Subtitling Translation of Medical Documentaries from the Perspective of Multimodal Discourse Analysis: A Case Study of *Baby Surgeons*

Anyue Zhou

Master's Student of Chongqing Medical University, Chongqing, China

Abstract

With the increasing cross-cultural dissemination of medical documentaries, their subtitling translation plays a pivotal role in promoting the popularization of professional knowledge and facilitating humanistic exchange. Grounded in the theory of multimodal discourse analysis, this study examines the subtitling translation of the medical documentary Baby Surgeons, exploring how medical information and narrative emotion are reconstructed and conveyed in translation through the synergistic interplay of visual, auditory, textual, and other multimodal semiotic resources. The research begins by outlining the core framework of multimodal discourse analysis, then proceeds to analyze representative excerpts from the documentary—including surgical scenes, doctor-patient interactions, and professional narration—to investigate how meaning is integrated across source and target languages within a multimodal context. Findings reveal that subtitling translation for medical documentaries must balance terminological accuracy, synchronization with narrative pacing, and cross-cultural adaptation of affective resonance. Through flexible application of strategies such as compensation, simplification, and explicitation, an effective equilibrium between medical scientificity and humanistic concern can be achieved. This study offers insights for the practice of medical audiovisual translation and contributes positively to the global dissemination of medical knowledge and its public comprehension.

Keywords

Multimodal Discourse Analysis; Medical Documentary; Subtitling Translation; Baby Surgeons; Cross-Cultural Dissemination.

1. Introduction

In the context of globalization, the dissemination of medical knowledge is increasingly transcending professional barriers and reaching the public through visual media. Medical documentaries, as an important vehicle for scientific documentation and humanistic storytelling, not only serve to popularize health knowledge and showcase medical advancements but also subtly shape public understanding of medical ethics, doctor-patient relationships, and perceptions of life. In recent years, the widespread popularity of domestic and international medical documentaries has highlighted the influence of this genre in cross-cultural communication. However, the construction of meaning in medical documentaries does not rely solely on verbal text; rather, it is accomplished through the complex interplay of multiple modalities, including visuals, live surgical scenes, sounds of medical instruments, medical imaging, voiceovers, character dialogues, subtitles, and even background music. This makes subtitling translation for such documentaries an exceptionally challenging task: translators must not only achieve linguistic accuracy but also accomplish the cross-cultural reconstruction of medical professionalism, narrative coherence, and emotional resonance within the collaborative constraints of a multimodal semiotic system.

While existing research has addressed the specialized nature of medical translation or the application of multimodal discourse analysis in audiovisual translation, systematic studies that integrate both perspectives to delve deeply into medical documentaries—a genre characterized by high information density, acute ethical sensitivity, and tightly interwoven multimodal collaboration—remain relatively scarce. Translating medical documentaries requires translators not only to overcome barriers posed by specialized terminology but also to possess a keen "modal awareness." This involves making judgments within limited temporal and spatial constraints: determining which information is directly conveyed through visuals, which requires supplementary explanation via subtitles, how to handle emotional tension and ethical implications in the original work, and ultimately achieving a balance between scientific accuracy and humanistic resonance in the target culture.

Therefore, this study aims to draw on the comprehensive theoretical framework of multimodal discourse analysis and apply it to the research on subtitling translation of medical documentaries. In doing so, it seeks to advance medical audiovisual translation in both theoretical and practical dimensions, guiding it toward a more scientific, humanistic, and cross-culturally adaptive direction.

2. An Overview of Multimodal Discourse Analysis Theory

The emergence of discourse analysis as an independent field of study can be traced back to the 1950s. In 1952, American structural linguist Z.S. Harris published the article "Discourse Analysis" in the journal *Language*, explicitly proposing this academic area for the first time and laying an important foundation for subsequent research. Thereafter, the study of signs and meaning gradually expanded beyond the realm of pure language. In 1977, French thinker Roland Barthes, in his work *Rhetoric of the Image*, delved into the semiotic systems within visual imagery and their mechanisms of meaning generation. His insights are later regarded as an early precursor to multimodal discourse analysis. By the 1990s, with the rapid evolution of digital media and information technology, the role of non-linguistic signs such as sound, image, and movement in meaning transmission became increasingly prominent. These elements intertwined and complemented verbal language, collectively participating in the construction of social meaning. Against this backdrop, multimodal discourse analysis gradually became a focal point at the intersection of linguistics, communication studies, and cultural studies. In 1996, Gunther Kress and Theo van Leeuwen, in their book *Reading Images: The Grammar of Visual Design*, grounded in the theoretical foundation of M.A.K. Halliday's systemic functional linguistics, formally introduced the term "multimodal discourse." They systematically explained how various semiotic resources work together in a rule-governed way to express meaning, thus marking the maturation of the theoretical framework of multimodal discourse analysis.

Entering the 21st century, domestic scholars in China have shown growing interest in multimodal discourse analysis, with related theoretical explorations gradually developing and becoming more systematic. In her 2003 work, *A Social Semiotic Analysis of Multimodal Discourse*, Li Zhanzi, adopting a social semiotic perspective, initially constructed a theoretical framework centered on visual analysis and highlighted the positive implications of this analytical approach for education and teaching. This research is regarded as the starting point for theoretical exploration in multimodal discourse analysis in China.

Subsequently, Zhu Yongsheng (2007), in his article "The Theoretical Foundation and Research Methods of Multimodal Discourse Analysis," further clarified the theoretical origins of this field. He explicitly stated that multimodal discourse analysis is rooted in systemic functional linguistics, systems theory, the metafunctional hypothesis, and register theory, while also pointing out three limitations in the research at the time, thereby charting a direction for deeper exploration in subsequent studies.Building upon this foundation, Zhang Delu (2009), in

Exploring a Comprehensive Theoretical Framework for Multimodal Discourse Analysis, established a well-structured and comprehensive theoretical framework grounded in systemic functional linguistics. This framework encompasses four fundamental levels: culture, context, content, and expression. The cultural level governs the traditions, forms, and technologies of communication; the contextual level determines the communicative genre within specific situations; the content level focuses on the formal systems of meaning construction; while the expression level pertains to the ways in which meaning is ultimately realized through material media. This framework provides an operational theoretical tool for the systematic analysis of multimodal discourse.

This study will primarily rely on the comprehensive framework proposed by Zhang Delu to analyze the subtitle translation practices of the documentary *Baby Surgeons* from the four dimensions of culture, context, content, and expression. It aims to provide reference for both research and practice in the subtitle translation of medical documentaries.

3. Introduction to Baby Surgeons

Baby Surgeons is a three-part medical documentary produced by the BBC in 2021. It authentically captures the intense and high-risk procedures performed by leading surgeons on fetuses still in the womb and newborn infants. The documentary focuses on the cutting-edge field of modern medicine—intrauterine fetal surgery—such as performing bypass surgery on a fetus's walnut-sized heart, using lasers to precisely block life-threatening tumor blood vessels through the mother's abdomen, or inserting a shunt into underdeveloped lungs. Through multiple specific cases, including families facing severe challenges such as congenital tumors, spina bifida, complex multiple pregnancies, and amniotic band syndrome, the documentary fully portrays the journey of doctors, parents, and fetuses striving together. It is not only a display of groundbreaking medical technology but also a profound emotional record, depicting the difficult decisions parents make in desperate situations, the mutual support between couples, and the professionalism and compassion of medical staff. The documentary blends the fragility and resilience of life, the grief of loss, and the hope of new beginnings. Directed by James W. Newton and meticulously produced, the documentary has deeply moved audiences with its authentic, direct, and humanistic narrative.

4. A Case Study of Chinese Subtitle Translation for the Documentary Baby Surgeons

4.1. Cultural Level

In cross-cultural communication, culture is not merely a backdrop but an intrinsic driver of meaning construction. As Zhang Delu (2009) notes, the cultural level forms the deep-seated foundation enabling multimodal communication, encompassing the ways of thinking, behavioral patterns, collective ideologies, and historical traditions of specific communities. This perspective resonates with the profound insight of sinologist John Deeney: "Language draws its vitality from cultural nourishment; thus, translation should not only involve the conversion of linguistic signs but also convey the intrinsic logic of thought and emotional expression between two cultures."In the practice of documentary subtitle translation, this implies that translators must keenly identify the cultural implicit information carried by the source language text and reconstruct it in a form that is natural, comprehensible, and acceptable within the target culture. This approach enables genuine meaning transmission and cultural dialogue in the context of intertwined audiovisual modalities (Yin Mingming, 2020).

Example 1: It's in the lap of the gods.

Translation: 现在只能听天由命。

The sentence "It's in the lap of the gods" is rooted in the context of Greek mythology, alluding to the idea that fate is controlled by the deities (according to legend, Zeus placed the lots of fate in an urn on his lap). This expression carries the collective memory of Western classical mythology, and its cultural imagery is characterized by personified, narrative-driven theological undertones. Since the original text references Greek mythology, a literal translation might confuse Chinese audiences. The translator therefore abandons a direct translation and opts for the idiom "听天由命" (ting tiān yóu mìng), which aligns with Chinese cultural cognition. Firstly, "听天由命" conveys a sense of passively accepting fate, corresponding to the original meaning of "leaving it to the gods to decide." Secondly, it avoids the need to explain foreign mythological references, making it more accessible for the audience.

Contextually and visually, this moment shows a doctor gazing with concern at the color ultrasound image of the fetus. The subtitle "听天由命" accentuates the tension and difficulty of the situation, effectively resonating with the scene's emotional weight.

4.2. Contextual Level

Within the multimodal discourse system, there exists a dynamic relationship of mutual construction between modes and context. Zhang Delu (2009) elaborates on this as an interactive model of active and passive modes. In composite texts such as films, symbols like imagery, sound, and text generally form active modal associations with the context—meaning that various signs actively participate in and reinforce the generation of contextual meaning. Consequently, during the process of subtitle translation, translators must fully consider and proactively engage contextual elements. This ensures that the textual translation organically aligns with the visual narrative and plot progression, thereby achieving intrinsic synergy and coherent meaning within the multimodal discourse system.

In the actual process of communication, contextual factors—including the field, tenor, and mode of discourse—profoundly constrain the expression and reception of meaning. Due to the transient nature of documentary subtitles, the coordination of multimodal elements, the selection and omission of translated information, and the logical coherence between narrative developments all directly influence the audience's interpretation of the content. Therefore, translators must carefully grasp the specific situational context, make necessary additions, deletions, and adjustments to information, and choose appropriate linguistic forms, thereby maintaining the coherence of subtitles and the fluidity of the viewing experience within limited temporal and spatial constraints.

Example 2: Zoe's case is incredibly rare, so it's one in about 20,000.

Translation: 佐伊的情况非常罕见，患病率只有两万分之一。

In the original text, "one in about 20,000" is a relatively colloquial expression in English. Considering that this is a subtitle translation for a medical documentary aimed at a general audience, a direct translation such as "两万分之一" might confuse viewers. By adding the term "患病率" (prevalence rate), the medical significance of this number is immediately clarified, aligning with the needs of science communication. The original phrase "incredibly rare" conveys a tone of emphasis within a professional context. The translation reinforces the emotional nuance with "非常罕见" (extremely rare) and highlights the low probability with the word "只有" (only). This approach maintains the objectivity of the narration while preserving its evaluative tone, fitting the documentary's stylistic feature of blending scientific education with storytelling.

Visually, the doctor is presenting Zoe's medical images and data, while the subtitles provide specific statistical figures—the two elements reinforce each other. The translator refrains from adding unnecessary explanations, maintaining the conciseness of the subtitles, which forms an effective complement to the visual information.

4.3. Content Level

The content level of multimodal discourse can be divided into two dimensions: meaning and form. The meaning dimension encompasses ideational, interpersonal, and textual meanings, all of which are constrained by factors such as the field, tenor, and mode of discourse. The form dimension is primarily composed of auditory modalities (e.g., language, music) and visual modalities (e.g., imagery, text) (Lü Jian et al., 2012). Therefore, when translating subtitles, translators should, based on a thorough understanding of the relationship between the source text and the target audience, integrate information from both auditory and visual modalities. Through appropriate supplementation or omission of information, they can achieve accurate transmission and effective reception of cross-modal meaning.

Example 3: All we're trying to do is nudge things from a place where parents' dreams and expectations are dashed into a place of hope.

Translation: 我们努力要做的就是，提醒父母们做好最坏打算，但同时也不放弃希望。

The spatial metaphor "from a place where... into a place of..." in the original text frames psychological states as movement between locations—an English cognitive pattern. The Chinese translation reinterprets this as "做好最坏打算…不放弃希望" ("prepare for the worst… without giving up hope"), aligning with Chinese conventions of linking inner states to attitudes when conveying emotional shifts.

The phrase "nudge things," reflecting the medical team's careful guidance, is rendered as "提醒…做好最坏打算" ("remind… to prepare for the worst"). This preserves professionalism while softening any sense of imposition, maintaining the trust-based tone typical of Chinese medical communication.

On screen, the doctor walks slowly toward the operating room—a visual metaphor for moving from despair toward hope. The subtitles synchronize with this imagery, reinforcing the emotional transition and aiding viewer comprehension.

4.4. Expression Level

At the expression level, documentaries rely primarily on auditory-verbal elements such as narration and interview dialogue to convey meaning. These are supported by nonverbal modalities, including explanatory subtitles, archival footage, animated graphics, and ambient sound. When translating subtitles, it is essential to account for the genre's commitment to authenticity and its characteristically high density of information. Within the strict temporal and spatial limits of subtitling, careful attention must be paid to typographic details—such as font size, on-screen duration, positioning, and layout—so that the text integrates seamlessly with the visual and auditory channels, ensuring coherent communication within the multimodal framework.

Example 4: It's amazing if you think that a sperm and an egg meet and out of those two single cells comes a perfect human being.
Translation: 想想看，一颗精子和一颗卵子偶遇，两个单细胞就能创造出一个完美的人，这真的很不可思议。

The original English sentence was relatively long and had to be divided into three segments for subtitling due to limited on-screen space. In translating, the translator restructured the sentence order: rendering "amazing" as "这真的很不可思议" and placing it at the end, while moving "think" forward with the natural Chinese phrase "想想看" to lead smoothly into what follows—a phrasing that reflects typical Chinese rhetorical flow.

Given the transient nature of subtitles, translation must quickly capture the audience's attention and emphasize key points. The original clause, "a sperm and an egg meet and out of those two single cells comes a perfect human being," presents a factual statement, while "amazing" conveys an evaluative attitude. By positioning the factual statement first, the translation helps the audience swiftly grasp the core message, achieving effective synergy between the audiovisual components and the text.

Operating within tight time and space limitations, the translator adapts sentence rhythm, employs expressive verbs, and rearranges information to ensure clarity for the viewer. This allows the subtitles to synchronize effectively with the visual and auditory elements while also evoking the documentary's central theme—the "miracle of life." Such an approach moves beyond literal translation, weaving together language, image, and sound into a cohesive communicative whole.

5. Conclusion

This study, grounded in the theoretical framework of Multimodal Discourse Analysis, systematically examines the practical features and theoretical implications of subtitling translation in the medical documentary *Baby Surgeons* from four dimensions: culture, context, content, and expression. At the cultural level, the translation of medical documentaries entails not only the conversion of terminology but also the cross-cultural adaptation of perspectives on medical culture, the human body, and the meaning of life.Translators must strike a balance between professional accuracy and public accessibility. This requires not only adhering to the discourse norms of the medical science community but also adapting Western-centric medical expressions into forms that resonate emotionally and cognitively with Chinese audiences. At the contextual level, subtitling translation is strictly constrained by the context. Translators must make deliberate choices regarding what to retain, omit, or restructure in the source text based on visual information, narrative pacing, and emotional progression.At the content level, translation must harmonize ideational, interpersonal, and textual meanings. The accurate transmission of medical information (ideational meaning) serves as the foundation, while the tone, stance, and emotion in doctor-patient communication (interpersonal meaning), along with the logical cohesion between subtitles and the audiovisual narrative (textual meaning), collectively shape the audience's reception. At the expression level, subtitles, as a linguistic mode, must actively interact with and complement non-verbal modes such as visual imagery and auditory elements. Translators need to consciously adjust sentence length, select appropriate verbs, and modulate tonal density so that the text resonates in sync with camera movement, surgical progression, musical atmosphere, and other multimodal cues.

Effective subtitling translation for medical documentaries necessitates not only the accurate conveyance of specialized information in medical texts but also the synchronized translation of procedural details, medical instruments, and imaging presented visually, as well as the medical ethics and emotional tension carried by auditory elements such as ambient sounds, tones in doctor-patient dialogues, and the pacing of narration. The theory of multimodal discourse analysis provides a systematic framework for interpreting such translation. However, in the highly specialized and culturally sensitive domain of medical communication, the integration of theory and practice remains an area for further exploration. It is hoped that future subtitling practices can achieve more refined and context-adapted approaches within medical

communication. This will better facilitate the cross-cultural dialogue between medical knowledge and humanistic care, thereby contributing to mutual understanding and empathy at the level of a global health community.

References

[1] KRESS G, LEEUWEN T V. Multimodality: A Social Semiotic Approach to Contemporary Communication [M]. London: Arnold, 2001.

[2] O'HALLORAN K L. Multimodal Discourse Analysis: Systemic Functional Perspectives [M]. London: Continuum, 2004.

[3] FENG Dezheng, ZHANG Delu, O'HALLORAN K L. Advances and Frontiers in Multimodal Discourse Analysis [J]. Contemporary Linguistics, 2014(1): 88–99.

[4] YIN Mingming. Research on Subtitle Translation Strategies of French Classic Art Films from the Perspective of Multimodal Discourse Analysis [J]. Contemporary Cinema, 2020(2): 164–168.

[5] ZHANG Delu. Exploring a Comprehensive Theoretical Framework for Multimodal Discourse Analysis [J]. Foreign Languages in China, 2009(1): 24–30.

[6] ZHU Yongsheng. The Theoretical Foundation and Research Methods of Multimodal Discourse Analysis [J]. Foreign Language Research, 2007(5): 82–86.