Asia Pacific Science Press

Critical Humanistic Social Theory
ISSN(O):3005-9291
ISSN(P):3005-9283
Vol. 3 No.4 (2026)

A Multimodal Perspective on Subtitle Translation of Documentaries: A Case Study of Masters in the Forbidden City

Yuying Zhang*, Chen Jiao

School of International Studies, Xi'an University of Finance and Economics, Xi'an, 710100, China

Corresponding author: Yuying Zhang, zhangyuying0317@163.com

Copyright: 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY-NC 4.0), permitting distribution and reproduction in any medium, provided the original author and source are credited, and explicitly prohibiting its use for commercial purposes.

Abstract: As a distinctive audiovisual genre, cultural documentaries serve as both a carrier of national heritage and a bridge for cross-cultural communication. Masters in the Forbidden City, a critically acclaimed Chinese documentary on cultural relic restorers in the Palace Museum, achieved remarkable domestic and overseas success, largely owing to its well-crafted English subtitles. Despite its exemplary status, scholarly attention to its subtitle translation remains scarce, with existing studies predominantly confined to linguistic equivalence while overlooking the dynamic semiotic interplay of verbal, visual and auditory modes. Based on Christopher Taylor's multimodal text analytical framework, this paper investigates the subtitle translation of Masters in the Forbidden City from narrative, linguistic, semiotic, and cultural considerations. Through a qualitative case analysis of representative subtitle excerpts, it examines how the translator achieves multimodal cohesion through strategic adoption of translation strategies including free translation, omission, specification and explanatory compensation. The findings reveal that effective documentary subtitle translation requires a shift from text-bound translation to screen-aware practice, wherein visual images, soundtracks, and narrative contexts dynamically constrain and enable translator decision-making. This study enriches the multimodal perspective on the subtitle translation of cultural documentaries and offers practical reference for audiovisual translation research and practice.

Keywords: Audiovisual Translation; Documentary Subtitle Translation; Masters in the Forbidden City; Multimodal Text Analysis; Multimodality

Published: Nov 12, 2026

DOI: https://doi.org/10.62177/chst.v3i4.1813

1. Introduction

Along with the acceleration of science and technology, audiovisual products have made much headway in the market and brought tremendous changes to social life. Various modalities such as visual images, sound, and action constantly spring up, hence the multimodal discourse appears and has gradually taken the place of the traditional single discourse mode [1]. Documentary is a distinctive audiovisual product whose multimodal characteristics deserve deeper investigation in academic study. It is worth noting that recently Chinese documentaries have entered into the international market with an eye-catching performance through subtitling, represented by the cultural documentary Masters in the Forbidden City. Broadcast on China Central Television (CCTV) in 2016, this documentary presents the work of cultural relic restoration masters in the Forbidden City and its excellent subtitling contributes considerably to its success in both domestic and overseas markets, which deserves

much more professional attention.

However, despite the fast development of audiovisual translation (AVT) practice, research on documentary translation has long been marginalized within AVT scholarship [2]. Current research is far from sufficient not only in terms of quantity and quality, but also in research scope. In dealing with subtitle translation of documentaries, most research is merely centered on linguistic level of the text whereas few studies have ever touched upon the influence of multimodal resources on the translator's choices. Based on Christopher Taylor's multimodal text analysis, this study investigates the subtitle translation of Masters in the Forbidden City from a multimodal perspective to analyze the possible effects of various modes on the translator's choices, in an attempt to provide new insights and shed light on multimodality in documentary translation.

2. Theoretical Basis

The concept of multimodality cuts across every discipline and has been defined by social semiotic scholars Kress and Van Leeuwen as "the use of several semiotic modes in the design of a semiotic product or event" [3]. Accordingly, Baldry and Thibault hold that multimodal texts refer to composite products with combined effects of all the semiotic resources for interpretation, and based on this insight they developed an innovative approach to multimodality, namely multimodal transcription, which paved the way for Christopher Taylor's multimodal text analytical framework [4]. Among scholars who have conducted comprehensive analysis in this field, Taylor has made great efforts in search for a multimodal approach to AVT research. Previous studies of multimodality provided an impetus to develop ideas on exploitation of multimodal analyses in AVT. Taylor thinks highly of Baldry and Thibault's innovative approach that divides a film into shots, frames, and phases for subtitling analysis, regarding it as a basis for further development. Based on multimodal transcription, Taylor has proposed his own model of multimodal text analysis from four distinct perspectives, including narrative consideration, linguistic consideration, semiotic consideration and cultural consideration [5]. This multimodal approach attempts to figure out how translators can be made sensitive to semiotic impact of a multimodal text and has largely expanded the scope of AVT research.

3. Case Analysis of Subtitle Translation of Masters in the Forbidden City

As a prelude to any serious translation task, text analysis in multimodal case study should move away from exclusive reference to words and sentences and recognize other well-established meaning-making units in film language [6]. Themed on excellent traditional Chinese culture, Masters in the Forbidden City is a typical multimodal discourse whose meaning construction relies on various meaning-making resources of more than one semiotic modality such as written modality, sound modality and visual modality. The following will mainly examine successful translation techniques adopted in the subtitle translation of this documentary from a narrative, linguistic, semiotic and cultural point of view, so as to explore how translator's decision making is influenced in the interplay of these different semiotic modes.

3.1 Narrative Consideration

Film and television series generally tell stories in a multimodal way [7]. Audience perceives filmic products not only through verbal language, such as dialogue, but also through transitions among various semiotic modes. According to Taylor, as viewers watch successive frames of a film, they form expectations about how the narrative will develop, and these expectations are fulfilled through an array of semiotic clues. [5]. In this light, translating verbal language alone is far from enough and narrative consideration becomes more than significant in audiovisual translation. As a cultural and historical documentary, the narration in Masters in the Forbidden City is a distinctive filmic language, whose major functions lie in visual image complementation, subject deepening as well as aesthetics improvement. "Translation, even of the intersemiotic kind, should be able to preserve narrativity and its basic mechanisms" [8]. A qualified audiovisual (AV) translator is supposed to consider the specific context in documentary narration as well as its distinctive narrative style, so as to reconstruct the story coherently for the target audience. Typical examples are listed below to show desirable translation techniques from the perspective of narrative consideration.

Example1:

Source Text: 真正咱们好的那个传统的东西，让你接笔完了以后，你根本看不出来。

Target Text: But for us, with the real good traditional techniques, you can never tell after the restoration.

As a specialized jargon in cultural relic restoration, “ 接笔 ” refers to a traditional restoration technique for ancient painting and calligraphy, which requires craftsmen to fill in missing parts of an antique to reproduce its beauty [9]. During this complicated process, craftsmen cannot rely on their imagination to draw at their own will nor can they recreate supplementary patterns directly, rather, they are required to add words, lines or colors with careful deliberation based on the original pictograph to restore its original appearance to the greatest extent. The narration carries dense information that may impose cognitive burdens on the target audience. Such an unfamiliar jargon might even cause obstacles for most Chinese audience without thorough background knowledge, not to mention about the target readers. Restricted by limited time and space, it is also unlikely to explain this complicated restoration technique in a brief and clear manner. In this case, the proper rendering of this jargon has become a great challenge for subtitle translation. It could be noted that the translator here has adopted the free translation strategy and the technique of omission to treat this difficulty with narrative consideration. Since visual and written modes in the narration have already provided comprehensive information about this term, the translator has omitted the redundant description. With clear visual modes and the specific context of the dialogue, it is not difficult for the audience to gain a proper understanding. The translator therefore extracts the core meaning of “ 接笔 ” and simplifies it into “restoration”, which maintains synchrony and enables the target audience to perceive this technical jargon in a more concise and comprehensible way.

Example 2:

Source Text: 行了，齐活。

Target Translation: All right, mission accomplished.

Narrative consideration requires the AV translator to gain a proper understanding of story creation as well as target audience’s expectations, so as to ensure that the translator could find words that successfully render the effect of the whole semiotic event in the target language [5]. The story of cultural relic restorers is narrated through dialogues with a strong Beijing flavor, contributing to the distinctive narrative style of Masters in the Forbidden City. “ 齐 活 ” is a daily expression in traditional Beijing dialect that means the work or task is completed. Through these two simple Chinese characters, audience could easily perceive the straightforward and simple personality of those Beijing cultural relic restorers. Under narrative consideration, the translator is supposed to consider target audience’s expectations of reproducing distinctive narrative style with Beijing flavor, thus the proper rendering of Beijing dialect becomes a major translation challenge. In this example, since the source text contains only four Chinese characters and correspondingly, the translator also uses four words to match the original language pattern and the rendering of “ 齐 活 ” into “mission accomplished” successfully reproduces the simple and direct style of Beijing dialect. By adopting this strategy, the translator retains the unique narrative style and delivers succinct, intelligible narration for target viewers.

3.2 Linguistic Consideration

In terms of the linguistic dimension of multimodal text analysis, the concepts of coherence and cohesion are major concerns. Taylor emphasized that the linguistic theories on which translation has been based need to be extended more fully to the multimodal sphere [5] . That is to say, translator should use linguistic technique to reinforce intersemiotic cohesion to connect written modes with various semiotic modes such as soundtracks and images, making full use of all available information to create a coherent linguistic-visual work as a whole [10]. The subtitle translation of Masters in the Forbidden City is a process of language transformation from Chinese into English, thus linguistic differences between two languages demand that the translator should make necessary adjustments to tackle linguistic obstacles. Besides, the cohesive ties between different modalities like verbal and visual modes in this documentary are more than difficult to be transferred for target audience, which requires the AV translator to select appropriate translation strategies to make it explicit in the target text. Typical examples are provided below to show satisfactory translation with linguistic consideration.

Example 3:

Source Text: 这些东西小修好，要大修可惜了，大修好多原来东西就保持不住了。

Target Text: Minor repair will do good at it, while major repair will ruin it as you cannot keep many of its original parts.

Chinese features parataxis which refers to the realization of internal text connection through semantic connectivity instead of the linguistic form, whereas English emphasizes hypotaxis which refers to the connection of words or sentences through linguistic form [11]. The transformation of syntactical structures is a typical example that often appears in subtitling. In Masters in the Forbidden City, there exists a lot of daily dialogues to record the real work experience of those masters, resulting in frequent appearance of relatively loose Chinese sentence structures. In the above example, despite the absence of conjunctions or prepositions connecting the three simple Chinese sentences, it is still coherent in semantic meaning and easy for Chinese audience to grasp the inherent logic. In Chinese-English subtitling, however, the translator is suggested to clarify the internal semantic logic in Chinese expressions and add appropriate conjunctive words to achieve formal cohesion in English sentence building. In terms of linguistic transference, conjunctions such as "while" and "as" are added to suggest that cultural relics can only be repaired slightly and to further indicate the underlying reason. Correspondingly, the original dialogue has been successfully restructured in the target language to help target audience understand how meaning is transferred across various modes in a film text.

Example 4:

Source Text: 陶瓷还没出来呢。

Target Text: Some of the pottery group are still in here.

The documentary's audiovisual discourse integrates a wide range of semiotic modes through multimodal cohesion, which hinders the target audience from identifying cohesive ties and interpreting the implied underlying meaning. In this scene, cultural relic restoration workers share their daily routine to the interviewer. After a day's hard work, designated gatekeepers must make sure each worker has left the Palace Museum before closing the door, and visual image shows that the pottery group workers are still working in the museum. For Chinese audience, multimodal cohesion between words and image on the screen poses a great translation challenge. From the clear visual image, the AV translator perceives the multimodal cohesion between verbal mode " 陶瓷 " to the visual mode and understands that " 陶瓷 " here actually refers to pottery group workers rather than pottery itself. Translating the words alone into "pottery" would cause serious misunderstandings, hence specification technique has been adopted to specify the verbal message " 陶 瓷 " into "pottery group" to make the implied cohesive ties more explicit to the target audience. As a result, the whole multimodal discourse becomes more coherent and comprehensible.

3.3 Semiotic Consideration

Taking dynamic interactions of various semiotic modes into consideration distinguishes multimodal text analysis from traditional text analysis. According to Taylor, the key concept for audiovisual translation is that semiotic resources can be seen as complementation or emphasis to verbal message [5]. The indispensable role of semiotic resources in subtitling has also been stressed by Fryer and Freeman, who regard them as key elements in audiovisual products and argue that they deserve to be fully perceived by viewers [12]. Specifically, certain gaps in original audiovisual text could be filled with information from visual images or soundtracks. Through subtitling of this documentary, awareness of cognitive information carried by semiotic modes drives the translator to absorb all the relevant information and adopt flexible translation strategies for the appropriate wording of the target text. A key scene is selected to discuss how the translator represents the meaning of the source text through the integration of semiotic modes.

Example 5:

Source Text: 把这颜色点全好，一个窟窿得全个三遍两遍的。这张全完了它就好了。

Target Text: Try to restore all the color spots, every color-loss spot needs to inpainting twice or more. When you finish it, it will get better.

In the original Chinese subtitle, the word " 全 " appears three times, however, it would be difficult for the target audience to distinguish its meaning from the written mode alone. As a term in Chinese calligraphy and painting restoration, " 全 " is used here to refer to three different restoration skills respectively. Its specific meaning could be perceived from the documentary image, which depicts a restorer repairing ancient calligraphy with a special pen to fill in color-loss spots in a cautious and careful manner. Considering the mode of moving images, the translator selects the appropriate meaning of each " 全 ", as the

first refers to the restoration skill at a macro level, while the second to the more specific skill of inpainting color-loss spots and the third to the completion of the task. Correspondingly, the loss of information in written mode has been compensated by visual mode in this documentary. This instance therefore illustrates how semiotic modes contribute their specific meaning to the process of subtitling. With the semiotic consideration, the translator fully comprehends the source text by integrating the meaning of visual images to give a proper interpretation to the target audience as that to the source audience.

3.4 Cultural Consideration

Since documentary often represents the culture of a nation, AV translators need to be careful to address cultural differences across languages, communities and nations [5]. To handle extralinguistic cultural references in subtitle translation, scholars represented by Pedersen have proposed several practical translation methods including omission and specification [13]. However,

attention to the transfer of cultural-loaded modalities remains insufficient. For culture-bound scenarios in multimodal discourse, Taylor suggests that a qualified translator should expand his or her encyclopedic knowledge, cultivate cultural awareness, and take into account the cognitive limitations of a potentially large audience. One factor contributing to the great success of Masters in the Forbidden City in documentary industry is that it serves as a unique cultural carrier to represent excellent traditional Chinese culture in cross-cultural communication. In the subtitling of this documentary, cultural references embedded in various verbal and visual modes constitute a shared translation challenge, and the translator must therefore decide to what extent cultural elements should be domesticated or foreignized to promote cultural exchange. From a cultural perspective, the translator has adopted appropriate strategies and two key examples are provided below for further analysis.

Example 6:

Source Text: 那个时候是接班，接班进来。

Target Text: The practice at that time is succession meaning to take over your parents' working position.

Culture and language intertwine with each other in multiple modalities of this cultural documentary. If written modes are translated separately without consideration of cultural connotations, the majority of foreign viewers would probably be at a loss and confused by such cultural-loaded information. In terms of cultural considerations, the translator employs free translation and compensation to render cultural references. " 接 班 " is a Chinese cultural-loaded expression used in the traditional handicraft industry during a particular period in China, which means that younger generations could take over the positions vacated by their predecessors after retirement and continue to work in the same industry. This documentary records the hard work of masters in cultural relics protection from generation to generation and the implied cultural connotation of this expression also reflects these masters' selfless dedication to restoration work. Considering different cultural backgrounds, the translator adopts free translation to domesticate this term into "succession", which retains the core connotation in a simple and understandable way. The compensation technique has also been selected to add the necessary explanation and further clarify this culture-bound scenario, thereby facilitating the target audience's comprehension of cultural differences.

Example 7:

Source Text: 下礼拜要送九件那个石渠宝笈展那个东西要来小修 [.....] 突然打了鸡血了。

Target Text: We are expecting 9 pieces next week for minor repairs and then sent for the Precious Collection of the Stone Moat Exhibition, [...] it makes you feel like getting stimulants injected.

This example contains two culturally loaded expressions. Such a considerable translation challenge requires the translator to fully comprehend the source text and strive to retain the original cultural connotation as much as possible, otherwise the dissemination of excellent traditional Chinese culture might be seriously undermined. " 石 渠 宝 笈 " is a comprehensive catalogue from the Qianlong and Jiaqing reigns of the Qing Dynasty, recording the essence of paintings and calligraphy collected in the Qing imperial palace and divided into nine categories. The large-scale exhibition of this ancient work held in the Forbidden City caused a great sensation [14]. Given that viewers could grasp most cultural information through visual images and soundtracks, the translator has omitted the detailed description to avoid redundancy in the audiovisual text and adopted literal translation to render it in a concise way, facilitating target audience's comprehension to a large extent.

The other expression “ 打了鸡血 ” is a typical Chinese slang. It does not literally mean injecting chicken blood; instead, it figuratively describes a state of intense excitement. From a cultural perspective, the translator adopts free translation to provide an appropriate interpretation in an effort to help target audience fully understand cultural connotation in this multimodal discourse.

Drawing on Christopher Taylor’s multimodal text analytical framework, this study offers a detailed analysis of the subtitle translation of Masters in the Forbidden City and finds that the multimodal nature of screen texts can be effectively exploited through this approach. The translator has taken narrative, linguistic, semiotic as well as cultural elements into consideration and flexibly adopted various translation methods such as free translation, omission and specification to enhance the multimodal cohesion and avoid multimodal redundancy. In this case, multimodal features of the original documentary have been rendered in a faithful and comprehensible version for the target audience, demonstrating the value of Taylor’s framework in guiding the full exercise of translators’ subjectivity.

4. Conclusion

This study investigates multimodal text analysis in the process of subtitle translation of a successful Chinese cultural documentary. As an archetypal multimodal discourse, documentary mainly transmits textual and communicative information through verbal, visual and auditory modes and these semiotic modes contribute specific meanings to the meaning-making of the whole documentary text. The flexible application of Christopher Taylor’s multimodal text analytical framework offers an innovative way to perceive subtitle translation in AVT study, which illustrates the significance of dynamic interplay among different modes as well as the necessity for AV translators to go beyond purely linguistic-based translation. Guided by this effective theory, faithful narrative style, linguistic and multimodal cohesion, multiple semiotic resources and accurate cultural spirit could all be reproduced through appropriate translation strategies. Given that AVT research, especially that on documentaries, and multimodality study have not yet been fully integrated, it is therefore hoped that further research could be conducted to achieve a more comprehensive and extensive understanding of multimodality.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Fu, X., & Feng, R. (2022). A Multimodal Approach to the Subtitle Translation of Cruella. International Journal of Frontiers in Sociology, 4(5), 19-24.

[2] Franco, E.P.C. (2001). Voiced-over television documentaries: Terminological and conceptual issues for their research. Target, International Journal of Translation Studies, 13(2), 289-304.

[3] Kress, G., & Van Leeuwen, T. (2001). Multimodal Discourse: The Modes and Media of Contemporary Communication, Arnold.

[4] Baldry, A., & Thibault, P. (2006). Multimodal Transcription and Text Analysis. Equinox.

[5] Taylor, C. (2016). The multimodal approach in audiovisual translation. Target. International Journal of Translation Studies, 28(2), 222-236.

[6] Pettit, Z. (2013). Language and image: Translating the audiovisual text in subtitling and dubbing. In Readings in intersemiosis and multimedia (pp. 213–232). Ibis Edizione.

[7] Remael, A. (2008). Screenwriting, scripted and unscripted language: What do subtitlers need to know? In The didactics of audiovisual translation (pp. 57–67). John Benjamins Publishing Company.

[8] Vandaele, J. (2019). Narratology and audiovisual translation. In The Routledge Handbook of Audiovisual Translation (pp. 225-241). Routledge.

[9] Wen, H., & Wan, T. (2021). Adaptation and selection in subtitle translation of Masters in the Forbidden City from the

eco-translatological perspective. Journal of Nanchang Hangkong University (Social Sciences), 23(2), 93-98.

[10] Díaz-Cintas, J., & Remael, A. (2014). Audiovisual Translation: Subtitling. Routledge.

[11] Zhang, S. J., & Zhang, B. R. (2001). Philosophical reflections on hypotaxis and parataxis. Chinese Translators Journal, 22(4), 13–18.

[12] Fryer, L., & Freeman, J. (2012). Presence in those with and without sight: implications for virtual reality and audio description. Journal of CyberTherapy & Rehabilitation, 5(1), 15-23.

[13] Pedersen, J. (2011). Subtitling Norms for Television: An Exploration Focusing on Extralinguistic Cultural References. John Benjamins Publishing Company.

[14] Liu B, 2021. New forms of documentary development in the context of Internet: A case study of Masters in the Forbidden City, China Radio & TV Academic Journal, (8), 77-79.