

Original Paper

A Multimodal Discourse Analysis of Subtitle Translation in the Film *Nezha: The Devil Boy Conquers the Sea*: Experience and Implications

Xinxin Chen¹

¹ Guangzhou College of Commerce, Guangzhou, China

Received: December 22, 2025 Accepted: January 28, 2026 Online Published: February 9, 2026
doi:10.22158/eltls.v8n1p43 URL: <http://dx.doi.org/10.22158/eltls.v8n1p43>

Abstract

*As an animated film carrying Chinese mythological narratives, *Nezha: The Devil Boy Conquers the Sea* explores themes of resistance to fate, identity construction, and value reconstruction, and has aroused wide resonance among audiences both at home and abroad. Exploring how subtitle translation can balance semantic accuracy and emotional transmission, ensuring that the film's value system transcends linguistic and cultural barriers and resonates with audiences from different cultural backgrounds, contributes to improving overseas audiences' overall perception and acceptance of Chinese animation. Taking *Nezha: The Devil Boy Conquers the Sea* as the research object, this paper, based on Multimodal Discourse Analysis (MDA), examines subtitle translation from four dimensions—culture, context, content, and expression. The study aims to provide experience and insights for film subtitle translation, enhance the artistic quality of audiovisual translation, and promote the global dissemination of Chinese animated works.*

Keywords

*Multimodal discourse analysis, animated film, subtitle translation, *Nezha: The Devil Boy Conquers the Sea*, cross-cultural communication*

1. Introduction

“Telling China’s stories well” has become an important mission in enhancing national cultural soft power, shaping China’s image, and strengthening cultural confidence amid profound changes in the global communication landscape. Film serves as a bridge and medium for cultural transmission. As a mass communication medium bearing the task of disseminating Chinese culture and shaping national image, Chinese cinema must actively participate in cross-cultural communication (Shao & Pan, 2006).

Animated films adapted from Chinese mythological stories constitute an important channel for telling Chinese stories. In recent years, Chinese animated films have developed rapidly. As a representative domestic production, *Nezha: The Devil Boy Conquers the Sea* (2025) has not only achieved remarkable box-office success in China but also gained wide attention in overseas markets, becoming an important symbol of Chinese animation “going global.” The film continues the narrative tradition of classical Chinese mythology and realizes a contemporary expression of traditional culture through modern visual language, audiovisual effects, and narrative techniques. With the increasing globalization of Chinese audiovisual products, subtitles, as a crucial medium of cross-cultural communication, have become a key bridge for foreign audiences to understand Chinese culture and values, and a critical step in the international dissemination of Chinese animation (Wang, 2022). Subtitle translation not only performs linguistic transfer but also participates in meaning construction through the coordination of visual, auditory, and emotional multimodal elements, playing an important role in promoting intercultural communication.

Multimodal discourse refers to communication phenomena that employ multiple sensory channels—such as auditory, visual, and tactile—through diverse semiotic resources including language, images, sounds, and actions (Zhang, 2009). With the multimodal development of audiovisual art, films are no longer single linguistic texts but comprehensive semiotic systems composed of images, sound, music, movement, and symbols. Traditional linguistically oriented translation studies are insufficient to explain meaning construction in audiovisual works. Multimodal Discourse Analysis (MDA), proposed by Kress and Van Leeuwen, has therefore emerged as an important research paradigm. It emphasizes that language is only one of many semiotic resources and that images, color, spatial layout, music, and movement also possess representational and meaning-making functions, highlighting the interaction among modes in meaning construction (Kress & Van Leeuwen, 2003). In China, multimodal discourse research is still developing. Against the background of enhancing cultural soft power and promoting Chinese culture globally, multimodal Chinese discourse studies are of positive significance in challenging Western centrism and constructing a Chinese discourse system (Li, Feng, & Deng, 2024). Zhang Delu, drawing on Halliday’s Systemic Functional Linguistics, proposed a comprehensive framework of multimodal discourse analysis consisting of four levels: culture, context, content, and expression, with the content level further divided into meaning and form (Zhang, 2009). This framework provides a new perspective for audiovisual translation by enabling researchers to understand the role of subtitles within multidimensional semiotic systems.

Based on this framework, this paper investigates how subtitle translation in *Nezha: The Devil Boy Conquers the Sea* achieves cross-modal coordination among language, image, sound, and movement across the four levels of culture, context, content, and expression, thereby reconstructing discourse meaning. The study aims to reveal the features and strategies of subtitle translation in multimodal discourse and to provide practical implications for the cross-cultural dissemination of Chinese animation.

2. Cultural Level

Within the multimodal framework, the cultural level constitutes the deep foundation of meaning construction. It involves ideologies shaped by thinking patterns, philosophies of life, social conventions, and genre potentials that realize these ideologies^[6]. Subtitle translation is not merely linguistic transfer but also ideological and cultural mediation. Translators should consider the interaction among multimodal elements and convey profound cultural meanings concisely within limited time and space.

The film draws extensively on Chinese mythology, reflecting distinctive Chinese thinking patterns, life philosophies, and cultural values in its storyline, character construction, and language use. Translating such culturally embedded elements into subtitles that can be understood and appreciated by international audiences is a key challenge.

Example 1:

Source: 你天生水象

Target: You were born of the water element.

This line is spoken by Taiyi Zhenren when reshaping Ao Bing's physical form. The concept of "water element" originates from the Chinese Five Elements (Wuxing) theory, which holds that all things in the universe are composed of five fundamental elements—wood, fire, earth, metal, and water. These elements interact dynamically and collectively constitute the material world. As one of the Five Elements, water possesses distinctive attributes and functions. Its nourishing capacity and downward flow symbolically represent humility, inclusiveness, and ultimate virtue.

The translator adopts a combined strategy of cultural equivalence and semantic expansion. Rather than rendering the phrase literally as "You are water-born" or "You belong to water," the translation introduces the term "element," a concept more familiar within Western discourse. This choice establishes a metaphorical correspondence between the Western notion of elemental systems and the Chinese Five Elements framework. In the corresponding scene, the visual modality reinforces the symbolism of water: Nezha and Ao Bing's spirits converge amid surrounding waves and blue radiance, while Ao Bing is dressed in blue. The dominant blue color palette, flowing water imagery, and low-toned background music jointly construct an atmosphere that blends softness with destiny. The subtitle thus works in complementary coordination with visual and auditory modes to convey meaning. Within limited subtitle space, the translation effectively transmits the symbolic implications of Eastern philosophy and achieves multimodal cultural mediation.

Examples 2 & 3:

Source: 师弟 / 大师兄

Target: Brother Taiyi / Brother Wuliang

These lines occur in a reunion scene between Taiyi Zhenren and Wuliang Xianweng, who is Taiyi's senior fellow disciple. In traditional Chinese culture, kinship terms within a master-disciple lineage, such as shixiong (senior fellow disciple) and shidi (junior fellow disciple), reflect a relationship structure centered on pedagogical inheritance. They embody Confucian social ethics that emphasize

hierarchical order and respect for seniority and teachers. Such titles function not only as markers of identity but also as carriers of emotional, ritual, and social significance. Since English lacks an exact equivalent for this intra-lineage relational system, cultural adaptation is required in translation depending on context and audience.

Here the translator employs an explicitation strategy by adding personal names—Taiyi and Wuliang—to the kinship term “Brother.” This approach is semantically clear and concise. It preserves the hierarchical distinction implied by the original terms while maintaining the cultural atmosphere and relational layering among characters. Target-language viewers can thus infer the social relationship and emotional distance between the characters through context, achieving both linguistic and cultural transmission.

Example 4:

Source: 给他点儿颜色瞧瞧

Target: and have a taste of my power.

This expression is a culturally marked Chinese idiom that metaphorically conveys the intention to teach someone a lesson or demonstrate one’s strength. The translation transforms the metaphor of “color” into the more explicit concept of “power,” and shifts the visual perception verb “look” into the gustatory metaphor “taste,” thereby accomplishing a cross-cultural transfer from figurative imagery to a more immediately intelligible concept.

At the multimodal level, the translated line aligns with the character’s facial expressions, gestures, and vocal delivery, reinforcing the semantic force of “displaying power” through visual and auditory channels. From a tonal perspective, the original phrase carries a colloquial, provocative, and confident flavor, with a slightly playful edge. The translation “have a taste of my power” preserves this provocative tone through the phrase “have a taste,” while the addition of the first-person possessive “my” heightens directness and self-assertion, strengthening the sense of confrontation and dominance. The shift from a relatively implicit metaphor to a more explicit expression reflects a cross-cultural tonal adjustment, which works in concert with performance modalities to reproduce the character’s confidence and intimidation effect.

3. Contextual Level

At the contextual level, multimodal discourse analysis treats context as a crucial mediating layer in meaning construction. Meaning does not emerge independently but is always supported and constrained by specific contextual conditions. In situated communication, discourse is shaped by contextual variables including field, tenor, and mode of discourse (Zhang, 2009). In audiovisual subtitle translation, the contextual layer is further manifested through the interaction of multiple semiotic resources, such as visual composition, facial expressions, tone of voice, and musical rhythm. For instance, when a character speaks in a low voice under dim lighting, even brief utterances can construct a complex emotional context. Translators must therefore fully recognize multimodal contextual

relations, preserve tonal and atmospheric cues in the target text, and conduct secondary analysis and reorganization of the source discourse to achieve semantic and affective equivalence. Otherwise, mismatches may arise between subtitles and images, or between subtitles and the film's intended meaning.

First, in subtitle translation, the control of discourse tenor is of central importance. Discourse tenor refers to the relationship between interlocutors and the attitudes and interpersonal stance expressed in communication. It reflects differences in social status, emotional orientation, and relational distance, and serves as a key factor in constructing character identity and speech style. Film subtitles not only transmit semantic content but also reproduce characters' emotions and relationships through linguistic style. If the translator neglects tenor, the target text may display tonal deviation or character distortion, thereby weakening the film's expressive power and the audience's emotional engagement. Translators should therefore base their decisions on a careful understanding of character relations and pragmatic intention, and flexibly adjust tone, lexical choice, and register to achieve unity between semantic equivalence and tonal representation.

Example 5:

Source: 若前方无路 / 我就踏出一条路 / 若天地不容 / 我就扭转这乾坤

Target: If no place to go / I will open up one myself! / If not accepted / I will overturn!

In this scene, Wuliang Xianweng attempts to coerce Nezha into submission despite being in a temporarily unfavorable position, relying on his supreme status in the celestial realm. Within the tripartite cosmology of immortals, demons, and humans, Wuliang Xianweng occupies the highest rank. Nezha, born as a demon child, holds low social status and is often treated as an outsider and potential threat. Although the power hierarchy strongly favors Wuliang Xianweng, Nezha openly resists authority and challenges fate, seeking to break the imposed label of "demon" and assert agency over his own destiny. The translation must therefore effectively convey Nezha's rebellious spirit, courage, and resolute tone.

The original employs a classical Chinese hypothetical structure marked by “若……,” producing a concise, solemn, and forceful rhetorical rhythm. The translation adopts the parallel structure “If no place to go / If not accepted,” likewise omitting the subject and the copular construction (“there is,” “I am”), thus preserving brevity and oral impact. The translator also omits culturally loaded cosmological terms such as “天地” and “乾坤” (roughly, “the world” or “the universe”) without impairing audience comprehension. This ellipsis and reduction strategy aligns with the rapid shot transitions and tense musical pacing, achieving rhythmic consistency across modes. The translation follows the multimodal discourse principle of intersemiotic complementarity, whereby subtitle language works together with visual and auditory signs to construct an emotional climax.

Verb selection in the English version further highlights agency and volition. Expressions such as “open up” and “overturn” are dynamic and directional, mapping onto the visual imagery of Nezha rising into the air amid flames and explosive motion. Multimodal discourse analysis holds that meaning can be

reconstructed through cross-modal semiotic co-articulation. Although culturally specific lexical items are reduced, the action-oriented verbs compensate through kinetic force, maintaining a dynamic balance among language, imagery, and emotion in cross-linguistic transfer.

Second, discourse field exerts a significant influence on lexical choice and stylistic presentation in subtitle translation. Field determines the subject matter and type of social action involved in communication and forms the basis of register variation. Different registers correspond to different linguistic features; therefore, subtitle translation should flexibly adjust style and expression according to plot content and register characteristics, ensuring that the translation remains consistent with the source text at semantic, affective, and aesthetic levels while also coordinating with multimodal elements such as imagery, music, and vocal delivery.

Example 6:

Source: 吃本王一斧 / 吃小爷一轮儿 / 吃小爷一百轮儿

Target: Take my axe! / Take my wheel! / Take my wheel times a hundred!

This segment occurs in a combat scene dominated by kinetic visual action and reinforced by high-impact sound design. The translation adopts direct and dynamic English expressions—“Take my axe!” and “Take my wheel!”—which conform to the short imperative patterns typical of English combat discourse. The imperative verb “Take” foregrounds force and immediacy, matching the expectations of the battle register. If rendered too literally, for example as “Receive the axe from me!”, the line would sound overly formal and dilute the intended intensity, thereby conflicting with the conventions of combat dialogue.

Because this segment belongs to an action-centered combat register, its primary linguistic function is to generate rhythm and momentum. Through brief, forceful imperatives, the translation successfully reproduces the original’s combative atmosphere and emotional intensity, achieving functional equivalence at the register level. This demonstrates that accurately identifying discourse field enables translators to adjust stylistic choices in accordance with scene type, thereby promoting multimodal coherence between language and image.

4. Content Level

At the content level, multimodal discourse analysis encompasses both the level of meaning and the level of form. At the level of discourse meaning, three metafunctions are involved—ideational (conceptual) meaning, interpersonal meaning, and textual meaning—each constrained by discourse field, tenor, and mode. At the level of form, the formal features of different modalities interact and jointly realize discourse meaning (Zhang, 2009).

Subtitle translation operates simultaneously on both levels. At the meaning level, it must reproduce ideational, interpersonal, and textual meanings by accurately conveying information, reconstructing character emotion and tone, and maintaining narrative coherence. At the formal level, it must coordinate typographic and technical features—such as font, color, spatial placement, and display

duration—with visual and auditory modalities, thereby enabling an integrated presentation of discourse meaning.

Example 7:

Source: 是红凤凰黄凤凰·粉红房上飞凤凰

Target: She sells seashells by the seashore...

This line functions as a coded verbal signal exchanged between Ao Bing (disguised as Ao Run) and Nezha (disguised as Ao Shun). The original Chinese phrase is a typical tongue twister that foregrounds rapid repetition of similar initials and complex phonetic clusters. Its literal semantic content (red phoenix, yellow phoenix, pink house, etc.) is not central; rather, its core function lies in phonetic playfulness and rhythmic performance. Subtitle translation in this case does not need to preserve literal meaning but should reproduce the function of linguistic play within the source modality.

The translator replaces the Chinese tongue twister with a widely recognized English one—"She sells seashells by the seashore"—and further extends it with "and the shells she sells all smell for sure," creating greater rhythmic symmetry and increasing articulatory difficulty. Although the semantic content is completely substituted, the translation achieves functional equivalence in phonetic density, rhythm, and humorous effect, demonstrating functional transferability in multimodal meaning construction.

From the perspective of audiovisual semiotic coordination, the characters' facial expressions, speech rate, intonation, and phonological rhythm jointly construct an interactive visual-auditory mode in this scene. The original phrase produces comic effect through repeated nasals and consonant clusters, while the English tongue twister likewise relies on alliteration and phonemic repetition (s, sh). Although the subtitle no longer conveys the literal imagery, it aligns with mouth movement patterns, delivery speed, and prosodic rhythm, thus achieving intermodal complementarity and coherence. Viewers perceive the playful language effect through the combined rhythm of sound and subtitle text, resulting in a reception experience comparable to that of the source-language audience.

Example 8:

Source: 莫惹事生非哈

Target: Not now.

In this scene, Nezha is mocked for his appearance and prepares to retaliate by breaking a stone pillar, but Taiyi Zhenren stops him and urges restraint. A literal rendering such as "Don't make trouble" would produce a strong imperative tone that does not match the relatively gentle and cautionary force of the original utterance. The translator instead adopts a sense-for-sense strategy, rendering the line as "Not now." This brief, decisive structure conveys urgency while remaining less confrontational than a direct imperative.

The softened yet firm tone aligns with the character's facial expression and the tense pacing of the scene, producing multimodal consistency. Compared with a prohibitive command, "Not now" sounds more tactful and situationally responsive, better reflecting the original interpersonal tone. Through this

context-sensitive reformulation, the subtitle achieves situational meaning reconstruction and maintains coherence with performance cues and scene rhythm across modalities.

5. Expressive Level

At the expression level, meaning is primarily realized through media, which can be broadly divided into linguistic and non-linguistic categories. Linguistic media include spoken dialogue features such as voice quality, volume, and intonation, as well as written features such as subtitle font and layout. Non-linguistic media include bodily movement, facial expression, props, and environmental setting. In subtitle translation, translators should therefore take into account both linguistic and non-linguistic media in order to produce contextually appropriate renderings ^[6]. At this level, subtitle translators must consider not only auditory and textual factors—such as pitch, prosody, subtitle size and shape, and screen-space distribution—but also the interaction between subtitles and non-verbal semiotic resources when translating source-language dialogue ^[7].

Example 9:

Source: 你打我噯 你打我噯

Target: Hit me! Come on!

In this scene, Taiyi Zhenren establishes a barrier while reconstructing the physical bodies of Nezha and Ao Bing. Shen Gongbao attempts but fails to break the barrier. Taiyi Zhenren then mocks him, first verbally asserting that he is incapable of breaking it, and then provocatively slapping his own face while repeating, “你打我噯 · 你打我噯” (“Hit me, Hit me!”). The repetition in Chinese intensifies the taunting effect. Instead of mechanically reproducing the repetition with two identical instances of “Hit me,” the translator renders the second clause as “Come on!” In English, “come on” can likewise function as a provocation, conveying a challenge equivalent to “go ahead” or “try it.”

This variation not only preserves the playful and provocative tone but also accords with English stylistic preferences that tend to avoid exact repetition in such contexts. The subtitle works in tandem with non-verbal media: Taiyi Zhenren’s self-slapping gesture, his smug and teasing facial expression, the sharp slapping sound effect, and Shen Gongbao’s frustrated reaction. Together, these verbal and non-verbal cues construct a coherent multimodal portrayal of provocation, enabling target-language viewers to grasp the intended characterization and dramatic effect more effectively.

Example 10:

Source: 受这点罪又算什么

Target: E-A-S-Y P-E-A-S-Y!

In this scene, Nezha forces himself to consume his own vomit in order to restore the medicinal effect he has expelled. The original line expresses a relaxed, nonchalant attitude and carries a sense of self-encouragement and bravado. The translation “E-A-S-Y P-E-A-S-Y!” adopts a common English colloquial expression. Through spelled-out lettering and full capitalization, the subtitle visually reinforces a light, confident tone and heightened expressiveness.

At the interpersonal level, the translation preserves the original utterance's casual and reassuring stance. At the formal level, the orthographic stylization aligns with prosodic rhythm and intonational emphasis in the spoken delivery, achieving coordination with the auditory modality. This spelling-based emphasis thus supports multimodal consistency in the representation of emotion and attitude.

6. Conclusion

From the four dimensions of multimodal discourse analysis, the subtitle translation of *Nezha: The Devil Child Conquers the Sea* demonstrates coordinated meaning construction through the interaction of linguistic, visual, and auditory semiotic systems. In the translation process, the translator attends not only to the transfer of linguistic information but also to the complementary relations among different modalities, enabling the subtitles to reproduce, to the greatest extent possible, the film's emotional intensity and aesthetic impact in cross-cultural communication.

At the cultural level, strategies such as cultural equivalence and semantic expansion are employed to transform language signs embedded with traditional Chinese philosophical and cultural concepts into expressions accessible to target-language audiences. In doing so, the subtitles retain distinctive cultural features while fostering cognitive and emotional resonance across cultures. At the contextual level, the translator demonstrates sensitivity to discourse tenor and tonal shifts, using techniques such as omission and reduction to align subtitle rhythm with visual pacing, musical cues, and character emotion, thereby achieving synchronicity between semantic content and affective force. At the content level, ideational, interpersonal, and textual meanings are jointly considered; expressions are selected in accordance with target-language usage conventions so as to effectively convey the original humor, evaluative stance, and emotional coloring. At the expression level, linguistic and non-linguistic modalities are integrated, with full use made of facial expressions, gestures, and sound effects to support meaning-making, enhance audiovisual impact, and strengthen audience immersion.

In summary, within a multimodal analytical framework, the translator adopts an integrated set of strategies—including cultural adaptation, contextual coordination, semantic reconstruction, and expressive enhancement—to ensure that subtitles function synergistically with image, sound, and rhythm. This multimodal coordination not only enhances the film's emotional expressiveness and communicative effectiveness but also offers a valuable strategic and practical model for the international translation of Chinese animated cinema.

References

- Dai, D. D., & Qi, W. H. (2023). A multimodal analysis of subtitle translation in *Nezha: Birth of the Demon Child*. *English Square*, 2023(08), 7-11.
- Kress, G., & Van Leeuwen, T. (2003). *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Li, M. J., Feng, D. Z., & Deng, Y. (2024). Developments and frontiers in international multimodal

- discourse analysis. *Modern Foreign Languages*, 47(03), 419-430.
- Shao, P. R., & Pan, X. H. (2006). On the cross-cultural communication strategies of Chinese films in the context of globalization. *Journal of Zhejiang University (Humanities and Social Sciences)*, 2006(01), 65-73.
- Wang, X. L. (2022). Cross-cultural communication strategies of Chinese animated films in the context of globalization. *Modern and Contemporary Cultural Creativity*, 2022(17), 84-86.
- Zhang, D. L. (2009). Exploring a comprehensive theoretical framework for multimodal discourse analysis. *Chinese Foreign Languages*, 6(1), 24-30.
- Zhang, D. L. (2009). Multimodal discourse theory and the application of media technology in foreign language teaching. *Foreign Language Education*, 30(04), 15-20.