videos Through Multimodal Discourse Analysis: A Case Study of Jiang Xunqian's "Queshan Iron Flower" Video

Pengyuan Wang

School of English and International Studies, Beijing Language and Culture University, Beijing, China

202211680362@stu.blcu.edu.cn

Abstract. In the new media era, short videos have gradually become an important medium for cultural communication due to their visual appeal and interactivity, presenting both opportunities and challenges for the inheritance and innovation of intangible cultural heritage (ICH). In recent years, some scholars have explored ICH-related short videos from perspectives such as symbolic interaction and narrative techniques, but few have explored the construction methods of different modalities within them. This study focuses on the short video "Queshan Iron Flower" produced by Jiang Xunqian on "Bilibili". Applying multimodal discourse analysis, the study combines Zhang Delu's analytical framework with Feng Dezheng's transcultural multimodal theory to systematically explore how ICH is communicated through short videos. The analysis unfolds across four key dimensions: culture, context, content, and expression, ultimately providing suggestions for enhancing ICH communication strategies in the short videos.

Keywords: multimodal discourse analysis, Jiang Xunqian, intangible cultural heritage, short video

1. Introduction

Intangible cultural heritage (ICH) represents the crystallization of wisdom accumulated throughout Chinese history. In August 2022, the 14th Five-Year Plan for Cultural Development has already emphasized the importance of ICH preservation, advocating innovative thematic campaigns and promoting its integration into schools, communities, and social media. Yet, in today's digital era, ICH communication still faces the challenges. How to make traditional crafts resonate with younger audiences has become pivotal to cultural preservation [1].

In January 2024, self-media blogger Jiang Xunqian's video "Queshan Iron Flower" gained widespread attention across major social media. As of April 19, 2025, the video had garnered over 27.5 million views on "Bilibili", while being reposted by official media outlets including People's Daily and CCTV News. This phenomenon serves as an exemplary model for the innovative

presentation of traditional culture in contemporary society. Given its representative significance, this study selects the video as its research subject. By applying Zhang Delu's framework of multimodal discourse analysis and Feng Dezheng's transcultural multimodal construction, the research aims to: examine how cultural meaning is constructed through multiple modalities in short videos; and analyze how different modal elements work synergistically to enhance ICH communication. The findings are expected to provide both theoretical insights and practical guidance for the modern communication of intangible cultural heritage.

2. Literature review

Scholars have explored ICH short videos from various theoretical perspectives. Bi Qiuling first linked ICH with short videos, emphasizing format compatibility [2]. Liu Ruoxi and Qiu Haoran examined ideological themes in ICH communication on these platforms [3]. Wang Ziyun analyzed Li Ziqi's videos using multimodal discourse analysis, highlighting synergies between language, visuals, and music [4]. In the research of Jiang Xunqian's ICH short videos, Hu Yinuo explored symbolic interactions among creators, ICH, and audiences [5], while He Yanxinyue stressed improving ICH-video format alignment through narrative techniques [6].

However, existing research lacks systematic multimodal analysis of newer creators like Jiang Xunqian and often rigidly applies frameworks like Zhang Delu's without adaptation. This study bridges the gap by integrating Zhang Delu's multimodal framework with Feng Dezheng's transcultural communication theory to analyze Jiang's "Queshan Iron Flower" video. It reveals ICH's unique social media communication features and suggests innovations for digital-era expression.

3. Methodology

The systemic functional grammar established by British linguist M.A.K. Halliday in the late 1950s comprises two closely interrelated components: systemic grammar and functional grammar [7]. Systemic grammar primarily focuses on analyzing the internal structural relationships within language as a complex system. Functional grammar, on the other hand, aims to explain the three fundamental metafunctions of language: the ideational function, the interpersonal function, and the textual function, emphasizing that language serves as a means of social interaction [7]. In contemporary multimodal discourse studies, Zhang Delu's framework emphasizes the semiotic synergy in a specific social and cultural context, providing a systematic approach for multimodal discourse analysis [8]. Feng Dezheng's transcultural framework focuses on global communication, regarding culture as semantic features with multimodal representations, particularly in new media [9]. Integrating these frameworks provides a tailored approach for analyzing Jiang Xunqian's ICH short videos, offering both theoretical foundation and analytical tools for a deeper understanding of ICH's modern communication.

4. Results

4.1. Culture

The cultural level refers to the social, historical, cultural, and ideological foundations that shape the production and interpretation of discourse, serving as a crucial background in multimodal discourse analysis [8]. Jiang Xunqian's "Queshan Iron Flower" video adopts a retrospective perspective to trace the history of this intangible cultural heritage, while simultaneously employing a forward-

looking perspective to facilitate public communication on social media, functioning both as a historical narrative and as a dialogue with the social public. Thus, the cultural context of this video can be analyzed along two key dimensions: historical dimension and social dimension. In the historical dimension, the video focuses on Queshan Iron Flower, using the historical marker “millennium-old artistry” in its title to highlight its deep historical accumulation. Meanwhile, it underscores the lineage of transmission by depicting the generational chain of “inheritor → master → grandmaster”. In the social dimension, the video transforms this traditionally perilous art into a symbol of “ultimate romance,” effectively dissolving the public’s sense of detachment from ancient art and instead stimulating their enthusiasm for rediscovering traditional culture, fostering a deeper sense of cultural identity in the society.

4.2. Context

The contextual level of discourse consists of three key components: the field, tenor, and mode of discourse [8]. First, the field of discourse addresses “what is being said” [10]. Jiang Xunqian’s video focuses on the cultural significance of the art form, particularly its historical roots as a folk ritual symbolizing good fortune and happiness. Second, the tenor of discourse refers to “who is speaking to whom,” closely related to the identities of both the blogger and the audience [10]. Quantitative analysis reveals Jiang’s dual social media identities: an ICH learner and equal-status communicator (Table 1). Meanwhile, its target audience mainly consists of Gen Z users who has a strong demands for interaction and personalized aesthetic preferences [6]. Based on these identities, the video adopts a natural and friendly tone to narrow the psychological distance with the audience. Finally, the mode of discourse addresses the issue of “how the message is conveyed” [10]. On platforms like “Bilibili”, the unique “bullet comment” culture fosters real-time interaction. Based on this feature, Jiang Xunqian uses inviting words phrase like “I’m going to take you to see...” to stimulate the audience’s enthusiasm for participation.

Table 1. The number of discourses reflecting identity characteristics

Identity	Number of discourses
Learner	11
Interlocutor	3

4.3. Content

The content level mainly focuses on how meaning is organized and expressed through form [8]. Form refers to the physical manifestation of symbols, including language, images, and sounds, while meaning is the semiotic connotation carried by form.

4.3.1. Language

Language, as the core element of discourse analysis, mainly focuses on the selection of vocabulary and the organization of grammar [8]. In Jiang Xunqian’s video, the use of language not only conveys information but also reinforces cultural expression through specific vocabulary and sentence structure, thereby strengthening audience engagement through both emotional resonance and cognitive experience.

At the lexical level, the video’s narration incorporates abundant “attitudinal words” and “novel metaphors” [9]. The main function of “attitudinal words” is to directly shape cultural images

through words with clear emotional and value judgments. Novel metaphor refers to expressing specific meanings through non-conventional and creative metaphors [9].

Example 1: “Millennium-old masterpiece”

The term “millennium-old masterpiece” belongs to the category of attitudinal words. The term “millennium” directly shows the long history of the iron flower, emphasizing its profound cultural heritage. The “masterpiece” highlights the technique's extraordinary skill level and implies its endangered status, thereby inspiring the audience's attention to the preservation of such intangible cultural heritage.

Example 2: “The ultimate romance”

This expression serves as a novel metaphor. Though the conventional meaning of “romance” seems contradictory to the inherent danger of iron flower, it can express the physical beauty of the spectacle, the spirit of bravery, and its cultural symbolism as peace and blessing.

At the grammatical level, the video uses a large amount of narrative discourse to depict behaviors. However, it is worth noting that narrative discourse provides the audience with a cognitive basis by “stating facts”, and often appears with a series of conjunctions and modal verbs, forming a narrative rhythm of “fact-emotion”.

Example 3: “Hold molten iron with the upper rod and strike with the lower rod.”

This declarative sentence adopts a parallel structure to describe the technical key points of the iron flower without subjective commentary or emotional judgement. When followed by adversative conjunctions like “however” or “in reality”, this simple description creates a dramatic contrast that strongly highlights the potential difficulty of the technique. Meanwhile, modal verbs further reinforce technical precision, as in “you must achieve a fine drizzling effect”. Under the guidance of the grammar, viewers undergo a psychological journey from cognitive understanding to value recognition and finally emotional resonance, thus forming an emotional connection with Jiang Xunqian.

4.3.2. Image

The visual symbols in Jiang Xunqian's “Iron Flower” video are composed of two key elements: human elements and environmental elements. Human elements mainly convey specific meanings through facial expressions and body languages. The environmental symbol system includes foregrounded symbols and background symbols. The work in concert to create a clear narrative structure that constructs a comprehensive cultural imagery space.

Example 4:

When ICH inheritor Yang Jianjun hears Jiang Xunqian's request to learn the iron-flower forging technique, his instantly visible look of astonishment forms a set of highly expressive human symbols. This symbolic representation not only directly communicates the inherent danger of iron-flower, but also implies the strict standards for the inheritance and the difficulty of learning it. The narrative tension builds precisely through this contrast, enhancing the video's emotional resonance and successfully engaging audience, which promote the communication of the cultural heritage.

Example 5:

The willow rods serve as key props throughout the video. By showing the changes in the form of willow sticks at different training stages, the symbol not only strengthen the theme of the video but also promote the natural development of the plot. The background scene symbols such as sparrows and eaves mainly serve the function of creating an artistic conception, enhancing the cultural atmosphere and enabling viewers to feel the original cultural context more intuitively.

4.3.3. Sound

The auditory symbols play an equally vital role in the video, working with visual symbols to create a highly immersive scene that strengthens viewer engagement. In the video, the auditory symbol system primarily consists of three key elements: background music, synchronized live sounds, and voice-over narration [11].

(1) Background music

Background music refers to non-verbal sound elements layered over the primary discourse. The video features three instrumental tracks: “Jiuzhoutong,” “Grace,” and “The History.” “Jiuzhoutong” combines ethnic symphonic and electronic elements, reinforcing themes of cultural unity. The other two tracks blend Chinese and Western instruments, creating a majestic, epic tone. Though stylistically distinct, all three enhance the video’s cultural depth.

(2) Synchronized live sounds

The synchronized live sounds refer to audio recorded on-site during filming, encompassing character dialogues, actions, and environmental noises [11]. The video preserves authentic on-site audio, including conversations between Jiang Xunqian and inheritor Yang Jianjun, as well as ambient sounds like the muffled strikes of willow rods. These synchronized sounds enhance documentary realism, amplify narrative tension, and deepen emotional engagement in the cultural portrayal.

(3) Dubbing

Dubbing narration refers to audio recorded post-production in a studio, including character dialogues, commentary, and voice-over narration [11]. In the video, Jiang Xunqian’s first-person narration dominates the dubbing. This narration provides key details (time, location, characters) while enhancing both factual and emotional dimensions of the visuals. Serving as a narrative bridge, it transforms fragmented footage into a cohesive storyline, improving video coherence and plot progression.

4.4. Expression

The expression level focuses on how media technologies facilitate the presentation of semiotic meaning—the mechanisms through which symbolic meaning is conveyed rather than the meaning itself. This dimension holds significant value when analyzing communication strategies.

4.4.1. Language

Regarding language expression, this level differs from the lexical and grammatical concerns of the content level, concentrating instead on how textual meaning is materially realized through accompanying media. In “Queshan Iron Flower” video, accompanying language plays an supplementary role in semantic transmission, mainly including accent, tone, font shape, typography (including font style and size), etc [8].

Example 6:

The subtitles in Jiang Xunqian’s video employ a white font centered at the bottom of the screen. The positioning not only conforms to the visual reading habits of the audience, but also avoids blocking the core content of the picture. The white typography establishes harmonious visual rapport with the video’s warm color, which prevents both the harsh contrast of dark subtitles and the legibility issues common with pale-colored text. Meanwhile, the choice of white color and simple

fonts create an optimal balance between readability and visual hierarchy. This way, they do not overly attract the audience's attention but provide necessary language supplements when needed.

4.4.2. Non-linguistic expression

At the expression level, non-linguistic elements encompass both visual imagery and sound [8], corresponding respectively to the visual and auditory semiotic representations at the content level. These components place particular emphasis on their modes of presentation, playing an even more crucial role in the field of communication.

(1) Image

In terms of visual systems, Kress and van Leeuwen extended Halliday's systemic functional grammar to the analysis of visual modalities, proposing three core meanings of visual grammar: representational meaning, interactive meaning, and compositional meaning [12]. In "Queshan Iron Flower" video, the representational meaning, interactive meaning proves particularly valuable for analyzing how creators achieve the visual construction of culture through image design.

① Interactive Meaning

Interactive meaning refers to the communicative relationship established between participants in an image and its viewers, manifested through four key dimensions: contact, social distance, perspective, and modality [13].

1) Contact

Contact in visual analysis examines whether image participants establish eye contact with viewers. Kress and van Leeuwen categorize this into demand images (direct gaze, creating engagement) and supply images (no eye contact, emphasizing objectivity) [11,14]. In Jiang Xunqian's "Queshan Iron Flower" video, supply images dominate by using third-person perspectives to maintain a professional.

Example 7:

The video deliberately avoids performers' eye contact with the camera, maintaining their complete focus on the striking movements. This reflects Jiang Xunqian's "learner's perspective" approach, which replaces traditional documentary's expert-audience hierarchy with an egalitarian dynamic. By positioning both narrator and viewers as ICH learners, the video eliminates the need for authoritative transmission or emotional connection through direct gaze. This method encourages audiences to subconsciously adopt the learner role, fostering deeper immersion in the cultural practice itself.

2) Social distance

Social distance refers to the psychological distance between the participants and the viewers in an image [12]. This sense of distance directly affects the closeness between the two sides and is mainly achieved through the shot types in visual representation. Specifically, camera distances can be categorized into three distinct levels: close shots convey intimate distance, creating a sense of personal connection. Medium shots reflect social distance, simulating everyday interpersonal interactions. Long shots establish public distance, maintaining formal observation space.

Example 8:

In the video, the close-ups of Jiang Xunqian striking sand and water focus on arm movement details, shortening psychological distance and transforming viewers into active participants. Medium shots during Jiang and Yang Jianjun's conversation create an "on-site observer" perspective, framing their interaction while maintaining realistic spatial distance. For the iron flower performance, long shots capture the magnificent night spectacle, with slow-motion emphasizing

technical details while situating the ritual in its grand, sacred context. Together, these techniques create an immersive “co-learning” experience that makes ICH transmission tangible.

3) Attitude

The elements of attitude are mainly constructed through the shooting perspectives. Perspective, as an important dimension of visual grammar, reflects the spatial positional relationship between the viewer and the participants in the image as well as the implicit power structure [11]. In specific application, shooting perspectives can be divided into three basic types: eye-level shots, high-angle shots, and low-angle shots, each of which can convey specific attitude tendencies.

Example 9:

The low-angle shot of blooming iron flowers creates a powerful visual hierarchy. Jiang Xunqian's upward gaze forms a guiding visual chain that: 1) highlights the iron flowers' central position; 2) imbues ICH with symbolic “sublimity”; and 3) visually conveys reverence for traditional culture. This cinematography naturally elevates viewers' cultural appreciation through visual experience.

4) Modality

Modality refers to the construction of an image's authenticity and emotional tone through visual elements like color and lighting [15]. High modality is usually achieved through highly saturated colors, indicating a relatively high degree of authenticity and credibility of the image content. The medium mood usually tends to the middle value in terms of image color, and the emotions it expresses are also relatively conservative and neutral. Low-mood tones often employ achromatic colors, conveying an emotional atmosphere of detachment or nostalgia. Most of the visual symbols in Jiang Xunqian's videos are of high modality.

Example 10:

The video mainly uses highly saturated orange and gold as the main tones, creating a strong visual impact. The golden iron flowers serve as the visual focus, complementing the red elements of the performers' costumes and jointly creating a brilliant and splendid scene. At the level of cultural symbols, gold has always symbolized wealth, good fortune and blessings, which is highly consistent with the original meaning of the iron flower: “driving away evil and bringing good fortune”, allowing the audience to naturally understand the spiritual connotation behind intangible cultural heritage in the aesthetic experience.

② Representational Meaning

The representational meaning mainly explores how images reproduce reality through visual elements (participants, environment, vectors). The participants in visual elements usually refer to the characters, objects, etc. that constitute the picture and participate in its reproduction. The environment is the spatial scene. Vector refers to the directional dynamic lines constructed through visual elements in an image [16]. It is the key criterion for distinguishing narrative representation from conceptual representation. When the representation behavior of the subject in an image is linked through vectors, it belongs to narrative representation. Conceptual representation does not have vectors and is mainly manifested in the static symbolic process.

1) Narrative Representation

Example 11:

There are obvious vectors in the video, constituting the action process. During the action process, the participant who initiates the action is the mover, and the participant pointed to by the vector is the target. In the figure, Jiang Xunqian strikes the upper rode with the lower one, propelling the sand towards the sky and forming an upward vector, which constitutes the action process.

2) Conceptual Representation

Example 12:

The willow sticks and sand, as the core participants, are presented independently in the form of close-up shots. There are no directional dynamic lines or obvious vectors in the picture, allowing the viewer's attention to be completely focused on the attributes of the participants themselves.

3) The collaborative mechanism for reproducing meaning

In videos, narrative representation and concept representation often do not occur independently. Instead, they are connected into a complete “plot” by superimposing and intersecting single dynamic and static visual images, thereby achieving a synergy mechanism of meaning representation.

Example 13:

By connecting the narrative representation and conceptual representation mentioned above, a complete plot of Jiang Xunqian's practice of sand-fighting can be formed, and their relationship can be illustrated by the following figure.

(2) Sound

Based on van Leeuwen's theoretical framework of auditory grammar, this study reconstructs a three-module auditory grammar framework mainly consisting of rhythm and melody, which is used to analyze the sound symbols in Jiang Xunqian's videos.

① Rhythm

Rhythm refers to the temporal arrangement of acoustic events, where structured patterns create distinct communicative effects. These arrangements can be categorized as countable or uncountable [16]. In Jiang Xunqian's videos, the two arrangement rules form variations, which, in coordination with image symbols and language symbols, influence the perception of the viewers.

Example 14: Video 02:08-02:20 [Sound of willow rods]

This part of the video demonstrates the rhythmic changes from countable to uncountable arrangements. At the beginning, the sound of the willow rods being struck constitutes the basic rhythm unit, forming a stable rhythm. Gradually, the time intervals shorten, overlapping sequences that defy counting. This change in rhythm echoes the narration “practice makes perfect”. The “countable → uncountable” transition constructs a “novice→proficiency→mastery” cognitive model. This sound strategy enables the audience not only to “see” the skills but also to “hear” it, thereby achieving a deep understanding of ICH inheritance at the multi-sensory level.

② Melody

The melody of sound mainly includes elements such as range, pitch and tone, which are mainly reflected in the background music. In Jiang Xunqian's video, the melody of the background music collaborates with visual and other auditory symbols to create multimodal synergy, shaping the cultural expression of this intangible heritage craft.

Example 15: Video 04:05-07:42 [Background music “The History”]

The background music “The History” mainly presents the structure of “beginning - development - transition - conclusion”. The music begins at 04:05 in the video. The melody of the music at this time is tranquil. Then the music transitions to the stage of “development”. The melody gradually intensifies, becoming the dominant auditory element. The stage of “transition” occurs precisely at 05:45—the moment of iron-flower explosions. Post 06:48, the “conclusion” phase emerges as the melody softens, accompanying visual symbol of “iron flowers falling to the ground” to achieve the complete structure of melody.

5. Conclusion

The study reveals that Jiang Xunqian's “Queshan Iron Flower” video adopts a “learner's perspective,” successfully transcending the authoritative communication model of intangible cultural heritage. This approach achieves multidimensional innovation in cultural expression: at the language

level, it maintains professional accuracy in depicting the craft while enhancing accessibility through strategic lexical and grammatical choices. At the visual level, the combination of narrative representation and conceptual representation profoundly interprets the cultural connotation. At the auditory level, the coordinated interplay of timbre, rhythm, and melody constructs a richly layered soundscape, providing viewers with a multi-sensory immersive experience. Based on the findings, this paper puts forward the following suggestions: firstly, the creation of short videos for intangible cultural heritage should enhance the synergy among different symbols. Secondly, communicators need to balance professionalism and popularity, and innovate the expression methods on the basis of ensuring the authenticity of the culture. Finally, the sensory experience of the viewers through multi-dimensional stimulation should be emphasized. Although this study has achieved certain results, there are still limitations: firstly, the limited sample size restricts the universality of the conclusion. Secondly, there is a lack of empirical analysis of audience feedback data. These limitations, however, point to productive directions for future research. Subsequent studies could expand the sample scope by building a comprehensive corpus of ICH short videos for comparative analysis, further advancing the innovative communication of intangible cultural heritage in the digital age.

References

- [1] Chen, S. Q. (2025). Research on strategies for new media to promote the dissemination of intangible cultural heritage. *Masterpieces Review*, (8), 20-22. doi: 10.20273/j.cnki.mzxs.2025.08.022.
- [2] Bi, Q. L. (2018). The dissemination of intangible cultural heritage based on short video platforms. *Cultural Industry*, (14), 46-48.
- [3] Liu, R. X., & Qiu, H. R. (2025). Ideological presentations in official promotion of intangible cultural heritage (ICH) on short-form video platforms: A multimodal content analysis. *Information Research-An International Electronic Journal*, (30), 413-424.
- [4] Wang, Z. Y. (2025). A multimodal discourse analysis of ICH short videos in international communication: A case study of Li Ziqi's videos. *Media Review*, (4), 43-45.
- [5] Hu, Y. N. (2024). Research on the communication strategies of Chinese traditional culture on Douyin from the perspective of symbolic interaction (Master's thesis, Capital University of Physical Education and Sports). <https://link-cnki-net-s.webvpn.blcu.edu.cn/doi/10.27340/d.cnki.gstxy.2024.000017>.
- [6] He, Y. X. Y. (2024). A narrative study of ICH short videos on Douyin: A case study of blogger Jiang Xunqian. *New Media Research*, 10(8), 74-77.
- [7] Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- [8] Zhang, D. L. (2009). Towards a comprehensive theoretical framework for multimodal discourse analysis. *Foreign Languages in China*, 6(1), 24-30.
- [9] Feng, D. Z., & Li, X. (2025). Multimodal construction and dissemination of transcultural Chinese discourse. *Foreign Language World*, (1), 21-28.
- [10] Zhang, D. L. (2023). The role of cultural context in multimodal discourse construction. *Journal of PLA University of Foreign Languages*, 46(5), 54-61+161.
- [11] Zhang, J. G., & Liu, X. (2024). A multimodal discourse analysis of the web micro-drama "Escape from the British Museum". *Media Forum*, 7(21), 65-69.
- [12] Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Routledge.
- [13] Li, L. H., & Niu, Y. Q. (2024). A multimodal discourse analysis of Chengdu city image promotional videos from the perspective of visual grammar. *Journal of Qiqihar University (Philosophy & Social Science Edition)*, (11), 115-119.
- [14] Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- [15] Yuan, X. L., & Nai, R. H. (2022). A study on the multimodal meaning construction of "Cultural China" in international communication. *Foreign Language Education*, 43(5), 23-29.
- [16] Zhang, Y. Y., & Chen, X. (2025). Research on the international dissemination of documentaries from a multimodal perspective: A case study of *Amazing China: Nature—Zhangjiajie* series. *Foreign Languages and Translation*, 32(1), 56-62.