

A Study of Multimodal Translation Models for Tourist Attractions in Guangzhou: A Case Study of Canton Tower

Lyuyao Jiang, Hua Qu

South China Business College, Guangdong University of Foreign Studies, Guangzhou, Guangdong, 510545, China

ABSTRACT

With the deepening integration of digital technology and the cultural-tourism industry, the cross-cultural communication of tourist attractions urgently needs to move beyond the limitations of single-text translation toward multimodal, coordinated expression. As an international cultural and tourism landmark, Canton Tower exemplifies multimodal translation through the integration of text, images, video, and audio. However, existing research primarily focuses on traditional textual translation and lacks a systematic discussion of the synergy among semiotic modes, technological support, and cultural adaptation strategies, thereby limiting the effectiveness of international communication. Based on multimodal translation theory, this study attempts to analyze the intermodal collaboration of text, visual, audio, video, and interactive modes in the translation of Canton Tower and other Guangzhou tourist attractions. It proposes four practical models: "text + image," "video + subtitles," "app interaction + audio guide," and "social media multimodal dissemination." The study dissects the multimodal composition and collaborative mechanisms of Canton Tower's translation practices, offering a replicable model for the global communication of Chinese cultural-tourism content.

KEYWORDS

Multimodal translation; Canton tower; Cultural translation; Translation model

1 Introduction

With the advancement of digital technology, the communication of tourist attractions has shifted from single-language texts to multimodal expressions. As Gambier note, "No text is, strictly speaking, monomodal" (Gambier, 2006)—meaning is constructed through the interplay of multiple semiotic resources. In the digital age, multimodality has become a fundamental feature of meaning-making, interpretation, and reception (Wu 2021). Canton Tower, the tallest tower in China, is not only a cultural icon of the city but also a prime example of international tourism integration. Its translation practices encompass various modes—text, imagery, sound, video, and interactivity—making it an ideal case for multimodal translation research. However, existing studies largely remain focused on traditional textual translation, with insufficient attention paid to the mechanisms of multimodal coordination.

2 Theoretical framework: multimodal translation theory

Multimodal Translation Theory, a significant branch of translation studies in the digital era, centers on the coordinated transformation of linguistic and non-linguistic signs (such as images, sound, gestures, etc.) to achieve accurate cross-cultural meaning transfer. This theory emerged in the late 20th century through interdisciplinary developments: traditional translation theories (e.g., Nida's functional equivalence, Toury's descriptive translation studies) provided a foundation for linguistic transformation, while Kress and Van Leeuwen's visual grammar extended semiotics to non-verbal modes, analyzing representational, interactive, and compositional meanings of images (Kress & Van Leeuwen, 2006).

In the digital age, the nature of translation has evolved from static text to dynamic multimodal discourse, demanding that translators possess skills in cross-modal coordination. The rise of fragmented knowledge and digital consumption has further elevated the appeal of multimodal texts—those that incorporate visuals, audio, and video—placing new demands on translators' abilities to integrate multiple modes effectively (Jiang & Li, 2024).

In China, following Li Zhanzi's introduction of multimodal discourse analysis, Zhu Yongsheng (2007) proposed identification criteria for multimodal discourse, emphasizing that multimodality must involve at least two modes or semiotic systems (e.g., comics comprising both words and images). Zhang Delu (2009) further integrated Systemic Functional Linguistics (SFL) into multimodal theory, establishing an operational framework for multimodal translation. Alonso, Molina, and Porto (2015) examine how various semiotic channels in digital storytelling—such as factual, emotional, and cultural cues—contribute distinct types of information that are ultimately integrated to form the overall meaning of a narrative. Burgess and Green (2018) emphasize the role of multimodal narratives in enabling content creators to construct online identities and tell stories. Djonov and Tseng (2021) argue that transmedia narratives actively engage audiences and enhance literacy education, especially for young learners. Pereira and Campos (2023) explore how focalisation within multiliteracies pedagogy fosters critical literacy skills, particularly in the analysis of multimodal

narratives. Osorio et al. (2024) investigate the function of co-speech gestures in naturalistic multimodal narrative processing, finding that such gestures support lexical retrieval and may aid semantic prediction at the neural level. Jabeen and Cheong (2024) analyze the construction of social meaning through multimodal narratives in TV advertisements, proposing that a multi-perspective approach enhances the interpretation of visual texts.

This study attempts to explore the intermodal collaboration of text, visual, audio, video, and interactive modes in the translation of Canton Tower and other Guangzhou tourist attractions and proposes four practical models: "text + image," "video + subtitles," "app interaction + audio guide," and "social media multimodal dissemination." It is hoped to offer a replicable model for the global communication of Chinese cultural-tourism content.

3 Multimodal composition of Canton Tower's translation discourse

3.1 Textual mode

Textual mode refers to how language is organized and structured into coherent and meaningful texts. It focuses on the form and flow of a message, helping it make sense to a reader or listener within a particular context. In Systemic Functional Linguistics (SFL), developed by M.A.K. Halliday, textual mode (or textual metafunction) is one of the three metafunctions of language, alongside ideational and interpersonal metafunctions. The textual metafunction is concerned with the creation of text, with the presentation of ideational and interpersonal meaning as information that can be shared through language (Halliday, & Matthiessen, 2014, 30). The official website and on-site signage at Canton Tower employ bilingual (Chinese-English) texts that focus on culturally adaptive translations. For example, "广州塔" is translated into Canton Tower rather than Guangzhou Tower, preserving historical and cultural connotations while aligning with international usage.

Additionally, its superlative features are effectively translated to attract tourists, such as: "世界最高的横向摩天轮" into "World's highest horizontal Ferris wheel", "世界最高户外观景平台(488米)" into "Highest outdoor observation deck (488m)", "世界最高垂直速降项目" into "Highest vertical drop thrill ride in the world". These translations, though purely textual, instantly capture the attention of international visitors.

3.2 Visual mode

The visual mode involves elements such as colour, vectors, perspective, and framing. It allows us to represent ideas, convey emotions, and structure information visually. (Kress & Van Leeuwen (2006, 2-3). The tower's light shows represent a synergy of symbolic coordination and cultural transference. For instance, on the evenings of February 12 and 13 (the third and fourth days of the Lunar New Year), at 7:30 PM, a total of 2,024 drones performed over the river between Haixinsha and Canton Tower. Centered around the theme "Lingnan Charm, Chinese Flavor, Global Sentiment," the show featured scenes of Spring Festival traditions and Lingnan culture, showcasing a fusion of heritage and modern technology. During the Qixi Festival in 2024, Canton Tower hosted an event themed on the legend of the Seven Fairies descending to earth. In combination with a coordinated drone performance, the event displayed breathtaking aerial patterns that highlighted the romance and allure of Lingnan intangible cultural heritage.

This affirms Guangzhou's identity as an international metropolis and caters to both foreign tourists and social media sharing. The format aligns with genre expectations of global landmarks (e.g., Eiffel Tower), where symbolic lighting functions as a form of ritual display. In this context, text anchors meaning, while dynamic lights construct emotional atmosphere and rhythm, demonstrating complementary intermodal relationships.

3.3 Audio mode

The audio mode includes sound such as music, ambient noise, sound effects, silence, and the qualities of speech like volume, pitch and intonation. These elements can strongly affect the emotional and interpretive impact of a message (Jewitt, 2009, 29).

The tower incorporates multiple audio elements: Audio guides in multiple languages are available on-site, combining English narration with original videos containing traditional cultural references. Promotional videos feature grand, rhythmic voiceovers accompanied by strong drumbeats or orchestral music to convey cultural richness. Commercial audio content often includes emotionally charged voiceovers and branded jingles to enhance tourism marketing effectiveness.

3.4 Video mode

The video mode combines moving image, spoken language, music, and visual design to create layered meanings. It is

inherently multimodal, bringing together different semiotic modes in time-based representation (Burn, 2003, 8). Through its website and on-site displays, Canton Tower promotes integrated tourism products like the “Princess Pearl River Cruise (珠江公主号),” where the cruise ship features Qilou-style architecture and visual motifs such as Cantonese opera masks, creating a visual dialogue between tradition and modernity. Notably, Canton Tower has ventured into digital crossovers, including: integration with the video game *Peacekeeper Elite*, launch of the “Canton Tower and intangible Cultural Heritage” NFT collection. These examples illustrate how video-driven strategies and technological tools empower cultural communication.

3.5 Interactive mode

The interactive mode involves the resources used for interaction between people—such as speech, gestures, facial expressions, and writing—which facilitate engagement and response (Jewitt, 2009, 25). Using cutting-edge digital twin technology, Haizhu District has replicated the Canton Tower in the metaverse, building Guangzhou’s first smart city digital twin platform.

With over 1,972 dynamic monitoring parameters, the platform digitally reconstructs the attraction on a 1:1 scale, allowing real-time forecasting of visitor flow, traffic congestion, and weather impacts. It also identifies urban issues like littering or vandalism through AI analysis. Additionally, WeChat mini-programs enable one-click ticket booking, streamlining visitor services and enhancing the overall user experience.

4 Multimodal translation models in Guangzhou’s tourism

4.1 Text + image model

This bilingual image-text model is widely used in brochures, airport displays, exhibitions, and city showcases.

Example 1

ST: 广州塔, 海拔600米, 是中国第二高塔, 拥有世界最高的户外观景平台与横向摩天轮。在这里, 您可以360°俯瞰广州城市风貌。

TT: Canton Tower stands 600 meters tall, making it the second tallest tower in China. It features the world’s highest outdoor observation deck and horizontal Ferris wheel, offering a stunning 360° panoramic view of Guangzhou.

From Example 1, The image grounds abstract concepts like “600 meters tall” or “360° panoramic view” in visual reality. The text, in turn, guides the viewer’s attention to specific features in the image (e.g., the Ferris wheel), enabling deeper interpretation and appreciation. This combination enhances accessibility and the impact of information through the synergy of visual (image) and linguistic (Chinese + English) modes.

4.2 Video + subtitle model

Promotional videos employ dynamic subtitles, such as color-matching text, to reduce cultural misunderstandings. For example, “广府文化” is translated as “Cantonese Traditions” to avoid literal translation ambiguity.

Example 2

ST: 她, 是广州的高度, 是珠江之畔的光影传奇。

TT: She is the height of Guangzhou — a shimmering legend by the Pearl River.

In Example 2, Personification (“She”) is supported by visuals that present the tower with reverence—framing it as a character. The subtitle line “a shimmering legend” likely coincides with shots of the tower lighting up—synchronizing text with visual metaphor. The Pearl River is mentioned just as the camera likely pans to or from the riverbank, grounding the metaphor in spatial context. Video visuals convey spatial aesthetics and emotion, while subtitles provide linguistic clarity, cultural context, and accessibility for the hearing-impaired and non-Chinese speakers. Background music (orchestral or ambient electronic) enhances immersion.

4.3 App interaction + audio guide model

This model combines touchscreen navigation with audio narration for a more immersive experience, ideal for smart tours and barrier-free services.

Upon entry, there is the audio guide “Welcome to Canton Tower. You are now at the 107-meter-high Science Exhibition Hall.” While selecting a feature, there is “At 455 meters is the world’s highest horizontal Ferris wheel, offering a 360° view of Guangzhou’s nightscape.” At the top: “You are approaching an outdoor high-altitude area. Please fasten your seatbelt and avoid leaning on the railings.”

The App Interaction + Audio Guide Model in this example shows how digital systems can provide just-in-time

multimodal communication that is: user-centered (driven by interaction and movement), context-sensitive (adaptive to real-world location), functionally diverse (informing, engaging, and protecting the user). This model is particularly effective in tourism, museums, and exhibitions, where layered information and spatial experience converge. The App interaction + Audio Guide Model shows how user interface (UI) interaction, audio narration, and contextual awareness combine to create a cohesive and immersive multimodal experience.

4.4 Social media multimodal dissemination

Canton Tower's cultural branding extends through short videos and hashtags on overseas platforms like Instagram, Facebook, and X. For instance, In October 2024, a foreign user shared night-show videos under the #CantonTower and #goodnight tags on Instagram, featuring the scrolling Goodnight text. The post received hundreds of likes and sparked cross-cultural resonance.

5 Conclusion

The multimodal translation practices of Canton Tower demonstrate how the integration of semiotic resources and digital technologies can simultaneously enhance cultural transmission and tourist experiences. Specifically, in terms of cultural adaptation, bilingual texts integrated with images reduce cultural barriers and misinterpretation. In technological integration, tools like AR and smart audio guides improve interactivity and immersion. In communication synergy, bridging offline attractions with online social platforms creates a seamless promotional network.

Future research should focus on two key areas. The first is to develop standardized multimodal corpora to systematically analyze semiotic synergy. The second one is to optimize translation strategies based on user feedback to better align with global audience expectations, thereby advancing the international reach of Chinese cultural-tourism brands.

Funding

The Department of Education of Guangdong Province (S202312620003); fund project : The Planning Office for Philosophy and Social Sciences of Guangdong Province (GD23WZXC01-11)

About the Author

Lyuyao Jiang, student, majoring in Translation, Class of 2022 at South China Business College, Guangdong University of Foreign Studies.

References

- [1] Alonso, I. Molina, S. and Porto, M. D. Multimodal Digital Storytelling Integrating Information, Emotion and Social Cognition, Review of Cognitive Linguistics [M]. 11:2. John Benjamins Publishing Company. 2015.
- [2] Burgess, J., & Green, J. YouTube: Online video and participatory culture. [M] London: Polity Press. 2018.
- [3] Burn, A., & Parker, D. Analysing Media Texts [M]. London: Continuum. 2003.
- [4] Gambier, Y. Multimodality and Audiovisual Translation [J]. Audiovisual Translation Scenarios: Proceedings of the Second MuTra Conference in Copenhagen, Denmark, 1-8. 2006.
- [5] Djonov, E. and Tseng, C.-I. Harnessing the Potential of Transmedia Narratives for Critical Multimodal Literacy [J]. Critical Discourse Studies, 18 (3), 349-367. 2021.
- [6] Halliday, M.A.K. & Matthiessen, C.M.I.M. (2014). Halliday's Introduction to Functional Grammar [M]. (4th ed.). London: Routledge.
- [7] Jewitt, C. (Ed.). (2009). The Routledge Handbook of Multimodal Analysis [M]. London: Routledge. 2009.
- [8] Jabeen, S. and Cheong, C. Y. M. Multimodal Narrativity in a Pakistani TV Advertisement: a Socio-semiotic and Narratological Analysis [J]. Social Semiotics, 34 (1), 61-81. 2024.
- [9] Jiang, H. Y. & Li, X. Y. A Study on Multimodal Translation of Ecological Civilization Texts in the Digital Age [J]. Modern Linguistics, 2024, 12 588-594. (2), 2024.
- [10] Kress, G. & Van Leeuwen, T. Reading Images: The Grammar of Visual Design [M] (2nd ed.). London: Routledge. 2006.
- [11] Pereira, Í.S.P and Campos, Â. Constructing the Pedagogy of Multiliteracies. The role of Focalisation in the Development of Critical Analysis of Multimodal Narratives [J]. Language and Education, 37 (1), 105-122. 2023.
- [12] Osorio, S., Straube, B., Meyer, L. and He, Y. F. The Role of Co-speech Gestures in Retrieval and Prediction during Naturalistic Multimodal Narrative Processing [J]. Language, Cognition and Neuroscience, 39 (3), 367-382. 2024.
- [13] Wu, B. The Media I Turn and Multimodal Translation Studies [J]. Journal of Foreign Languages, 44(1), 115-123. 2021.
- [14] Zhang, D. L. On A Synthetic Theoretical Framework for Multimodal Discourse Analysis [J]. Foreign Languages in China, 6(1), 24-30. 2009.
- [15] Zhu, Y. S. Theory and Methodology of Multimodal Discourse Analysis [J]. Foreign Language Research, 5, 82-86. 2007.