International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 25, No. 1, 2026

A Study on the Subtitling of the Animated Film Mulan from the Perspective of Multimodal Discourse Analysis

Kexin Zhu[a], Yajun Qi[b]

Institute of Humanities, Huzhou College, Huzhou 313000, China
[a]2392611916@qq.com, [b]jsszqyj@163.com

Abstract: Prevailing audiovisual subtitle research prioritizes verbal comparison while neglecting the multimodal interplay of images and sound, a gap particularly evident in cross-cultural animation studies lacking systematic frameworks. Adopting Zhang Delu's four-tier multimodal discourse analysis framework-cultural, contextual, semantic, and expressive-this study examines the Chinese and English subtitles of Disney's Mulan (1998) through close analysis of dialogues and lyrics, exploring how subtitles coordinate with visual and auditory channels in cross-cultural meaning construction. Findings demonstrate that Zhang's framework offers robust explanatory power, and that multimodal strategies effectively reconcile linguistic, cultural, and audiovisual challenges, providing a replicable model and practical guidance for future research.

Keywords: Multimodal Discourse Analysis, Mulan, Subtitle Translation, Cross-cultural Communication, Modal Coordination.

1. Introduction

In the context of globalisation, film and television media have become central vehicles for cross-cultural communication; as a crucial link in the transmission of meaning in audiovisual texts, the quality of subtitle translation directly influences the effectiveness of cultural communication. The 1998 Disney animated feature film Mulan, adapted from the classic Chinese narrative The Ballad of Mulan, reconstructs Eastern cultural symbols from a Western creative perspective, forming a hybrid textual characterisation of Chinese elements + Western values. In this context, the translation of Chinese subtitles must not only accomplish conversion at the linguistic level but also achieve synergistic adaptation across different cultural contexts and multimodal signs. The theory of multimodal discourse analysis breaks through the limitations of text-centrism, pointing out that film and television works are systems of meaning constructed through the synergy of multimodal signs such as language, images, and sound (Hu Zhuanglin, 2007: 1). Zhu Yongsheng (2007: 82) clearly defines the concept of multimodal discourse, emphasising that its core lies in the diversity of modalities and sign systems; subtitling must therefore balance linguistic accuracy with the unity of multimodal meaning. The comprehensive framework for multimodal discourse analysis developed by Zhang Delu (2009: 24), which encompasses four dimensions-culture, context, content, and expression-provides a practical theoretical approach for the analysis of cross-cultural film and television subtitles. Existing research has largely focused on translation analysis at a single modality or purely linguistic level, neglecting the synergistic effects of multimodal signs. Baseon Zhang Delu's four-level framework, this paper takes the Chinese subtitles of Mulan as its subject of study. By examining specific translation examples, it explores the strategies for cross-cultural adaptation and the logic of multimodal meaning reconstruction, with the aim of enriching the practical dimension of multimodal translation research and providing a reference for similar cross-cultural film and television subtitling projects.

2. Theoretical Framework

2.1. Theoretical Core of Multimodal Discourse Analysis

Multimodal discourse analysis originates from Halliday's systemic functional linguistics, the core of which lies in revealing the synergistic mechanisms by which multimodal signs-such as language, images, sound, and movement-contribute to the construction of meaning. The three macro-functions of language proposed by Halliday (1994)-conceptual, interpersonal, and organizational-provide a foundational framework for modal analysis; Kress and van Leeuwen (1996), in Reading Images: The Grammar of Visual Design, put forward the theory of visual grammar, extending the three macro-functions to the visual modality and thereby breaking through the limitations of logocentrism. Building upon systemic functional linguistics, Zhang Delu (2009) constructed a comprehensive framework for multimodal discourse analysis encompassing four levels-culture, context, content, and expression-and clarified the decisive role of ideology and genre in modal selection. This framework provides comprehensive coverage ranging from macro-level culture to micro-level expression, wherein the cultural level serves as the foundation of meaning, the contextual level represents the concrete instantiation of culture, the content level constitutes the core of meaning, and the expressive level acts as the material vehicle for meaning. These four levels build upon one another and interact synergistically, providing a systematic tool for the analysis of cross-cultural film and television subtitle. translation.

2.2. The Need for Multimodal Adaptation in Subtitle Translation

As an integral part of audiovisual texts, the translation of film and television subtitles is not merely a linguistic conversion, but a process of synergistic adaptation involving multimodal signs. Wang Hongfang and Qiao Mengqi (2018: 95) point out that film and television subtitles must achieve a deep integration of linguistic signs with the audiovisual medium; Yang Zengcheng (2022: 105) emphasises that

multimodal text-composition mechanisms can ensure the synergistic expression of linguistic and non-linguistic modalities. High-quality subtitle translation must balance linguistic accuracy, cultural adaptability, and multimodal coordination, thereby achieving alignment between the meaning of the source language and the cognitive understanding of the target audience, and completing the reconstruction of multimodal meaning.

3. Multimodal Discourse Analysis of the Subtitle Translation for Mulan

Whilst Disney's Mulan reimagines the traditional image of Mulan, it retains her core cultural essence of filial piety and patriotism; consequently, the translation of its subtitles must, in addition to linguistic conversion, achieve cultural adaptation and multimodal synergy. Drawing on Zhang Delu's four-level framework and incorporating specific case studies, this paper examines the film's translation strategies and adaptation logic.

3.1. Cultural Level

Zhang Delu (2009) points out that subtitle translation must achieve coordination between linguistic and non-linguistic modalities at the cultural level to convey complete multimodal meaning. As a Western production on a Chinese theme, the subtitles for Mulan must be adapted to the cultural perceptions of Chinese audiences, resolving cultural differences through the equivalent translation of cultural symbols and metaphors.

Example 1 (47:29):

So carry on before I report you

干活儿去吧 不然我参你一本.

The source language phrase report you is an expression of denunciation within an English hierarchical context; the target language phrase 参你一本 aligns with Chinese feudal bureaucratic culture and echoes the source language's logic of intimidation. 干活儿去吧 conforms to Chinese colloquial usage and creates a contrast in tone with 参你一本, fitting the character's feigned anger and achieving dual alignment with both culture and context.

Example 2 (38:47):

Tranquil as a forest but on fire within

其徐如林 侵略如火.

The source text uses natural imagery to embody Western military ethos, whilst the target text employs a classic phrase from The Art of War by Sun Tzu, aligning with Chinese military culture and the story's local setting. This not only conveys the core meaning but also reinforces cultural appropriateness, achieving an equivalent translation of cultural symbols.

Example 3 (6:22):

We are gonna turn this sow's ear into a silk purse

我会把这只灰色的麻雀打扮成金凤凰.

The source text is a classic English idiom that uses sow's ear and silk purse to construct a metaphor for turning the crude into the refined; the target text employs the equivalent Chinese metaphors of grey sparrow and golden phoenix, which align with Chinese cultural perceptions, retain the source text's confident, tongue-in-cheek tone, and avoid misunderstandings (Liu Chengping, 2008).

3.2. Contextual Level

Situational context is the manifestation of latent cultural meaning in specific communicative situations; it is constituted by the scope, tone, and manner of discourse. Subtitle translation must precisely adapt to the social conventions, role relationships, and immediate demands of the specific situation (Zhang Delu, 2009). The subtitles for Mulan achieve synergy between linguistic modality and situational context through strategies such as adjustments to forms of address and adaptation of tone.

Example 1 (15:29):

Please, sir, my father has already fought bravely...

官爷，求求你，我爹他为国家效力了一辈子……

The source language is a polite request from a person in a position of weakness to an authority figure. The target term 官爷 aligns with the respectful title used by commoners towards officials in ancient China; 求求你 reinforces the earnest tone; and 为国家效力了一辈子—compared to the source expression-better aligns with the Chinese cultural context's emphasis on sacrifice, thereby enhancing the persuasiveness of the request and adapting to the social conventions and status disparity inherent in the conscription scene.

Example 2 (28:36):

I'm travel-size for your convenience

我缩小了，给你行个方便.

The source language uses the colloquial analogy of travel-size to address the question about his stature; as there is no equivalent concept in Chinese, the translation 缩小了 directly addresses the query, whilst 给你行个方便 conveys a gesture of goodwill. This aligns with the dynamic between Mulan and Mushu during their first encounter and complements the visual motif of Mushu's miniature form.

Example 3 (33:09):

Soldiers!

弟兄们!

In the source language, this is a collective term for soldiers in a military camp context; the target language phrase 弟兄们 aligns with the interpersonal dynamics of Chinese military camps, reflecting both the superior-subordinate relationship and the bond between comrades-in-arms. It rapidly evokes a sense of group identity, suits the immediate scene of chaotic combat amongst soldiers, and achieves the same communicative purpose as the source language.

3.3. Content Level

The content level is central to multimodal meaning, linking the contextual and expressive levels. The key to its analysis lies in examining the synergy between subtitles and the visual and auditory modalities, clarifying the relationship between text and imagery, and jointly realising the meaning of the discourse (Zhu Yongsheng, 2007). The subtitles in Mulan achieve a complementary construction of multimodal meaning by accurately conveying semantics and aligning with the audiovisual modalities.

Example 1 (29:31):

Two, three. Break it down. Hup, two, three. And work it!

一 二，神气点，一 二 ，扭起来!

The source language contains three types of instructions-rhythm, state, and demeanour-and the target language precisely corresponds to the core semantics; 扭起来 aligns with the comical visual modality of Mulan imitating a man's gait, echoing the on-screen imagery. This relationship of textual and visual restatement simplifies the audience's information processing and avoids cognitive load (Chen

Yuping and Zhang Caihua, 2017: 105).

Example 2 (40:41):

Be a man! — With all the strength of a raging fire

男子汉！——满腔热血像那野火压境.

The source text contains dual meanings: an indication of identity and a description of a state. The target text's 男子汉 clearly indicates identity, whilst 野火压境 visualises fighting spirit. This works in tandem with the visual imagery of Mulan climbing the tall pillar as the sky gradually brightens, reinforcing the meaning of Mulan pushing beyond her limits and enhancing the emotional impact.

Example 3 (51:14):

The general.

将军的头盔

The surface meaning of the source text merely refers to the general, whilst its implied meaning relies on the visual metaphor of the firelight on the snowy plain and the supplementary, tragic auditory effects; the target text explicitly identifies the concrete symbol of the helmet, echoing the visual modality of the helmet in the scene. Through this plain expression, it carries forward the sombre mood of the auditory modality; via multimodal synergy, it conveys both the brutality of war and a sense of mourning.

3.4. Expressive Level

The expressive level, as the material vehicle of meaning, encompasses both linguistic and non-linguistic media; its core lies in achieving spatiotemporal synergy and semantic complementarity between linguistic and non-linguistic modalities (Zhang Delu, 2009). The subtitles for Mulan strike a balance between linguistic form and multimodal rhythm, achieving a deep integration of linguistic and audiovisual modalities.

Example 1 (48:10):

I want her paler than the moon, with eyes that shine like stars

我的她洁白似月亮，眼眸赛星光.

The source language uses a clear metaphor to visualise the ideal female figure; the target language retains this metaphorical strategy, presenting the translation in a concise and rhythmic manner. This resonates with the elegant atmosphere of the Chinese painting and the visual modality of the full moon, whilst harmonising with the gentle auditory rhythm of the lyrics, collectively creating a dreamlike atmosphere.

Example 2 (31:50):

Number one in his class, extensive knowledge of training techniques, an impressive military lineage

武艺是班里的尖子生，熟知练兵之法，又是行伍世家出身.

The source text lists his strengths using parallel clauses; the target text retains this structure. The phrase 行伍世家出身 is more natural in Chinese than a literal rendering of military lineage, and aligns with the general's subtle visual expression of delight, striking a balance between objective information and implied emotion.

Example 3 (12:41):

Who is that girl I see?

倒影里的姑娘是谁？

The source language uses a question to convey Mulan's confusion; the target language retains the interrogative structure and the gentle rhythm, clarifying the reference to in the reflection. This works in tandem with the visual modality of Mulan gazing at her reflection on the lake and the low, soft auditory modality, fully conveying Mulan's self-doubt and confusion.

4. Discussion and Implications

Based on Zhang Delu's multimodal discourse analysis framework and drawing on the subtitling case study of Mulan, the following insights can be drawn at both theoretical and practical levels, providing guidance for cross-cultural film and television subtitling. At the cultural level, cross-cultural subtitling must achieve an equivalent translation of cultural symbols, using strategies such as the substitution of local symbols and the adaptation of metaphors to bridge cultural differences. The subtitles for Mulan did not employ the transplantation of cultural symbols; instead, they took into account the cognitive framework of the Chinese audience, transforming English slang and military terminology into proverbs and expressions from classical literature familiar to Chinese speakers. This approach not only preserved the core meaning of the source language but also enhanced cultural appropriateness, thereby confirming the feasibility of the global localisation concept in cross-cultural translation (Feng Wei, 2023). At the contextual level, situational elements are the core drivers of translation strategies; subtitle translation must closely align with scene conventions, character relationships, and immediate demands to achieve an equivalent transmission of communicative intent. Examples such as translating Soldiers as 弟兄们 (brothers) and Please, sir as 官爷，求求你 (Sir, I beg you) demonstrate that flexibly adjusting forms of address and phrasing can achieve synergy between linguistic modality and situational context, thereby helping the audience to quickly grasp the intended meaning of the discourse. At the content level, subtitle translation must focus on semantic transmission and multimodal coordination, constructing a complete semantic field by supplementing implied meanings and aligning with audiovisual modalities. Examples such as the translation of The general. as 将军的头盔 (the general's helmet) demonstrate that the semantic transmission of multimodal texts relies on modal synergy; subtitles must complement the visual and auditory modalities to simplify the audience's cognitive process and enhance the efficiency of meaning transmission (Chen Yuping and Zhang Caihua, 2017). At the level of expression, subtitle translation must adhere to the source language's expressive strategies whilst balancing linguistic rhythm with the on-screen atmosphere. In Mulan, both the lyrics and dialogue retain the core expressive forms of the source language, whilst adjusting details of expression to align with Chinese conventions and audiovisual modalities, thereby achieving expressive synergy between linguistic and non-linguistic modalities (Yang Zengcheng, 2022), which provides practical guidance for the translation of film and television subtitles of a similar nature.

5. Conclusion

Based on Zhang Delu's four-level framework for multimodal discourse analysis, this paper conducts a systematic analysis of the Chinese subtitles for the Disney animated film Mulan, finding that the subtitle translation achieves synergistic adaptation between the source and target languages across four levels: culture, context, content, and expression. At the cultural level, cultural differences are bridged through symbolic substitution and metaphoric translation; at the contextual level, expressions are adjusted

based on situational elements to suit the scenes and characters; at the content level, semantic transmission and modal coordination construct a complete field of meaning; at the expression level, the source language's form is followed whilst taking into account rhythm and atmosphere. This study validates the applicability and effectiveness of Zhang Delu's multimodal discourse analysis framework in the study of film and television subtitling, providing theoretical support and a practical paradigm for cross-cultural film and television subtitling. The study also has certain limitations: the case studies focused on core scenes and did not cover all discourse types; the research perspective was primarily text-based, lacking empirical data on the audience. Future research could expand the scope of study, integrate empirical research methods to enhance the persuasiveness of the conclusions; whilst also focusing on China's film and television outreach to explore strategies for the international dissemination of cultural symbols, thereby helping China's outstanding culture to "go global."

References

[1] Halliday, M. A. K. (1994). An introduction to functional grammar. Edward Arnold.

[2] Kress, G., & van Leeuwen, T. (1996). Reading images: The grammar of visual design. Routledge.

[3] Chen, Y. P., & Zhang, C. H. (2017). Chinese translation of English film subtitles: A multimodal analysis focusing on the relationship between text and image. China Translation, (5), 105–110.

[4] Feng, W. (2023). Global localisation and local globalisation: New trends in the cross-cultural communication of national images. Journal of Shanxi University (Philosophy and Social Sciences Edition), 46(3), 76–81.

[5] Hu, Z. L. (2007). Multimodality in the study of social semiotics. Language Teaching and Research, (1), 1–10.

[6] Liu, C. P. (2008). The image of Hua Mulan from a variationist perspective: Taking Tang Tingting's The female warrior as an example. Journal of Leshan Normal University, 23(7), 24–26, 41.

[7] Wang, H. F., & Qiao, M. Q. (2018). Audiovisual translation, multimedia translation and multimodal translation: Analysis and reflections. Research on Foreign Languages and Literatures, 4(6), 95–104.

[8] Yang, Z. C. (2022). Constructing a research paradigm for multimodal translation. Chinese Foreign Languages, (4), 105–111.

[9] Zhang, D. L. (2009). Exploring a comprehensive theoretical framework for multimodal discourse analysis. Chinese Foreign Languages, 6(1), 24–30.

[10] Zhu, Y. S. (2007). Theoretical foundations and research methods of multimodal discourse analysis. Journal of Foreign Languages, (5), 82–86.