

The Mechanism of Metaphor Generation and Meaning Construction through Multimodal Collaboration: A Case Study of Folk Media Li Ziqi in Cross-Cultural Communication

Hui Wang

Department of English, School of Foreign Languages, Beijing Forestry University, Beijing, 100083, China

Keywords: Cross-cultural communication; Folk media; Li Ziqi; Multimodal metaphor; Modal collaboration

Abstract: In the contemporary era of deep integration between digitization and globalization, short videos have become a primary medium for cross-cultural communication. Through the multimodal weaving of audiovisual signs, they can trigger cognitive resonance and emotional consensus at multiple levels. This study employs multimodal metaphor theory from multimodal discourse analysis, and uses Li Ziqi's five most-viewed videos on YouTube as research objects. Combining multimodal annotation using ELAN with audience commentary analysis, it systematically examines folk media's metaphor generation mechanism through multimodal collaboration and its cultural dissemination effects from the perspective of cross-cultural communication. The study finds that this folk media's metaphor generation through multimodal collaboration follows a three-layer progressive mechanism: visual-linguistic collaboration accomplishes semantic anchoring of surface-level cultural symbols; visual-auditory collaboration triggers audience's embodied cognition at the level of perceiving cultural symbols; visual-editing and auditory-editing collaborations internalize deep cultural metaphors within the audience's emotional experience. The "cognition-perception-emotion" analytical framework constructed in this study clarifies the specific mechanism by which folk media dissects deep cultural metaphors into concrete and perceptible audiovisual expressions. It provides a refined analytical tool for understanding the "lifeworld translation" of Chinese culture in the context of globalization, and also provides modal configuration ideas and content balancing strategies that can be used as a reference for other video creators committed to cross-cultural expression.

1. Introduction

With the growth of globalization and digital media, video has become an important medium for cross-cultural communication. Through multimodal symbols, it shapes how the audience understand and emotionally respond to cultural content. Within multimodal discourse analysis, multimodal metaphor theory focuses on how different semiotic modes (e.g. visual, auditory and linguistic)

function as source domains that construct metaphorical meaning as the target domain, thereby helping the audience understand and relate to cultural meaning.

Although research on multimodal metaphors in cross-cultural video communication is growing, several issues remain underexplored. First, previous studies have primarily focused on institutional or officially produced media content, such as state-sponsored tourism promotional films and documentaries, while offering relatively limited attention of content created by folk media creators, which makes up a large proportion of content on platforms like YouTube. Second, most studies analyze visual, linguistic, and auditory modes separately, examining their independent functions, while paying insufficient attention to how these modes work together to produce metaphorical meaning.

Li Ziqi, a Chinese self-media creator with considerable international influence on YouTube, employs an idyllic, pastoral narrative style to present elements of traditional Chinese culture through everyday visual and auditory scenes. Her videos have attracted emotional responses and cultural interest among overseas audiences. This study selects the five highest-viewed Li Ziqi videos on YouTube as its corpus. Through multimodal annotation and audience comment analysis, it investigates two central questions: (1) How do visual, linguistic, auditory, and editing modes work together to construct cultural metaphors? (2) How do overseas audiences interpret these metaphors? The study also proposes an analytical framework focusing on cognition, perception, and emotion in cross-cultural video communication and to provide some implications for cross-cultural video production.

2. Literature Review

2.1 The Development of Multimodal Metaphor Theory

Multimodal discourse analysis is based on the idea that meaning is created through multiple semiotic resources [1]. Building on systemic functional grammar, Kress and van Leeuwen extended this approach to image analysis, proposing a social semiotic framework grounded in “visual grammar” with three metafunctions—representation, interaction, and composition. This work contributed to the development of multimodal discourse analysis [2]. Chinese scholar Zhang Delu further defined multimodal discourse as “forms of communication involving multiple sensory channels (auditory, visual, tactile, and so on.) and various semiotic resources (language, images, sound, action, etc.)” [3].

Influenced by developments in social semiotics, scholars in cognitive linguistics combined insights from metaphor theory, metonymy, and relevance theory to argue that metaphors can also be expressed through non-linguistic modes. They proposed the concept of multimodal metaphor: a metaphorical phenomenon in which the source and target domains are presented (wholly or primarily) through two different semiotic modes [4].

In China, studies on multimodal metaphor have mainly centered on theoretical reviews and case studies. Previous reviews have discussed the origins, development [5], and characteristics [6] of multimodal metaphor research. Case studies have covered political cartoons [7], automobile advertisements [8], documentaries, posters, self-media content, and other genres. However, these studies largely treat modes as independent analytical units and pay limited attention to how different modes work together to generate metaphorical meaning.

2.2 Research on Multimodal Metaphors in Video Communication from a Cross-Cultural Perspective

With significant development in globalization, multimodal metaphor theory has increasingly

been used in cross-cultural communication studies. Some research has noted that multimodal metaphors, through shared embodied experiences, facilitate cross-cultural understanding and help people from different cultural backgrounds understand each other [9].

One study examined multimodal metaphorical representations and interactive relationships in the *Publicity Film for Tourism in Chongqing*, offering a fresh perspective and new pattern for applying metaphor theory to video research in cross-cultural contexts [10]. Another proposed an analytical framework—from textual metaphor identification to multimodal translation—using YouTube English-subtitled short videos from the series *China in the Classics: The Book of Documents* [11]. Yet, most of these studies focus primarily on officially produced media content, with limited attention to folk media.

2.3 Research on Multimodal Metaphors in Li Ziqi's Videos from a Cross-Cultural Perspective

Li Ziqi's videos on YouTube have accumulated over 3.42 billion views. Even after a three-year hiatus from uploading, they continue to generate high engagement, suggesting that her videos continue to attract international audience.

Existing studies have analyzed Li Ziqi's videos from perspectives such as semiotics, narratology, and audience reception. Some scholars, drawing on Roland Barthes' semiotic theory, have distinguished visual signs in her videos into layers of “explicit (informative—denotation) — cultural-public (connotation) — significance of the role” (connotation) [12]. Audience studies have discussed audience responses to the interaction of visual and auditory elements within multimodal narratives [13]. Others have explored the interactions among visual elements, sound segments, linguistic text, and other modes [14]. Nevertheless, most studies still analyze different modes separately. Deeper inquiry is still needed into how modes interact with one another to “collaborate” in generating metaphorical meaning.

Overall, previous studies have mainly focused on official media and separate analyses of individual modes. Therefore, this study takes Li Ziqi's videos as its sample to examine how visual, linguistic, auditory and editing modes work together to express meanings related to traditional Chinese culture.

3. Methodology

3.1 Research Sample

This study adopts purposive sampling to select five of Li Ziqi's highest-viewed videos on YouTube. These videos represent several common multimodal features in her videos and include: *Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival* (130 million views), *A Pot of Wine Among Watermelon Fields. So...The life of watermelons and grapes?* (92.24 million views), *Liuzhou “Luosifen”* (89.98 million views), *It's a red mountain, and in the fall, it's natural to make some sweet persimmons* (74.56 million views), and *The Life of Purple Rice, Pumpkins, and ... Peanuts* (67.61 million views).

3.2 Research Design

The study employs the multimodal annotation tool ELAN to annotate the videos, which were coded according to four main modes: visual, linguistic, auditory, and editing. Audience comments with high engagement or representative viewpoints were selected to analyze audience interpretation and response. This approach allows the study to integrate multimodal metaphor generation analysis and audience feedback research for mutual corroboration.

4. Results and Discussion

Based on the coding results, this study categorizes multimodal collaboration in Li Ziqi's videos into three types: visual-linguistic collaboration, visual-auditory collaboration, and visual-editing or auditory-editing collaboration. These three forms of collaboration are interconnected and contribute to meaning construction in different ways. The first-level "visual-linguistic collaboration" lays down a framework for the cross-cultural audience to move from "unfamiliar" to "recognizable" and even "comprehensible". The second-level "visual-auditory collaboration" retains cultural subjectivity while triggering the audience's embodied cognitive experience, transforming symbols from other cultural backgrounds into perceivable bodily experiences. The third-level "visual-editing or auditory-editing collaboration" builds upon the cognitive and perceptual foundations of the first two levels to encourage emotional identification with the cultural metaphor presented in the videos. Analysis of audience comments also suggests that these multimodal strategies are effective.

4.1 Visual-Linguistic Collaboration: Semantic Anchoring of Surface-Level Cultural Symbols

4.1.1 Visuals and Titles: Initial Construction of Cross-Cultural Cognitive Frameworks

All sample videos feature bilingual Chinese-English titles that help explain the cultural meaning of the videos. For example, the English title *Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival* conveys the core cultural concept of "New Year snacks" and corresponds with the food preparation scenes shown in the video. This enables the overseas audience to understand the cultural significance behind these activities—namely, creating a festive atmosphere and reinforcing a sense of occasion. As one commenter (@kimisoa5794) noted:

Can we appreciate how she writes her title, name, and description in english and how she put english captions. She didn't need to do so but she went above and beyond.

The comment indicates that English titles make the videos easier for the overseas audience to understand in different cultural contexts, allowing the audience to develop an initial understanding of these cultural symbols.

4.1.2 Visuals and Subtitles: Knowledge-Oriented Explanation and Meaning Articulation

The interaction between the visual mode and subtitles lies in providing additional explanation and contextual meaning. In key narrative moments, the sample videos include Chinese subtitles, complemented by YouTube's auto-generated English subtitles. This bilingual approach offers ongoing explanatory support for cross-cultural the audience. Subtitles appear in scenes marking the transitions of the twenty-four solar terms, naming farming and cooking procedures, and titling finished products. In *A Pot of Wine Among Watermelon Fields. So...The life of watermelons and grapes?*, for instance, subtitles for solar terms such as "February 3,4 or 5" and "February 2020" alternate with shots of misty mountains and growing crops in the fields, together emphasizing the seasonal passage of time. In food-preparation videos such as *Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival*, subtitles at critical steps of each process supply step-by-step guidance while helping the audience better understand and remember the process of unfamiliar culinary procedures. The final titling of completed dishes provides semantic closure, integrating scattered procedural segments into a recognizable image of Chinese food culture.

4.2 Visual-Auditory Collaboration: Constructing the Audience's Embodied Cognition

4.2.1 Visual Mode and Dialect: Preservation of Cultural Subjectivity and Triggering of Bodily Sensations

The retention of dialect in Li Ziqi's videos first reflects a conscious stance of cultural subjectivity. In global communication practices, the creator preserves this locally distinctive semiotic resource, allowing it to complement the platform-generated English subtitles—dialect creates cultural atmosphere, while subtitles convey semantic information.

The dialect forms symbolic linkages with visual elements that evoke everyday life and pastoral imagery. In a segment from *The Life of Purple Rice, Pumpkins, and ... Peanuts*, the visual mode uses close-up shots to anchor a “labor” scene, while the auditory mode features brief dialogue in Sichuan dialect. The rises and falls in the dialect's intonation and rhythm create non-semantic resonance with the visual pacing. Although the audience cannot directly access the semantic content of the dialect, they can bodily sense the presence of a different culture through its tone, rhythm, and prosody. As one commenter (@mineswaripatwari2004) wrote:

What is she cooking- No idea...

What is she saying- No idea...

What are they eating- no idea...

But, hey, I always watch her video. Its gives me peace!!

This comment demonstrates that the lack of semantic comprehension does not prevent the audience from “feeling” and “appreciating” the different culture at a non-semantic level. The resulting sense of “peace” emotionally compensates for the semantic gap and helps achieve the goal of affective dissemination in cross-cultural communication.

4.2.2 Visual Mode and Ambient Sound: Triggering Synesthetic Metaphors

Another key sub-dimension of visual-auditory collaboration is the interaction between the visual mode and ambient sound, which constructs an immersive perceptual field that invites audience's bodily engagement. This mechanism can be explained through Embodied Cognition theory, which posits that human understanding of the world relies not only on abstract symbol processing but also on perceptual schemas accumulated through bodily senses.

In Li Ziqi's videos, there is repeated collaboration between color tone and shot scale (visual) and ambient sound (auditory). In the stir-frying scene from *Liuzhou “Luosifen”*, high-contrast, highly saturated images of food preparation and sustained close-up shots aim to evoke imagination of flavors through visual signs. Meanwhile, the ambient sounds of sizzling oil and ingredients colliding enhance the sense of authenticity and tactility. Together, they activate audience's past dining experiences, triggering synesthetic metaphors that map from the visual and auditory domains onto gustatory and olfactory sensations, enabling audience to construct a perceptual experience of deliciousness in their imagination. As one commenter (@jessdoritowhale) noted:

...i can even SMELL whats going on through the screen...(excerpt)

The audience's mention of being able to “smell” the food on screen demonstrates the successful operation of the synesthetic metaphor.

4.2.3 Visual Mode and Light Music: Construction of Emotional Space

The final sub-dimension of visual-auditory collaboration is the interaction between the visual mode and light music, which primarily functions to evoke emotional atmospheres and provide an affective tone and psychological container for audience's bodily projection.

Piano music, for instance, creates a relaxed and peaceful ambiance that brings healing and

comfort to the mind. In the wine-pouring scene from *A Pot of Wine Among Watermelon Fields*. So...*The life of watermelons and grapes?*, the melodic undulations of the piano guide the free flow of the audience's thoughts, transforming the leisurely visual scene into a multisensory, immersive psychological space. This collaboration triggers emotional bodily experiences and offers the audience a therapeutic mental refuge—an extension of embodied cognition. As one commenter (@Yogartist_PalakMehra) shared:

When I was on my periods and facing a very bad cramps I just take hot water bottle and watch lizi's video it's so therapeutic keep going and growing lizi !

During physical discomfort, the audience found a “therapeutic” experiential response within the audiovisual space constructed by Li Ziqi's videos.

When traditional Chinese instruments are used with the visual mode, the videos become an organic whole that allows immersive perception of classical Chinese aesthetics. In a segment from *The Life of Purple Rice, Pumpkins, and ... Peanuts*, the visual mode shows Li Ziqi in traditional attire performing various procedures, accompanied by the rising and falling melody of a bamboo flute in the auditory mode. As a representative traditional Chinese instrument, the flute's sound creates intertextual resonance with traditional attire, ancient craftsmanship, and natural landscapes in the frame. Together, the audiovisual signs from different senses weave a Chinese classical aesthetic field accessible to bodily experience, allowing the audience to “feel” an aesthetic quality of tranquility, composure, and harmony with natural rhythms at a corporeal level. One commenter (@timsoftx) wrote:

May the peace be with everyone who is watching (excerpt)

The word “peace” directly echoes the tranquil aesthetic quality. Another commenter (@jesswangpastry) remarked:

I feel like I'm watching an ancient Chinese drama. (excerpt)

The description “ancient Chinese drama” encapsulates the content, indicating that the audience captured the spiritual core of traditional Chinese culture from the labor scenes.

4.3 Visual-Auditory-Editing Collaboration: Internalization from Cultural Metaphor to Emotional Experience

4.3.1 Visual-Editing Collaboration: Regulation of Narrative Rhythm, Transformation of Temporal Perception, and Construction of the “Agricultural Temporality” Metaphor

This study further subdivides visual-editing collaboration into three subtypes: visual with editing transitions, visual with empty shots, and visual with time-lapse photography.

First, the collaboration between visuals and editing transitions serves to regulate rhythm and enhance information density. The sample videos frequently employ a “parallel editing” technique. In *Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival*, for example, while one snack is being prepared, the footage inserts close-ups of a finished product from a previous step, reinforcing visual memory of prior achievements. This strategy overlays multiple meaning units within the same time frame, allowing the audience to accumulate a sense of progressive accomplishment during the continuous production process.

In the latter part of the videos, the creator also uses “rapid switching” to depict the final stages of New Year goods preparation. Here, high-contrast, highly saturated color tones in the visual mode collaborate with accelerated editing rhythms to evoke the vibrant, bustling ritual atmosphere of traditional Chinese festivals, metaphorically conveying Chinese people's anticipation and well-wishes for the occasion. As one commenter (@tracy_1983) observed:

“Her videos came to my attention by my 10-year-old daughter, she loved watching her videos on YouTube. Now we are watching it together and we both are fascinated and awe-inspiring by her

works, we are learning Mandarin and this video will help us to know a little more about the traditional Chinese food etc.”

The commenter’s mention of being “fascinated” and “awe-inspired” echoes the “sense of progressive accomplishment” created by editing transitions. At a deeper level, the shift from individual viewing to intergenerational co-viewing and engagement with language learning indicates that this sense of achievement further motivates sustained exploration of and actual investment in Chinese traditional culture.

Second, the collaboration between the visual mode and empty shots functions to create narrative breathing space and transform temporal perception. Empty shots are typically inserted at narrative junctures. In *A Pot of Wine Among Watermelon Fields. So...The life of watermelons and grapes?*, after completing the watermelon rind pickle and before moving to the next preparation, the creator inserts an empty shot focused on a sunflower in a natural setting. This provides the audience a moment of respite within the otherwise compact narrative rhythm. More profoundly, the visual-empty shot collaboration points to a cultural translation of temporality: empty shots restore time as a flowing, inhabitable experience synchronized with natural rhythms—precisely the audiovisual embodiment of the traditional Chinese concepts of “acting in accordance with natural rhythms” and “harmony between man and nature.” As one commenter (@alexawagner2514) noted:

Every video feels like stepping into a timeless painting... the harmony between nature, food, and tradition here is pure poetry. (excerpt)

The phrase “stepping into a timeless painting” captures the quality imparted by visual-empty shot collaboration: time is slowed and extended, and everyday labor is elevated into an artistic tableau for contemplation. The comment explicitly identifies “the harmony between nature, food, and tradition” and summarizes it as “pure poetry,” reflecting conscious recognition and emotional identification with the “harmony between man and nature” ideal among the audience.

Within the visual-editing framework, visual-time-lapse photography collaboration further manifests in the construction of the “agricultural temporality” metaphor. Time-lapse sequences in Li Ziqi’s videos typically compress natural processes spanning days or months into mere seconds of footage. In *A Pot of Wine Among Watermelon Fields. So...The life of watermelons and grapes?*, between shots of human labor activities such as sowing, watering, and tilling, the creator inserts time-lapse sequences of watermelon seedlings emerging and leaves unfurling. The audience thus perceives, through the rhythmic alternation between “human labor” and “natural growth,” that human effort should align with natural rhythms. By making time the narrative subject, time-lapse photography highlights its cyclical nature that synchronizes with rhythms of life—this is the core meaning of the “agricultural temporality” metaphor. As evidenced by commenter @fulviab:

Her videos take you to a beautiful place and you experience the passage of time with her as she teaches you how things are done. (excerpt)

Although the commenter does not use the term “agricultural temporality,” they perceptually recognize a time flow distinct from everyday experience—one synchronized with labor and nature. This constitutes an indirect manifestation of the “agricultural temporality” metaphor at the reception side.

4.3.2 Auditory-Editing Collaboration: Construction of the “Pastoral Utopia”

The core mechanism of auditory-editing collaboration lies in the linkage between light music and time-lapse photography. This pattern not only extends the earlier discussion of temporal perception but also, through deep collaboration of auditory signs and editing rhythms, completes the construction of a complete “utopian” spacetime that the audience can imaginatively inhabit. In *Liuzhou “Luosifen”*, gentle xiao (vertical flute) notes lay a serene and peaceful emotional foundation, while time-lapse photography captures the shifting light and clouds across the sky,

compressing hours of natural change into seconds. This creates a heterogeneous spacetime characterized by “false prolongation” [15] that remains synchronized with natural rhythms. Accompanied by soothing light music, the audience gains a leisurely, “lingering” psychological space and become immersed in a slow-paced “pastoral utopia.” Commenter @digantamajumder5900 confirmed the success of this construction:

The whole thing just looks like a fairytale.... isn't it?

The term “fairytale” precisely captures the idealized and poetic quality of the experience—an imaginative space distanced from the real world yet emotionally aspirational. This is exactly the inhabitable “utopian” spacetime we describe, one that stands apart from the linear time of industrialization.

5. Conclusion

5.1 Theoretical Contribution: A Hierarchical Progressive Mechanism of Multimodal Collaborative Metaphor Generation

This study shows that different modes in Li Ziqi’s videos work together in layered ways to shape cultural meaning and influence audience response. At the first level, collaboration between visuals and titles/subtitles provides initial cognitive guidance for the cross-cultural audience. At the second level, collaboration between visuals and dialect, ambient sound, and music preserves cultural subjectivity while constructing the audience’s embodied perceptual experiences of cultural symbols. At the third level, further involvement of visual-editing and auditory-editing modes occurs: visual-editing collaboration regulates narrative rhythm through visual modes and editing transitions, creates narrative breathing space and reconstructs temporal perception via visual modes and empty shots, and reveals natural rhythms and the “agricultural temporality” metaphor through visual modes and time-lapse photography; auditory-editing collaboration uses light music and time-lapse photography to construct idealized spacetime and a pastoral utopia. Together, these multimodal strategies help the audience move from basic understanding to emotional engagement with the cultural values presented in the videos. The three levels exhibit a logic of sequential progression and gradual deepening.

This “cognition-perception-emotion” analytical framework for multimodal collaborative metaphor generation elucidates the specific mechanisms by which folk media deconstruct abstract, deep cultural metaphors into concrete, lifeworld audiovisual expressions. It provides a refined analytical tool for understanding the “lifeworld translation” of Chinese culture in the context of globalization, which is precisely the core cross-cultural communication advantage that distinguishes folk media from official grand narratives. These findings may also contribute to ongoing discussions about how cultural content with Chinese characteristics is communicated in global digital media environments.

5.2 Practical Implications: Reference Value for Other Media in Cross-Cultural Communication

By combining visual, linguistic, auditory and editing modes, Li Ziqi’s videos present traditional Chinese culture in ways that feel concrete, lifeworld and relatable to the overseas audience. This approach can help the audience gradually develop understanding and resonance by following the aforementioned hierarchical mechanism. The findings suggest that cross-cultural communication is not always dependent on direct explanation. Instead, through well-designed multimodal collaboration strategies, deep cultural metaphors and meanings can be internalized and embraced by gradually influencing the audience’s “perception” and “emotion.”

These findings hold referential value for video creators engaged in cross-cultural expression.

First, in terms of modal configuration strategy, creators can design video collaboration following the logic of “visual-linguistic collaboration to build cognitive frameworks, visual-auditory collaboration to construct embodied experiences, and visual-editing and auditory-editing collaboration to deepen emotional resonance,” thereby replicating Li Ziqi’s success in constructing cultural metaphors and meanings from shallow to deep in a layered manner. Second, regarding the balance between cultural subjectivity and cross-cultural adaptation, Li Ziqi’s videos demonstrate a viable strategy: indigenous cultural elements such as the solar term system, dialect, and traditional instruments are not weakened or replaced in the process of cross-cultural dissemination. Instead, through collaboration with other modes, they become easier for the overseas audience to understand and emotionally connect with. This strategy shows that cultural communication need not come at the cost of diluting its own characteristics; cultural differences themselves can become appealing when presented through effective audiovisual storytelling.

5.3 Research Limitations and Future Directions

This study has some limitations in three aspects. First, the representativeness of comment data is limited: purposive sampling was used to select comment samples, resulting in a relatively small number that cannot fully cover decoding patterns across audiences from different cultural backgrounds. Second, the sample size is still relatively limited: although the five selected videos’ immense popularity partly demonstrates the superiority and representativeness of their modal configuration strategies, they cannot cover the full repertoire of multimodal strategies used across Li Ziqi’s videos. Third, the depth of theoretical integration needs improvement: while this study attempts to combine multimodal metaphor theory with embodied cognition, deeper integration of the two theories within cross-cultural communication contexts remains insufficient. Future research could incorporate additional perspectives such as symbolic interactionism and reception aesthetics to construct a more comprehensive theoretical framework, and provide a more nuanced understanding of how multimodal communication operates in cross-cultural contexts.

References

- [1] Halliday, M. A. K., & Matthiessen, C. M. (1994). *An Introduction to Functional Grammar*. London/New York: Arnold.
- [2] Kress, G. R., Van Leeuwen, T., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Psychology Press.
- [3] Zhang, D. L. (2009). *Exploration of a Comprehensive Theoretical Framework for Multimodal Discourse Analysis*. *Foreign Languages in China*, 6(01), 24-30. <https://doi.org/10.13564/j.cnki.issn.1672-9382.2009.01.004>.
- [4] Forceville, C. (1996). *Pictorial Metaphor in Advertising*. Psychology Press.
- [5] Zhao, X. F. (2011). *New Developments in Conceptual Metaphor Research: Multimodal Metaphor Research: A Review of Multimodal Metaphor by Forceville & Urios-Aparisi*. *Foreign Languages Research*, (01), 1-10+112. <https://doi.org/10.13978/j.cnki.wyyj.2011.01.005>.
- [6] Xie, J. X. (2011). *Metaphor from a Multimodal Perspective: A Review of Charles Forceville’s Metaphor Research*. *Foreign Language Research*, (05), 49-52. <https://doi.org/10.16263/j.cnki.23-1071/h.2011.05.009>.
- [7] Pan, Y. Y. (2011). *Multimodal Metaphors and Identity Construction in Political Cartoons*. *Foreign Languages Research*, (01), 11-15. <https://doi.org/10.13978/j.cnki.wyyj.2011.01.006>.
- [8] Feng, D. Z., & Xing, C. Y. (2011). *Spatial Metaphor and Multimodal Meaning Construction: A Case Study of Car Advertisements*. *Journal of Foreign Languages (Journal of Shanghai International Studies University)*, 34(03), 56-61.
- [9] Ladilova, A. (2020). *Multimodal Metaphors of Interculturereality/Metaforas multimodais da interculturalidade*. *Revista de Estudos da Linguagem*, 28(2), 917-955.
- [10] Wang, Y. (2017). *The Application of “Three Views” Theory in Post-Linguistic Philosophy to Video Communication: A Case Study of the “Chongqing Tourism Image Promotion Video.”* *Journal of Nantong University (Social Sciences Edition)*, 33(06), 144-149.
- [11] Song, H. L., Zhao, Z. H., & Song, Y. Q. (2025). *Visual Rebirth: A Study on the Translation of Multimodal Intercultural Metaphors in Chinese Classics*. *Journalism & Communication Review*, 78(06), 112-124.

<https://doi.org/10.14086/j.cnki.xwycbpl.2025.06.010>.

[12] Qu, Y. (2025). *Li Ziqi's Chinese idyllic dream: Analysis of the symbolic characteristics of vernacular culture from symbolic aggregation theory/semiotic theory*. *TAMA: Journal of Visual Arts*, 3(1), 35-45.

[13] Li, J., & Adnan, H. M. (2023). *Exploring the role of productive audiences in promoting intercultural understanding: A study of Chinese YouTube influencer Liziqi's channel*. *Studies in Media and Communication*, 11(4), 173-184.

[14] Li, L. T., Hu, Q. Y., Peng, Y. P., Zheng, J. X., & Lu, M. D. (2025). *Research on the Communication of "Chinoiserie" Culture from the Perspective of Multimodal Metaphor: A Case Study of Li Ziqi's Videos*. In *Innovative Research on Foreign Language Education and Translation Development* (17) (pp. 312-314). School of Foreign Languages, Chengdu University of Information Technology. <https://doi.org/10.26914/c.cnkihy.2025.011178>.

[15] Li, T. Y., & Jiao, J. X. (2024). *Overseas Communication Strategies for Traditional Chinese Culture: A Case Study of Rural-Themed Short Videos on YouTube*. *China Publishing Journal*, (13), 40-44.