ARTICLE

On Audiovisual Translation from the Perspective of Multimodal Discourse Analysis: A Case Study of *Wolf Warrior 2*

Jing Ning [1] 🄙 , *Xiaoxin Mo* [2*]

[1] *School of Foreign Studies, Southern Medical University, Guangzhou 510515, China*

[2] *School of Laboratory Medicine and Biotechnology, Southern Medical University, Guangzhou 510515, China*

ABSTRACT

In today's globalized world, the exchange of audiovisual products across cultures has become increasingly frequent. This research, by utilizing a comprehensive framework of Multimodal Discourse Analysis (MDA), aims to examine the subtitle translation of the Chinese action film *Wolf Warrior 2* in the context of cross-cultural audiovisual translation by integrating linguistic components (words, grammar, semantics) with non-verbal components (facial expressions, body language, sound effects) to bridge cultural disparities. Subtitle translation, as an important means of audiovisual translation, not only needs to convey the literal meaning of the original text but also has to consider the integration with various non-verbal elements in the film. These non-verbal elements play a crucial role in expressing emotions, attitudes, and cultural connotations. Through a meticulous analysis of MDA principles and multimodal interactions in the film, this study puts forward effective translation strategies that take into account the cultural background of the target audience, the intended meaning of the original text, and the audiovisual context to enhance cross-cultural communication and foster a more profound global comprehension of the film's content. This research not only makes contributions to the field of audiovisual translation but also enriches the discourse on cross-cultural communication in the globalized media environment.

Keywords: Multimodal Discourse Analysis; Cross-Cultural Audiovisual Translation; Subtitle Translation, Technological Advancements; Audience Reception

*CORRESPONDING AUTHOR:

Xiaoxin Mo, School of Laboratory Medicine and Biotechnology, Southern Medical University, Guangzhou 510515, China;
Email: moxiaoxin_smu@163.com

1. Introduction

In the globalized era, the dissemination of films across different cultures has become increasingly prevalent, necessitating a deeper understanding of the complexities involved in audiovisual translation. Compared to conventional discussions on textual translation strategies of subtitles or film names, in recent years using multimodal pathway to introduce Chinese stories and enhance their dissemination and effectiveness has become a hot research focus for domestic scholars. The selection of *Wolf Warrior 2* as the subject of research holds significant importance for two principal reasons: firstly, its commercial success and international acclaim serve as a paradigm for the complexities and prospects inherent in the cross-cultural dissemination of cinematic works, rendering it an optimal case study to examine the translation of Chinese cultural narratives across various media; secondly, the film's intentional fusion of patriotic motifs with vivid visual and action-oriented sequences, along with emotionally charged soundtracks, offers a fertile terrain for the analysis of the interrelation between linguistic and non-linguistic components within the realm of subtitle translation. The success of *Wolf Warrior 2* is closely tied to its profound embodiment of the Chinese spirit of patriotism. It meticulously captures the heroic deeds of soldiers defending their homeland, while also depicting the unity and collective will of the nation.

The aim of this research is to substantiate the hypothesis that Multimodal Discourse Analysis (MDA) can systematically elucidate how subtitle translation strategies in the film *Wolf Warrior 2* amalgamate linguistic, visual, and auditory components to surmount cultural divides. MDA provides a comprehensive framework that integrates linguistic, visual, and auditory elements, offering a holistic approach to understanding how meaning is constructed and conveyed in audiovisual media. By scrutinizing the film's manifestation within various cultural contexts, through analytical lenses encompassing culture, context, content, and expression, this research identifies the impediments and efficacious methodologies inherent in cross-cultural audiovisual translation. It substantiates that the integration of multimodal elements enhances the precision of cultural message conveyance and fosters audience engagement.

2. Literature Review

The term "audiovisual translation" evolves with technological advancements and can be categorized into film translation, screen translation, audiovisual translation, and multi-media translation[1]. The 2009 edition of the *Routledge Encyclopedia of Translation Studies* merged subtitling, revoicing, and dubbing under the entry of audiovisual translation. The classification of these key terms in a certain degree reflects the four stages of AVT studies: the birth of AVT (film translation), the expansion of AVT (screen translation), the maturation of AVT (audiovisual translation), and the renewal of AVT (multimedia translation)[2]. In this study, "audiovisual translation" (AVT) is adopted as the umbrella term for subtitle translation.

2.1. Audiovisual Translation and Subtitling

International research on AVT has a long history, with early studies focusing on film dubbing and subtitling. Delabastita argued that the translation of mass media, film, and television serves as evidence of cultural interaction within a specific social and cultural context[3]. Gottlieb provided a comprehensive overview of subtitling, discussing its definition, process, types, and challenges[4]. Since then, research has expanded to various aspects of AVT, such as quality assessment, audience reception, and the impact of technology on translation practices. Díaz-Cintas combined multimodality with AVT, identifying four key elements that define audiovisual texts: audio-language, audio-non-verbal, visual-nonverbal, and visual-language[5]. Pérez-González discussed the advantages and limitations of automated subtitling, arguing that while it can increase productivity and reduce costs, it also requires careful post-editing to ensure quality and accuracy[6]. Chaume posited four principal transformations within the realm of film and television translation studies: a technological transformation (the utilization of AI tools), an audience transformation (research into cognitive reception), a cultural transformation (analysis of power discourse), and an industrial transformation (the emerging demands of streaming platforms)[7]. These transformations offer methodological frameworks for contemporary empirical research, including eye-tracking studies. In recent times, a growing cohort of international scholars has concentrated on machine models, such as the creation of multimodal large-scale models, the employment of machine models for the integration

of visual signal analysis, and the modification of subtitle tonality in accordance with the ambiance of the scene.

Domestic research on audiovisual translation has primarily focused on the introduction and reflection of western and domestic AVT research developments[8–12]. This research initiative has constructed theoretical paradigms through the incorporation of international theoretical perspectives, thereby promoting discourse between domestic and international Audiovisual Translation (AVT) academic circles. Nevertheless, it has faced criticism for an excessive dependence on foreign theories, which may lead to an oversight of the requisite contextualized adjustments pertinent to Chinese media practices. Some other scholars delved into the subtitle translation strategies applied to specific film or dramas[13–17]. In recent years, the demand for subtitle translation is increasing rapidly against the backdrop of the Belt and Road Initiative. How to quickly translate and produce a large number of film and television works within a limited time has become an urgent problem to be solved. Artificial intelligence-empowered translation technology has provided basic and comprehensive technical support for subtitle translation, and the importance and urgency of subtitle translation technology research are increasingly prominent[18]. Li and Zhang proposed an improvement plan under the perspective of digital humanities: to perfect the theoretical system based on bibliometric tools and corpus methods, to build large-scale and high-quality databases to enhance standardization, to conduct multimodal and semiotic research on audiovisual translation combined with discourse analysis, to track the dissemination paths of audiovisual translation products with the help of geographic information system technology, and to study the reception through computer programming such as Python and R[19]. This interdisciplinary methodology surpasses conventional qualitative research by incorporating quantitative analysis and spatiotemporal tracking into Audiovisual Translation (AVT) studies. It has precipitated a paradigmatic shift towards empirical research that is driven by data. Apart from theoretical discussion, practical translation strategies and digital humanities, the topic of multimodal translation and the international dissemination of Chinese culture is gaining momentum. Wu and Mou investigated how to enhance the international dissemination of Chinese stories through multimodal translation strategies, suggesting shifting from monomodal to multimodal translation to strengthen the shaping power and appeal of Chinese stories[20]. Zhang and Kang conducted an analysis of contemporary multimodal translation methodologies, underscoring the interplay between linguistic and non-linguistic symbols. Utilizing Chinese martial arts communication as a case in point, their research illuminated the role of non-linguistic elements as conduits for cross-cultural comprehension. The study demonstrated that multimodal translation adeptly transmits implicit cultural connotations and functions as a pragmatic paradigm for augmenting national cultural soft power[21].

2.2. Multimodal Discourse Analysis

Multimodal discourse analysis (MDA) originated in the 1990s and has gradually developed with the interdisciplinary integration of linguistics, especially with fields such as cognitive science, communication studies, and semiotics, focusing on how different modes of communication interact to construct meaning. Kress and Van Leeuwen laid the foundation for MDA by developing a social semiotic approach to the analysis of visual images. Since then, MDA has been applied to various contexts, including advertising, education, and media studies. In the field of translation studies, MDA has provided new insights into the complexities of audiovisual and subtitling translations. O' Halloran argued that MDA can help translators better understand the multimodal nature of audiovisual texts and the interactions between different modes of communication[22]. Similarly, Baldry and Thibault proposed a framework for analyzing multimodal transcripts of audiovisual texts, emphasizing the importance of considering the temporal and spatial organization of different modes[23]. Recent studies have employed MDA to investigate the translation of multimodal features in audiovisual texts, such as the interplay between language and images, the use of sound effects and music, and the role of non-verbal cues in communication. For example, Pettit argued that the meaning of multimodal texts is not expressed in a single channel, and translators should consider the interplay between different modes when translating[24]. She proposed a framework for analyzing the multimodal aspects of audiovisual texts and their implications for translation. As times change, technology drives theoretical innovation. Multimodal discourse analysis is also beginning to explore artificial intelligence. Norris proposed a method framework based on multimodal discourse analysis for analyzing com-

munication strategies in video conferences[25]. Multimodal large language models (MM-LLMs) can process not only text but also image, video, and audio information. This opens up possibilities for applications such as text-to-video generation, image description, and text-to-speech.

Domestic research has also recognized the value of integrating audiovisual translation with MDA. Many researchers analyzed the multimodal features of audiovisual texts, including films, television series, and documentaries by applying MDA to the translation practice. For example, Lv and Wu analyzed the subtitle translations of *The Flowers of War* from a multimodal perspective, examining how the translators handled the interactions between the spoken dialogue, visual images, and sound effects[26]. Liang explored the translation of English subtitles in American and British TV dramas from the perspective of multimodal discourse analysis[27]. Yin researched on the translation strategies of three French classic art films under the perspective of multimodal theory in order to provide references for Sino-French film and television cooperation[28]. In recent years, domestic scholars shifted to the study of audiovisual translation in the aspect of cultural dissemination to promote the spread of Chinese culture overseas. Wu examined the multimodal translation strategies used in the documentary *Wild China*, focusing on how these strategies shape China's national image: BBC constructs a negative image of China through extension, selection, exclusion, emphasis, and elaboration, while CCTV reshapes a more positive image through addition, condensation, omission, and compilation[29]. Wu and Li used ELAN software to analyze the Chinese film *My People, My Country*, proposed strategies such as addition, deletion, substitution, and conversion to optimize subtitle translation, and highlighted the importance of flexible selection, accurate output, dynamic equivalence, efficient delivery, and instant reflection in enhancing the transmission of Chinese film and television culture[30]. Niu examined the translation strategies employed at cultural, contextual, content, and expressive levels of the film *Three Thousand Miles of Chang'an*, emphasizing the importance of balancing cultural fidelity with audience accessibility[31].

3. Framework of Multimodal Discourse Analysis

The theoretical foundation of multimodal discourse analysis is deeply rooted in the pioneering work of the British linguist M.A.K. Halliday. The theory of Systemic Functional Linguistics (SFL) proposed by Halliday emphasizes that language is a tool for achieving communicative purposes and constructing meaning. Language is not merely a symbolic system but serves functional and social roles in interactions. SFL focuses on the relationship between the function, structure, and meaning of language, emphasizing how language is used to express and achieve different communicative purposes. Thus, within his framework of SFL, Halliday describes language as "social semiotics". He posits that besides verbal language, other systems of signs also serve as carriers of meaning[32]. Scholars like Kress and Van Leeuwen further developed this concept based on Halliday's fundamental principles, particularly in the realm of visual communication[33–35]. They meticulously crafted a comprehensive visual grammar that resonates with Halliday's tripartite functions and incorporates insights from film studies. The theory of visual grammar provides an essential theoretical framework and methodology for studying images and visual communication. The grammatical structure, symbolic systems, and ways of constructing meaning of visual elements aid in better understanding and interpreting the role and significance of images in communication and expression, thus offering a critical theoretical foundation for multimodal communication and visual design. With the development of the times, the theoretical core has been deepened and expanded. Norris proposed Interactional Discourse Analysis, which broke through static text analysis for the first time and introduced the concept of Action Hierarchy, incorporating dynamic modalities such as gestures, postures, and spatial movements into the three meta-functional frameworks, with particular emphasis on the coordinating role of interpersonal functions in real-time interaction[25]. Furthermore, Halliday's framework continues to expand in breadth, with new modalities such as music and digital media being incorporated into the theory. Moskvil proposed the concept of meaning construction in music[36]. For example, rhythm functions as an independent subsystem, regulating emotional intensity through beat and tempo, and reinforcing narrative tension through the synchronization of sound and image rhythm in films.

Drawing upon the theoretical framework of systemic functional linguistics proposed by Halliday and further investigating relevant studies, early theoretical exploration of MDA in China can be found in the paper of Li, who compared the visual grammar with Hallidayan system-functional grammar, and made the theoretical introduction to representation meaning, interaction meaning, and composition meaning, with pictures shown as examples[37]. Zhu concluded five kinds of modality senses and pointed out that the closest modalities are visual modality and acoustic modality[38]. Zhang proposed a "multimodal discourse analysis comprehensive framework"[39]. Accordingly, multimodal discourse utilizes a series of sensory perceptions of the human body, such as auditory and visual, and then interacts with the resources of language, sound, and other modalities. The framework is divided into four main dimensions and their subcategories: cultural level, contextual level, content level (further including discourse meaning and form level), and expression level. Through this hierarchical structure, a more comprehensive understanding and analysis of the use and meaning of multimodal discourse can be achieved, gaining insights into how various modes of communication interact in different contexts and facilitate the conveyance of meaning. This approach underscores the complexity and richness of communication, highlighting the importance of considering multiple modes beyond traditional linguistic analysis to fully comprehend the subtle differences and dynamic changes in discourse in contemporary environments. Over time, domestic scholars have further developed and restructured the core framework. Zhang refined the original "complementary/non-complementary relationship" in the framework[39]. Feng first established a three-tier classification system for modal relationships[40], enhancing the explanatory power of cross-media communication of cultural symbols: the complementary relationship added a "reinforcement hierarchy," such as the "primary-secondary reinforcement" formed by artifact images and explanatory text. The non-complementary relationship introduced "cross-modal metaphorical mapping,"such as the "abstract-concrete" mapping formed by bronze ware decorations and English explanations. This paper still employs Zhang's classical theory for subtitle analysis for two main reasons. First, the classical theory more systematically and comprehensively caters to the multidimensional requirements of subtitle translation.

Its analytical framework, which covers cultural, contextual, content, and expressive layers, enables a thorough examination of how subtitles balance cross-cultural communication, situational adaptation, information fidelity, and formal constraints. Second, the classical theory has been more thoroughly validated in practical applications. In contrast, emerging theories may overlook the unique characteristics of subtitle translation—due to their untested applicability or focus on abstract constructs rather than operational details. This makes the classical theory a more reliable and comprehensive tool for addressing the subtitle translation.

4. Case Analysis

The 2017 film *Wolf Warrior 2* tells the story of Chinese soldier Leng Feng experiencing various challenges while carrying out a rescue mission in a certain country in Africa. With intense battle scenes, high difficulty stunts, and a profound exploration of humanity and justice, the film offers an engaging exploration of patriotism and heroism. The success of *Wolf Warrior 2* can fundamentally be seen as a successful amalgamation of East and West. In case analysis, this study delves into various aspects such as culture, context, expression methods, and the interaction between images and language, exploring how the translator integrate linguistic and non-linguistic elements for meaning construction and enrich the viewing experience by bridging cultural gaps.

4.1. At Cultural Level

Translation is not merely a process of converting one language into another; it is a matter of achieving cross-cultural communication and dissemination. Throughout the entire translation process, translators must consider the inherent expressions, customs, and values of different cultural backgrounds to ensure that the translated work accurately conveys the original information and its emotional resonance. The cultural dimension is crucial to achieving multimodal discourse analysis. To achieve the goals and functions of subtitle translation, translators must not only adhere to relevant translation principles and adopt specific translation strategies but also closely monitor cultural phenomena during the translation process. Effective translation requires a deep cultural resonance and understanding, not just superficial words. It demands a meticulous comprehension of the subtleties that

define and differentiate cultures. This approach ensures that translation conveys not only the words but also the essence and context that give meaning to the words, thereby fostering a deeper cross-cultural dialogue and understanding.

Example 1:

ST: 十二年卧薪尝胆。(Twelve years of hard work and perseverance)

TT: 12 years of hard work.

The Chinese culture has a long and profound history, with idioms playing a crucial role. They encapsulate rich philosophies and cultural meanings, offering deeper insights into cultural values and thus protecting and promoting traditions. The idiom "十二年卧薪尝胆 (Twelve years of hard work and perseverance)" originates from an ancient Chinese story. The story tells of King Goujian of Yue who was defeated by the state of Wu. He decided to endure humiliation and work towards revenge by sleeping on straw and frequently tasting bitter gall, constantly reminding himself of the sufferings he endured. This idiom symbolizes profound endurance and the importance of sticking to one's original intention, indicating the willingness to endure long-term hardships and difficulties to achieve one's goals. From a cultural perspective, translating the Chinese idiom "十二年卧薪尝胆 (Twelve years of hard work and perseverance)" as "12 years of hard work", which, although simplified, captures the core idea of "diligence and perseverance" in the original sentence and avoids misunderstandings caused by cultural differences. This illustrates that in a multimodal context, translators need to balance cultural connotations and audience comprehension, choosing appropriate translation strategies to achieve effective cross-cultural communication.

This quote is from a Chinese businessman during a Chinese evacuation operation, looking at the country with fighting scenes in front of him, coincidentally an image of a body floating on the river surface. At this special moment, the businessman and the protagonist are both leaning forward, leaning on the railing of the warship. This common posture clearly indicates that they are both focused on listening to each other's conversation. This moment captures the essence of sincere and genuine communication between the two characters. The background of this scene is filled with the hustle and bustle of evacuation, adding a sense of urgency and chaos. In this noisy background, the plight of these people - after enduring twelve years of relentless work on unfamiliar land, finding themselves impoverished and struggling - becomes even more tragic and poignant. Such scenes and sounds complement the subtitle translation of "12 years of hard work", enhancing the audience's ability to grasp the profound meaning conveyed in the dialogue. The stunning visual effects, combined with the bewildered expression of the businessman and the subtitles, allow the audience to immediately understand the devastating impact of war - how it erases all efforts.

MDA focuses on how different modes (such as language, images, sounds, etc.) jointly construct meaning. In the text, it is mentioned that in a scene, the subtitle translation "12 years of hard work" is combined with visual elements (such as the characters' postures, expressions, and the chaotic evacuation background) and auditory elements (such as background music) to convey the profound theme of how war destroys people's efforts. This synergy between multimodal elements allows the audience to more comprehensively and profoundly understand the cultural connotations and emotional significance embedded in the discourse. It demonstrates that in a multimodal context, cultural translation needs to consider how to interact with visual, auditory, and other modal elements to create a complete and emotionally engaging narrative space.

Example 2:

ST: 血债血偿。(An eye for an eye, a tooth for a tooth)

TT: Blood for blood.

In the film, the protagonist Leng Feng discovers that the mercenary veteran who has had multiple conflicts with him is actually the mastermind behind the kidnapping of his wife. Filled with anger, Leng Feng confronts the mercenary and demands "an eye for an eye", leading to an intense hand-to-hand combat. In the case of "血债血偿" (An eye for an eye, a tooth for a tooth), the original phrase corresponds to the Western cultural reference from the Bible, "an eye for an eye, a tooth for a tooth." This translation leverages the cultural symbolism from the Bible to resonate with Western audiences. It shows that in multimodal discourse, the use of cultural symbols (such as allusions) can enhance the persuasiveness and emotional impact of the message. Translators find similar or corresponding cultural symbols between different cultures to facilitate communication in a cross-cultural context.

By combining subtitle translation with cultural back-

ground, the film facilitates deeper connections and resonances among Western audiences, making the dialogue more friendly and engaging. In this particular scene, Leng Feng raises a bullet high, displaying an expression of anger and ferocity. This powerful image not only conveys his unrestrained anger but also highlights his deep personal animosity towards the mercenaries. The tense and thrilling background music further emphasizes the high tension of the scene and serves to accentuate the narrative climax. The combination of these elements effectively enhances the audience's emotional involvement, pushing their emotions towards a strong climax. It allows the audience to deeply understand the spirit of "Blood for blood," immersing them in the irreconcilable hatred between the protagonist and the mercenary.

Example 3:

ST: I want you to round' em up like cattle.

TT: 把他们当作饺子给我包起来. (Wrap them up like dumplings for me)

This sentence was said by the leader of the mercenaries when ordering the troops to surround the factory hostages. In this picture, the mercenaries are arranged facing down with tasks. At this moment, his relaxed expression and humorous words further demonstrate his determination to win and confidence in his abilities. The translation of "I want you to round' em up like cattle" as "pack them up like dumplings for me" demonstrates the wonderful interaction between cultural context and multimodal discourse analysis.

From a cultural perspective, the original phrase "I want you to round' em up like cattle" evokes the image of herding, a metaphor deeply rooted in Western culture and historical background, especially in contexts related to ranches and cowboy culture. This metaphor implies a dehumanizing way of control, treating the hostages simply as livestock to be rounded up. The Chinese translation "Wrap them up like dumplings for me" carries significant Chinese cultural colors. In China, dumplings are a familiar food with important cultural significance. The action of wrapping dumplings is meticulous and thorough, implying a complete surround and escape-proof situation for the hostages. This translation not only preserves the dehumanizing implications of the original text but also adds a layer of cultural resonance for Chinese readers. Therefore, the translation "Wrap them up like dumplings for me" is a subtle interpretation of the cultural differences and appropriate contexts of the original sen-

tence. It maintains the dehumanizing tone of the original text through cultural metaphors, while enhancing the audience's connection to the scene through familiar cultural imagery. The multimodal elements of visual cues further enhance the effectiveness of the translation, making it a convincing and effective adaptation. This indicates that in multimodal discourse analysis, the transformation of cultural symbols is an essential means of cross-cultural translation. Translators need to have a deep understanding of both cultures and find suitable symbols to substitute or complement, in order to achieve the best communication effect.

4.2. At Contextual Level

In the field of discourse, the dimension of context primarily includes the scope of discourse, the tone of discourse, the manner of discourse, and the structural framework that constitutes the context (Zhang 2009). Zhang believes that communication in a specific context is inevitably influenced by contextual factors. In the context of films, subtitles and the surrounding context play a crucial role in advancing the plot. The context determines the translation of subtitles. From the perspective of context, subtitles should serve as a link between the preceding and the subsequent, avoiding the situation of taking sayings out of context, thus ensuring the natural and gradual development of the film narrative. Therefore, translation at the contextual level needs to consider how to interact with other elements in the multimodal context to create a unified and coherent narrative structure, enabling the audience to smoothly switch and understand between different modes for a better viewing experience.

Example 4:

ST: That could be the sound of the future.

TT: 那也有可能是你的下场. (That could also be your fate)

MDA posits that the meaning of discourse is determined by its context. The text points out that in the context of film, subtitle translation needs to consider the scope, tone, manner, and structural framework of the discourse. In this scene, after executing a hostage, the mercenary conveyed a message to his employer through the phone, with a tone full of dissatisfaction and veiled threats. It can be inferred from the context that the mercenary is not happy with the way the employer communicates. His subsequent statement, "Did you hear that"? is not merely questioning the reception of sound

but also a stern warning. This signifies a severe warning to his employer, implying that if the current course of action continues, the consequences will be dire. Therefore, from the context, it is evident that the mercenaries' s statement is a direct threat to his employer, suggesting that persisting in the current approach will lead to a grim future. The scene vividly captures the mercenaries' arrogance through their triumphant expressions and the lifeless bodies scattered behind them. The mercenaries' smug demeanor and the lifeless bodies strewn behind them visually reinforce their cruelty and confidence. In this case, the translation "That could also be your fate" is adapted from the original phrase "that could be the sound of the future". Given the threatening tone and expression of the mercenary in the film, as well as the overall atmosphere of the scene, the translator chose a more direct and contextually appropriate phrase in Chinese to convey the mercenary's dissatisfaction and warning to his employer.

Besides, the directness of the Chinese translation perfectly matches the visual clues. The clear threat in "That could also be your fate" echoes the arrogant expression of the mercenaries, enhancing the consistency between language and visual cues. At this moment, the background music suddenly introduces a jarring accent, aiming to emphasize the importance of this scene and the paired dialogue. The sudden and jarring accent in the background music highlights the seriousness of the scene and the importance of the subsequent dialogue. The explicit threat in the translation, combined with the intensified background music, increases the audience's sense of unease and foreboding. The sudden auditory cues amplify the threat in the mercenaries' words, making the sense of threat more direct and real.

Therefore, the translation "That could also be your fate" effectively captures and magnifies the original message through its directness and consistency with the multimodal elements of the scene. In multimodal discourse, different modes need to maintain coherence and consistency to form an organic whole. In the scene mentioned in the text, the subtitle translation is coordinated with visual elements (such as the mercenary's arrogant expression and the corpses behind him) and auditory elements (such as the sudden change in background music) to enhance the threatening atmosphere. This coordination between multimodal elements allows the audience to more naturally understand the plot development and character emotions, enhancing the coherence and fluency of the narrative.

Example 5:

ST: 一朝是战狼，终生是战狼！ (Once a wolf warrior, forever a wolf warrior)

TT: Once a wolf, always a wolf!

In the visual modality, Leng Feng's "determined expression" and "roaring" gestures convey anger and unyielding resolve. These visually encoded emotions resonate emotionally with the translated subtitle's concise phrasing and the repetitive rhythm of "Once...always...", which mirrors the short, forceful tone of his roar. The monosyllabic "wolf" better aligns with the physiological characteristics of a roar, synchronizing with the character's lip movements and facial muscle tension to enhance the scene's impact. In the auditory modality, the tense atmosphere created by battle sound effects and the character's roar is further intensified by the translation. The plosive sound of "wolf" and the stressed syllable in "always" generate a strong rhythmic flow, integrating the subtitle language seamlessly into the scene's acoustic system rather than treating it as isolated text. In terms of contextual coherence, the translation employs the metaphorical continuity of "wolf" to directly address the mercenary's challenge to Leng Feng's identity. By emphasizing the core semantic of "inherent nature prevails" through this animal metaphor, the subtitle forms a logical closure in the "challenge-rebuttal" conflict, propelling the plot toward its climax. Additionally, omitting "warrior" and using the wilder, more aggressive imagery of "wolf" implicitly aligns with the character's "lone wolf" combat style while echoing the film's themes of patriotism and heroism. This stylistic choice transforms the subtitle into an extension of the character's personality rather than a mere linguistic conversion. Drawing on Zhang's (2009) Multimodal Discourse Analysis framework, this translation achieves synchrony between the linguistic modality and visual-auditory elements, enabling viewers to grasp the character's emotions without diverting attention from the screen. By using a "de-identifying" metaphor ("wolf" over "wolf warrior"), it circumvents potential cultural barriers related to military identity, instead invoking a universal concept of "inherent nature" that resonates with diverse audiences—consistent with the MDA principle that context shapes meaning construction. In conclusion, the translation demonstrates contextual integration by harmonizing multimodal elements and reinforcing narrative logic.

It functions not merely as a linguistic transposition but as a contextual reconstruction of scene atmosphere, character emotion, and plot coherence, exemplifying the theoretical value of "context as the core of meaning construction" in multimodal discourse analysis.

4.3. At Content Level

MDA focuses on the semantics and forms of discourse content and their presentation in different modes. The text mentions that film subtitles are immediate and subject to time and space constraints. Therefore, translators can adjust the content as needed, such as adding or deleting details, to ensure the accuracy and conciseness of the subtitles.

Example 6:
ST: 再喝! (Drink again!)
TT: Again!

In this scene, the protagonist Leng Feng participates in a heated drinking contest, which occupies a significant portion of the film. Intertwined with the emotional flashbacks of his fiancée, the full minute of drinking captures the mix of emotions. In this scene, Leng Feng seeks solace in alcohol. Considering the fervent atmosphere of the drinking scene, Leng Feng recalls his fiancée's tearful sobs, as well as factors such as sound, visuals, color, etc. The translation "Again!" is adapted from the original phrase "Drink again!". In the context of the drinking scene in the film, the translator simplified it to fit the subtitle word limit while accurately conveying the original meaning and matching the characters' actions, expressions, and scene atmosphere. This demonstrates that in multimodal discourse analysis, content adjustment is necessary to better adapt to multimodal expression. Translators need to adjust the content according to the characteristics of different modes and the audience's comprehension ability to achieve the best communication effect.

Example 7:
ST: I sure hope these Chinese people don't die.
TT: 中国佬，你可别死. (Chinese dude, don't you dare die.)

MDA emphasizes the importance of integrating content from different modes for meaning construction. In the scene mentioned in the text, language (subtitles), visual elements (characters' actions, expressions, scene, etc.), and auditory elements (background music, sound effects, etc.) together form a complete meaning space.

The scenario involves the leader of mercenaries expressing his desire to find the protagonist and his group, as Leng Feng and others may possess the live vaccine he seeks. The original text, which does not directly translate to "I sure hope these Chinese people don't die" should be rephrased to provide a more subtle and emotionally resonant explanation, given the hostile nature of both sides. A proper translation could be "Chinese dude, don't you dare die" which not only adds depth but also reflects the subtle undercurrents in the dialogue. This phrase conveys not only a begrudging respect for the protagonist's survival skills but also reflects the mercenary leader's strong desire for the live vaccine. From a multimodal discourse analysis perspective, translating "I sure hope these Chinese people don't die" to "Chinese dude, don't you dare die" has several notable aspects that enhance the overall narrative and emotional impact of the scene.

This translation is coordinated with the visual elements (the mercenary's arrogant posture and the tense scene) and auditory elements (the suspenseful background music) to create a hostile and dangerous atmosphere. On the visual mode level, the mercenary crosses his hands with a disdainful facial expression. This body language and facial expression convey arrogance and obvious contempt towards the protagonist. Although the original line "I of course hope these Chinese people don't die" carries a sense of irony, it lacks the directness and intensity of the translation. The translation "Chinese guy, don't you dare die" using the derogatory term "Chinese guy" (可译为 "支那人") immediately amplifies the disdain and hostility of the mercenary. Such wording is closer to the image of arrogance and contempt of the character expression. On the auditory mode level, this scene is accompanied by the tense and suspenseful music typical of horror films, enhancing the audience's emotional involvement. The background music emphasizes the imminent danger and vulnerability of the protagonist. The translated lines, which are direct and aggressive in language, resonate more strongly with the tension and hostility conveyed by the music. The abrupt and rough nature of "Chinese guy, don't you dare die" echoes the sharp and unsettling tones of the background music, thereby enhancing the overall sense of threat and urgency.

Additionally, the communicative relationship between the visual and auditory modes is crucial here. The mercenary's body language, facial expression, and suspenseful

music combine to form a multimodal ensemble, conveying a complex emotional landscape. The translation choice strengthens this landscape by providing language clues that match the visual and auditory elements. For example, the use of the derogatory term "Chinese guy" not only reflects the mercenary's disdain but also arouses a stronger emotional response from the audience, making their emotions more closely linked to the protagonist's plight. Therefore, the translation from "I of course hope these Chinese people don't die" to "Chinese guy, don't you dare die" effectively utilizes multimodal discourse elements to enhance the emotional impact of the scene. By choosing a more direct and aggressive word, the translation combines the language mode with the visual and auditory modes to form an organic meaning whole, enabling the audience to establish connections between different modes and more comprehensively and accurately understand the meaning conveyed by the discourse.

4.4. At Expression Level

MDA regards linguistic and non-linguistic modes (such as images, sounds, actions, etc.) as joint participants in meaning construction. The text mentions that subtitle translation needs to consider the presentation of linguistic text, such as font size, position, display time, etc., while also paying attention to non-linguistic elements that accompany the language mode, such as the characters' tone, intonation, actions, and background music.

Example 8:
ST: 他们立刻就这样 (They immediately did so.)
TT: All their cares go away.

In this vivid scene, Leng Feng and the old platoon leader stood on a high point, overlooking the crowd of people dancing and singing around the bonfire. Moments earlier, Leng Feng had just received support from a Chinese warship and assured the factory workers who were trapped that they could safely evacuate. Upon receiving this reassuring news, the initially tense crowd immediately burst into celebratory singing and dancing. The platoon leader jokingly praised the resilience of the African people, saying with a lighthearted tone that with the bonfire, "all their cares go away". In the translation "all their cares go away," adapted from "they immediately did so", the translation is not only concise and clear in language but also matches the visual scene of people dancing and singing around the bonfire and the cheerful at-

mosphere. The translator selects appropriate vocabulary and sentence structures to make the subtitles interact with visual and auditory elements, jointly conveying the relief and joy of the people after learning the good news of safe evacuation.

Additionally, the long vowel sound in "go away" gives a feeling of light-heartedness and pleasure, which is consistent with the cheerful scene. This optimization of language not only considers the semantics and expression effect of the language itself but also fully considers the coordination and consistency with visual, auditory, and other modes. In addition, when selecting vocabulary, the translator also focuses on its emotional connotations and cultural implications. The word "cares" in the translation not only contains the original meaning of "worries" but also has a broader emotional meaning, which can resonate with audiences from different cultural backgrounds. Therefore, the translation reads, "All their cares go away" successfully conveys the emotions and intentions of the original text through concise language, a cheerful tone, harmonious sight and sound, and precise word choices, thus achieving the essence of a good translation.

This illustrates the importance of the synergy between linguistic and non-linguistic modes in multimodal discourse analysis. Translators need to consider the characteristics and expression methods of multiple modes comprehensively, enabling them to complement and reinforce each other to achieve the best expression effect.

Example 9:
ST: 你醒了 (You're awake.)
TT: Welcome back.

The original line "你醒了" (You're awake) merely states an objective fact straightforwardly, while "welcome back" conveys a stronger emotional resonance through its emotive linguistic expression. This translation fosters a deep interaction with the multimodal elements of the scene. In this sequence, Leng Feng gradually regains consciousness in a convoy en route to join the Chinese military, having just emerged victorious from a fierce battle. As soft light filters into the vehicle and soothing background music plays, the heroine Rachel looks at him with a relieved smile. The "welcome" in the linguistic modality echoes the friendly expressions and relaxed body language in the visual modality, collectively conveying the emotion of "welcoming the protagonist back from danger to new life". This approach highlights the emotional tension of the scene more effectively

than the literal "你醒了".

From the auditory modality perspective, the heroine uses a gentle and reassuring tone, and the pronunciation of "welcome back" perfectly aligns with the rhythm of her voice, further enhancing the sense of intimacy. By contrast, the declarative tone of "你醒了" is relatively bland, failing to fully reflect the profound emotional bond forged through their shared life-and-death experiences.

The "welcome" in the linguistic modality forms a striking contrast with the "peaceful atmosphere after the battle" in the visual imagery, subtly accentuating the protagonist's transition from danger to safety. Meanwhile, the word "back" implies a strong sense of national belonging, which deeply resonates with the patriotic core of the film. This translation choice fully embodies the expressive level's requirement to "optimize emotional transmission through linguistic forms", achieving an organic integration and synergistic effect among linguistic, visual, and auditory modalities.

Example 10:

ST: 小日子过得挺美的呀 (Your daily life is quite nice.)

TT: I see you're enjoying your life.

The Chinese expression "小日子过得挺美的呀" is a typical colloquial banter that conveys a light-hearted and humorous tone in interpersonal communication. Its English translation, "I see you're enjoying your life," ingeniously employs the epistemic construction "I see + object clause" in everyday English conversation, paired with the emotionally charged verb "enjoy." This approach not only preserves the original bantering tone but also forms a deep synergy with the multimodal elements of the scene.

More specifically, in terms of visual modality, the frame depicts Leng Feng strolling leisurely with a dog in the snowy landscape. Dressed in bulky outdoor attire, he exhibits a relaxed posture, with a broad smile revealing his teeth and a look of relief in his eyes after the battle. The visual signifier of "leisurely vacation" directly maps onto the semantic content of "enjoying your life" in the translation, creating an emotional resonance between the linguistic text and the visual imagery. By contrast, a literal translation such as "Your daily life is quite nice" would not only weaken the vividness of the colloquial banter due to its formal diction but also fail to establish a meaningful intermodal resonance with the protagonist's relaxed body language and joyful facial expressions.

In terms of auditory modality, the soothing background music in the scene creates a tranquil atmosphere with its gentle melodic lines and low-frequency sound effects, echoing the rhythmic cadence of the translation's concise syntactic structure. In the phonetic flow of "I see you're enjoying your life," the stress on "see" and "enjoying" naturally aligns with the commander's bantering tone—conveying both playful concern for the subordinate's rare respite and implicit relief that he can finally relax after going through hardships. This tonal treatment forms a cross-modal emotional echo with Leng Feng's smile in the frame, collectively constructing the narrative node of "a moment of relaxation after intense combat."

The bantering tone in the linguistic modality, the smiling expression in the visual modality, and the soothing music in the auditory modality are organically integrated, collectively serving the narrative transition in the film from "peaceful rest" to "call of duty." This fully demonstrates the core requirement of the expressive level to coordinate multimodal elements and shape the scene atmosphere through linguistic form optimization.

5. Conclusions

This study is the first to systematically apply the Multimodal Discourse Analysis (MDA) framework to the investigation of subtitle translation in *Wolf Warrior 2*, a mainstream Chinese film. It pioneers a research paradigm with both theoretical depth and practical significance, introducing a novel academic perspective to the field of cross-cultural audiovisual translation. Unlike traditional subtitle translation studies, which predominantly focus on linguistic transfer, this research breaks new ground by revealing the synergistic mechanisms among linguistic, visual, and auditory signs, while validating the "bridging function" of non-verbal elements in cross-cultural communication.

Furthermore, through detailed case studies, this paper delves into the dynamic interplay of multiple dimensions, including cultural schema adaptation, contextual logic adjustment, audiovisual resource integration, and emotional expression coordination, thereby expanding the applicability of MDA theory in cross-cultural contexts. This approach not only enriches the theoretical framework of multimodal anal-

ysis but also provides empirical support for the development of AI-powered multimodal translation models, highlighting the dual value of this study in terms of theoretical innovation and practical guidance.

Additionally, this research offers both a new theoretical toolkit for multimodal analysis of subtitle translation and concrete operational strategies for practitioners. By conducting an in-depth examination of *Wolf Warrior 2*'s subtitle translation, the study demonstrates how multimodal strategies can facilitate effective cultural transfer and emotional resonance in cross-cultural contexts. When addressing cultural disparities, subtitlers should employ creative translation strategies, such as "metaphorical substitution," to achieve equivalent cultural schema conversion. The study also emphasizes the importance of contextual logic in translation and how semantic adjustments can enhance narrative coherence. Moreover, it explores how visual information can be leveraged to streamline subtitles and how synchronization with audio elements can strengthen emotional impact.

The findings of this study hold significant implications not only for translation studies but also for film studies, cross-cultural communication research, and the construction of cultural soft power. By applying MDA theory to subtitle translation, this research provides new insights into how films, as cross-cultural media, convey meaning—particularly in a globalized context where cinema serves as a medium for fostering mutual understanding and respect among cultures. Furthermore, this study outlines future directions for theoretical and applied research in cross-cultural translation, especially in the era of AI-assisted translation technology, by demonstrating how multimodal approaches can enhance the effectiveness of audiovisual translational analysis can enhance translation quality and deepen the scope of cross-cultural exchange.

Author Contributions

Author contributions are as follows: Conceptualization, J.N. and X.M.; Methodology and validation, J.N. and X.M.; formal analysis, X.M.; writing—original draft preparation, X.M.; writing—review and editing, J.N.; visualization, J.N. and X.M.; supervision, J.N.. All authors have read and agreed to the published version of the manuscript.

Funding

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data supporting reported results can be found in the article.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] Gambier, Y., 2003. Introduction: Screen Transadaptation: Perception and Reception. The Translator. 9(2), 171 – 189. DOI: https://doi.org/10.1080/13556509.2003.10799152

[2] Liu, D.Y., 2010. The evolution of AVT terminology: From film translation to multimedia translation. Shanghai Journal of Translators. (4), 61 – 65.

[3] Delabastita, D., 1989. Translation and mass-communication. Babel Revue internationale de la traduction/International Journal of Translation. 35(4), 193

– 218. DOI: https://doi.org/10.1075/babel.35.4.02del

[4] Gottlieb, H., 1994. Subtitling: Diagonal translation. Perspectives. 2(1), 101 – 121. DOI: https://doi.org/10.1080/0907676x.1994.9961227

[5] Cintas, J.D., Remael, A., 2014. Introduction to subtitling. In: Cintas, J.D., Remael, A. (eds.). Audiovisual Translation: Subtitling. Routledge: London, UK. pp. 8 – 29.

[6] Pérez-González, L., 2014. Audiovisual translation: theories, methods and issues. Routledge: London, UK. p. 356.

[7] Chaume, F., 2018. An overview of audiovisual translation: Four methodological turns in a mature discipline. Journal of Audiovisual Translation. 1(1), 40 – 63. DOI: https://doi.org/10.47476/jat.v1i1.43

[8] Wang, C.S., 2017. A knowledge graph analysis of foreign audiovisual translation research. Foreign Language Research. (1), 80 – 84. DOI: https://doi.org/10.16263/j.cnki.23-1071/h.2017.01.014

[9] Wang, H.F., Qiao, M.Q., 2018. Audiovisual translation, multimedia translation, and multimodal translation: Differentiation and reflection. Foreign Language and Literature Research. 4(6), 95 – 104. DOI: https://doi.org/10.16651/j.cnki.fllr.2018.0083

[10] Wang, J.H., Li, J., 2021. A review of quantitative evaluation research on foreign audiovisual translation: Status, limitations, and implications. Foreign Language. 44(3), 110 – 118.

[11] Yan, X.S., Lan, H.J., 2021. A review of domestic audiovisual translation research (2000 – 2020): Based on the analysis of translation research-related journals. Language Translation. (2), 64 – 70.

[12] Fu, J.M., Zhang, K.Z., 2021. A 20-year review of audiovisual translation research (2000 – 2019): A knowledge graph analysis and literature comparison based on Chinese and foreign journals. Foreign Language China. 18(3), 88 – 97. DOI: https://doi.org/10.13564/j.cnki.issn.1672-9382.2021.03.012

[13] Chen, Y.P., Zhang, C.H., 2017. Chinese subtitle translation of English films: A multimodal analysis focusing on the relationship between images and texts. China Translators Journal. 38(5), 105 – 110.

[14] Liu, X.H., Zhang, L., 2017. Subtitle translation and cross-cultural communication: A case study of the American version of Empresses in the Palace. View on Publishing. (6), 64 – 66. DOI: https://doi.org/10.16491/j.cnki.cn45-1216/g2.2017.0229

[15] Lu, J.P., Zhu, Q.W., 2023. Cultural element translation strategies in Chinese films from the perspective of international communication: A comparative analysis of two English versions of the subtitles of Farewell My Concubine. Foreign Language Teaching. (6), 134 – 144, 150. DOI: https://doi.org/10.13458/j.cnki.flatt.005023

[16] Wang, J.G., Dai, J.X., 2023. The influence of pragmatic differences between Chinese and English on film subtitle translation: A case study of the subtitle translation of Dreams Come True. China Translation Journal. 44(3), 148 – 155.

[17] Yin, X.F., 2024. Subtitle translation strategies from the perspective of skopos theory. Film Literature (18), 130 – 134.

[18] Wang, H.S., Li, Y., 2020. Research on subtitle translation technology: Status, problems, and suggestions. Technology Enhanced Foreign Language Education. (6), 80 – 85, 86. DOI: https://doi.org/10.20139/j.issn.1001-5795.2020.06.013

[19] Li, C.H., Zhang, Z., 2023. Research paradigms and approaches in audiovisual translation based on digital humanities. Foreign Language Research. (1), 61 – 65. DOI: https://doi.org/10.16263/j.cnki.23-1071/h.2023.01.008

[20] Wu, Y., Mou, Y.W., 2022. Research on multimodal national translation strategies for telling Chinese stories. Foreign Language Education. 43(1), 76 – 82. DOI: https://doi.org/10.16362/j.cnki.cn61-1023/h.2022.01.012

[21] Zhang, W.H., Kang, L.M., 2023. Multimodal translation and international dissemination of Chinese martial arts. Journal of Wuhan Sports University. 57(10), 55 – 62. DOI: https://doi.org/10.15930/j.cnki.wtxb.2023.10.007

[22] Jewitt, C., 2006. K. O'Halloran: Multimodal Discourse Analysis: Systemic Functional Perspectives. Continuum, 2004. Applied Linguistics. 27(2), 335 – 337. DOI: https://doi.org/10.1093/applin/aml002

[23] Baldry, A., Thibault, P.J., Lemke, J., 2006. Multimodal transcription and text analysis: A multimodal toolkit and Coursebook. University of Toronto Press: Toronto, ON, Canada. pp. 18 – 22.

[24] Pettit, Z., 2009. 3: Connecting cultures: Cultural transfer in subtitling and dubbing. In: Díaz Cintas, J. (eds.). New Trends in Audiovisual Translation. Multilingual Matters: Bristol, UK. pp. 44 – 57.

[25] Norris, S., Pirini, J.P., 2017. Communicating knowledge, getting attention, and negotiating disagreement via videoconferencing technology: A multimodal analysis. Journal of Organizational Knowledge Communication. 3(1), 23. DOI: https://doi.org/10.7146/jookc.v3i1.23876

[26] Lv, J., Wu, W.Z., 2012. A study on subtitle translation of the film The Flowers of War from the perspective of multimodal discourse analysis. Shanghai Journal of Translators. (4), 36 – 38.

[27] Liang, Y.Z., 2016. Subtitle translation of British and American films and TV series from the perspective of multimodal discourse analysis. Contemporary TV. (12), 80 – 81. DOI: https://doi.org/10.16531/j.cnki.1000-8977.2016.12.043

[28] Yin, M.M., 2020. An exploration of subtitle translation strategies for French classic art films from the per-

spective of multimodal discourse analysis. Contemp Cinema. (2), 164 – 167.

[29] Wu, J.G., Li, D.Q., Zhang, H.S., 2021. Multimodal translation of the commentary in Amazing China and the construction of national image. Shandong Foreign Language Teaching. 42(5), 31 – 41. DOI: https://doi.org/10.16482/j.sdwy37-1026.2021-05-004

[30] Wu, J.G., Li, Y.J., 2024. Multimodal context reconstruction and the dissemination of Chinese film and television culture: A case study of the subtitle translation of My People, My Country. Shandong Foreign Language Teaching. 45(2), 11 – 21. DOI: https://doi.org/10.16482/j.sdwy37-1026.2024-02-002

[31] Niu, Q., 2025. Experiences and insights from subtitle translation of the film Chang'an Three Miles under multimodal discourse analysis. Film Literature. (5), 145 – 149.

[32] Halliday, M.A.K., 1994. An introduction to functional grammar. Edward Arnold: London, UK. pp. 1 – 3.

[33] Kress, G.R., van Leeuwen, T., 2017. Multimodal discourse: The modes and media of Contemporary Communication. Bloomsbury Academic: Bloomsbury, UK. pp. 1 – 142.

[34] Kress, G., van Leeuwen, T., 2020. Reading images: The Grammar of Visual Design. Routledge: London, UK. pp. 1 – 310.

[35] Kress, G., Bezemer, J., 2023. Multimodal Discourse Analysis. In: Handford, M., Gee, J.P. (eds.). The Routledge Handbook of Discourse Analysis. Routledge: London, UK. pp. 139 – 155.

[36] Moskvil, J.Ø., 2021. Multimodal meningsskaping I NRK-produksjonen sangfoni. Sakprosa. 13(1). DOI: https://doi.org/10.5617/sakprosa.8126

[37] Li, Z.Z., 2003. A sociosemiotic analysis of multimodal discourse. Foreign Language Research. (5), 1 – 8, 80.

[38] Zhu, Y.S., 2007. The theoretical foundation and research methods of multimodal discourse analysis. Foreign Language Research. (5), 82 – 86. DOI: https://doi.org/10.16263/j.cnki.23-1071/h.2007.05.034

[39] Zhang, D.L., 2009. An exploratory study on the comprehensive theoretical framework of multimodal discourse analysis. China Foreign Language. 6(1), 24 – 30.

[40] Feng, H., 2024. Exploring the identity of Chinese cultural relics on British Museum websites through Multimodal Discourse Analysis: A case study of the official website. Communications in Humanities Research. 40(1), 50 – 60. DOI: https://doi.org/10.54254/2753-7064/40/20242219