
| RESEARCH ARTICLE

A Study on the Subtitle Translation of the Film *Cliff Walkers* from the Perspective of Multimodal Discourse Analysis

Caiyan Zou

College of Foreign Studies, Jinan University, Guangzhou, PR China

Corresponding Author: Caiyan Zou, E-mail: 2376639585@qq.com

| ABSTRACT

This article explores the subtitle translation of *Cliff Walkers* from the perspective of multimodal discourse analysis, using visual grammar as the theoretical framework. The study reveals that subtitle translation is a multimodal practice, where language choices are closely related to visual, auditory and narrative elements. The translator's adaptation strategies, such as using expressions with cultural resonance, condensing themes in chapter titles, and coordinating with visual and auditory cues, demonstrate a profound understanding of the ways meanings are constructed in different modes. These strategies not only ensure the accuracy of information but also enhance the expressiveness and cultural accessibility of the film.

| KEYWORDS

Multimodal discourse analysis; subtitle translation; visual grammar theory; *Cliff Walkers*

| ARTICLE INFORMATION

ACCEPTED: 15 March 2026

PUBLISHED: 02 April 2026

DOI: 10.32996/ijllt.2026.9.4.3

1. Introduction

In recent years, with the continuous advancement of the national Going Global strategy, film and television works, as carriers of the country's cultural soft power, have played an indispensable role in shaping the national image, presenting social realities, and conveying cultural values. Subtitles, as a language medium in film and television dissemination, serve as a bridge between cultural input and output, and their translation quality directly affects the target audience's understanding and acceptance of the source culture. How to accurately, effectively, and acceptably convey the complex historical context, ideological content, and national cultural imagery in Chinese mainstream films in cross-cultural communication has become an important issue in subtitle translation research. However, previous related studies have mostly focused on single-modal discourse, with less attention paid to multimodal discourse.

Directed by Zhang Yimou, the film *Cliff Walkers* is set against the backdrop of the 1930s in the puppet state of Manchukuo, with the story centered around four agents carrying out the Utrennya secret mission. As a representative of the new era of spy films, *Cliff Walkers* not only has a strong sense of the times and political undertones but also integrates fast-paced editing and high-density dialogues. The subtitle translation of this film demonstrates the translator's sensitive adaptation to language and culture. Compared to ordinary dramas, the subtitle translation more prominently highlights how to achieve the best information presentation effect within the limited character space and complex semantic load. Therefore, this paper explores the subtitle translation of *Cliff Walkers* from the perspective of multimodal discourse analysis and further discusses how character images are constructed through the collaborative effect of multimodality.

2. Literature review

The multimodal analysis is a approach that incorporates non-verbal symbols used by humans into the process of meaning communication. These symbols include images, actions, eye contact, postures, and so on, and it also pays attention to the interrelationship between verbal and non-verbal symbols (Jewitt 2009).

In the 1990s, the publication of *The Language of Displayed Art* (O'Toole 1994) and *Reading Images: The Grammar of Visual Design* (Kress & van Leeuwen 1996) marks that multimodal discourse analysis (MDA) has entered a new stage of vigorous development (Jewitt et al 2016), further emphasizing the social nature of multimodal discourse and comprehensively establishing the multimodal analysis method of social semiotics. However, most of the relevant studies abroad focus on education and instructional design (Kress & Selander, 2012; Roohani & Hosseini, 2021), with only a few scholars discussing cultural dissemination. For instance, Dash et al. (2016) used MDA to explore how Indian television advertisements re-established new discourses and cultural identities between global and local audiences, in order to position and promote the brand in different cultural spaces. Hosoda & Aline (2021) paid more attention to the interactive role of multimodal discourse resources between tour guides and tourists, which aroused the latter's interest in local culture and ultimately facilitated cultural dissemination. It can be seen that the integration and interaction of modalities enhance cultural diversity, build the cultural identity of the audience, and ultimately promote cultural exchange and dissemination.

The research on multimodal discourse in China started relatively late, but has developed rapidly in the later stage. Many scholars have successively conducted theoretical exploration around the definition, research methods and research significance of multimodal discourse (Li Zhanzi & Lu Danyun 2012; Huang Lihete 2015; Zhang Delu & Zhao Jing 2021), and clarified that multimodal discourse originates from human perception of different modalities in communication, composed of multiple symbol systems in social practice, and its theoretical basis is closely related to systemic functional linguistics and semiotics. At the same time, the significance of MDA in cultural dissemination has become increasingly prominent. Lan Jie et al. (2017) focused on exploring the multi-modal construction of Xinjiang's external publicity forms and contents and proposed corresponding translation strategies; Wu Jianguo et al. (2021) analyzed the multi-modal relationships in documentary external publicity to explore how translators can use multi-modal translation strategies to construct the national image. It is easy to see that MDA enjoys a relatively wide research field in China, especially playing a key role in the dissemination of Chinese culture.

One of the most significant differences between subtitle translation and traditional text translation lies in the fact that it often involves multiple symbols coexisting. In the process of subtitle translation, the information of one symbol is ultimately very likely to be presented in the form of another symbol (Taylor 2004). Many studies have analyzed the multimodal characteristics of subtitle translation (Chuang 2006; Gottlieb 1992&2005; Chaume 2004), arguing that subtitle translation must be combined with the multimodal characteristics of the audio-visual text. However, previous studies on the multimodal aspects of subtitle translation have mostly focused on how multiple symbols interact in static images, while neglecting the analysis of subtitle translation in dynamic images. The dynamic images in audio-visual texts are composed of static images, but they are not the simple addition of the meanings of static images; instead, they reflect the logical connections between static images. Film and television subtitle translation is not only related to the content of static images, but is also closely related to the logical relationships between these static images.

Theoretical framework--Visual Grammar

Among the main research paths of multimodal discourse, sociolinguistics places great emphasis on the social significance of symbols, arguing that the way symbol users utilize symbols reflects their attitudes and purposes. Visual grammar is the main theory within sociolinguistics. This study adopts the visual grammar theory as the framework and serves as the primary reference indicator for analyzing the multimodal discourse in *Cliff Walkers*. Kress & Van Leeuwen (1996) systematically analyzed various visual products such as graphics, images, and symbols, and proposed the corresponding concepts of visual grammar. This visual grammar framework is based on Halliday (1994)'s systemic functional linguistics and is modeled after the three metafunctions in systemic functional linguistics, namely the ideational function, the interpersonal function, and the textual function. Among them, representational meaning corresponds to the ideational function, interactive meaning corresponds to the interpersonal function, and compositional meaning corresponds to the textual function.

3.1 Representational meaning

Representational meaning consists of two processes: one is the narrative process, and the other is the conceptual process. The narrative process describes the implementation process of a certain action or the dynamic presentation in the picture. The conceptual process is characterized by its static nature. When viewers observe this process, they mainly consider the analytical process, classification process, and symbolic process of the picture elements. Representational meaning presents the identity, behavior, and relationships of the characters through the narrative function of visual symbols.

3.2 Interactive meaning

Interactive meaning is defined as the interaction between the viewer and the image creator. Interactive meaning is mainly composed of the following three indicators: contact, social distance, and power relation, which collectively express the interaction relationship between the participants and the viewer. Interactive meaning guides the audience through visual language and sound to establish the psychological distance and attitude towards the characters.

3.3 Compositional meaning

Compositional meaning refers to the entire set of elements that collectively convey the complete meaning of an image. Compositional meaning consists of three elements: information value, salience, and framing. These elements focus on the distribution of information points and key elements in the image. Compositional meaning guides the viewer's attention and interpretation towards the characters through the organization of the image elements.

3. Analysis

3.1 Analysis on representational meaning

3.1.1 Narrative process

The narrative process connects the participants in the scene through vectors, expressing the content of the scene, and also revealing some of the plot and character relationships of the film. For example, in this scene from *Cliff Walkers*:



Figure 1

Firstly, from a linguistic perspective, the translation “rat out” for this line adopts an informal English expression meaning “to expose”. This translation has a more authentic spoken language tone. Although “rat out” is not a formal term, it is precise in meaning and clear in context, adhering to the principle of efficient information transmission. From the perspective of ecological translation studies, this translation caters to the cultural understanding of the target audience regarding spy-related plots. “Rat out” is an idiomatic term in the American spy drama context, which enhances the authenticity of the context in the translation. In terms of communication, the translation not only accurately conveys the original meaning but also strengthens the narrative theme of the film through language choices.

From the perspective of visual grammar, this line of dialogue forms a profound intertextual relationship with the scene, jointly depicting the survival rules and interpersonal relationships of the underground workers in that particular era. This dialogue constructs a clear logical vector through the conditional sentence “if one of us gets caught”, closely linking the two key actions of “being caught” and “not betraying”, creating a necessary causal relationship. The use of the personal pronouns “us” and “we” creates a closed group vector, incorporating the speaker and the listener into the same fate community, while forming an opposition with “others”. This linguistic vector relationship is often manifested on the screen as the close composition of the characters, their lines intersecting with each other, forming an interactive process of solicitation and reaction. When the speaker gazes at other members, this line vector not only conveys information but also is inquiring for an unspoken commitment. More importantly, this opening dialogue actually sets the narrative tone for the entire film. It is both a rule and a prophecy, foretelling the life-and-death tests that the characters will face in the subsequent plot.

3.1.2 Conceptual process

The symbolic meanings of snow and Utrennya in *Cliff Walkers* reflect the construction of the conceptual process.



Figure 2

It is necessary to clarify here the historical background of the story. In the early stage of the War of Resistance against Japanese Aggression, the liberated areas were under the double suppression of Japan and the Nationalist Government. Under the limited conditions, people translated and introduced many literary and theoretical works mainly from the Soviet Union. Utrennya is not only the name of the action but also symbolizes the ultimate goal of faith, hope and sacrifice. This subtitle treatment demonstrates the adaptability to the cultural-loaded language. Utrennya is a variant of the Russian word утро, and it involves the understanding of Utrennya among different ethnic groups. In Chinese, it is translated as dawn, and in China, dawn represents hope and a vibrant atmosphere. While dawn in English also means the beginning stage and becoming clear, the imagery expression here does not require excessive supplementation or alteration because the meaning of this image is almost the same in any language. Under the condition of semantic equivalence, direct translation is sufficient.

In this snowy night dialogue, Utrennya and the snowflakes form a profound symbolic opposition through a multimodal symbol system. The snow keeps falling, but it also stops. Since the group of agents landed at the destination shown in the opening scene, the snow has never stopped, and the snowfall has not been reduced. The picture has been dominated by cold tones such as black, white and gray, with a majority of black and white scenes, expressing that the agent team is in a dark and perilous situation, and the tension, coldness and oppression of the mission are overwhelming. When Zhang Xianchen asks about the meaning of Utrennya, the violent snowstorm outside the window occupies the picture with overwhelming visual power, just like the spirit portrayal of revolutionaries holding onto the light in the darkness. The low-pitched piano sound of the background music is like the roaring cold wind, symbolizing the confrontation between darkness and dawn in the auditory level. This audio-visual coordination forms a complete conceptual system. The oppressive existence of the snow symbolizes the current perilous struggle environment, while Utrennya represents the future dawn that pierces through the darkness.

The core part of the film's development is precisely the Utrennya operation carried out by the agent team. Utrennya is translated in Russian as dawn, and it also symbolizes the successful completion of the mission, the phased victory of the revolution, and the peace and light that revolutionaries are all looking forward to, as Zhang Xianchen said: "It would be better if it were dawn." This snowy night dialogue, through the collaborative effect of language, picture and sound, elevates the natural phenomenon to a spiritual image, giving the process of waiting for dawn a powerful and touching ritualistic feeling. Under the framework of visual grammar, the relationship between snow and dawn not only drives the narrative but also shapes the unique aesthetic characteristics and philosophical depth of the film.

3.2 Analysis on interactive meaning

Interestingly, at the end of the film, there is the same dialogue as in Figure 2, both asking about the meaning of Utrennya. However, the language processing method, visual effects, and background music are all different.

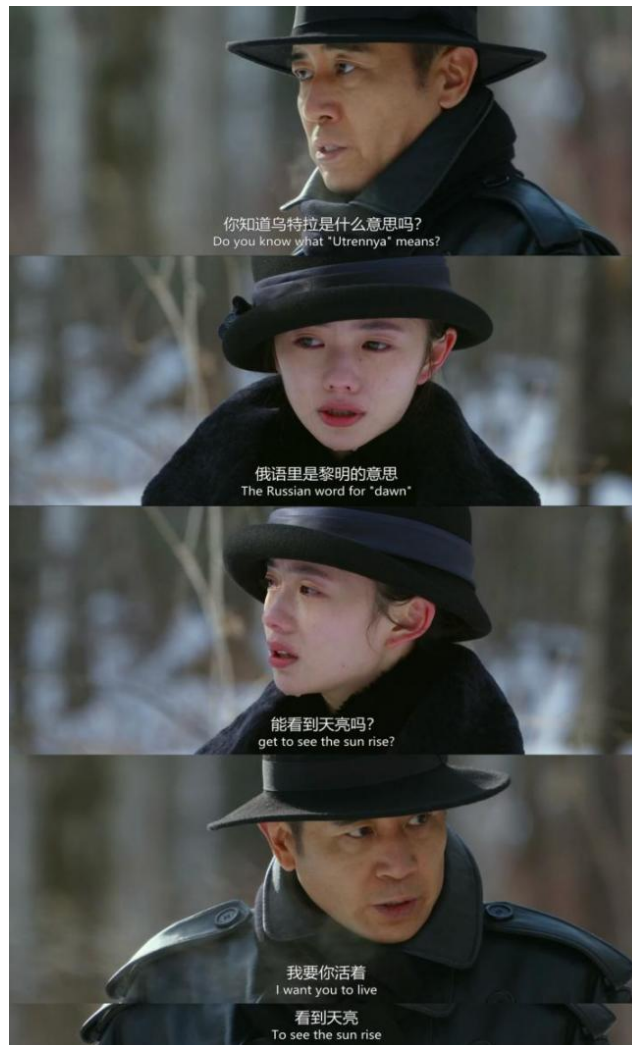


Figure 3

3.2.1 Contact

The first Utrennya dialogue took place on a snowy night. Zhang Xianchen and Xiaolan stood side by side but did not look at each other. Their gazes both turned towards the endless darkness outside the window. This visual indirect contact belongs to the "providing" type of image. The characters not only avoided establishing direct eye contact with the audience but instead directed the audience's attention towards the snowstorm and night outside the window, which symbolize danger and the unknown. This handling implies that their conversation transcends individual emotions and becomes the transmission of a more symbolic collective belief. Especially when Xiaolan asked, "Can we see the dawn?" Utrennya's action was first closely linked to the collective life experience of "us". And in the ending part, Zhou Yi, with a close-up shot, faced Xiaolan directly, forming a strong "demand" type of image. He looked straight at the camera as if breaking the fourth wall, and the gaze with Xiaolan formed a close visual interaction, forcing the audience to accept the emotional instruction of "I want you to live". This transformation from indirect expression to direct gaze not only strengthened the emotional tension but also demonstrated how revolutionary beliefs were concretized from grand ideals into deep concern for individual lives.

3.2.2 Social relation

In terms of creating social distance, the director skillfully uses camera language to regulate the psychological distance between the audience and the characters. The medium shot used in the first dialogue maintained a moderate social distance, neither overly close nor deliberately distant, allowing the audience to observe calmly as bystanders, thereby enabling them to

more profoundly feel the mutual belief in the dawn that does not require words, between the characters. In the final scene, Zhou Yi's body slightly leans forward, and the camera zooms in to a close-up. The social distance is compressed to an intimate range, and the originally abstract dawn is transformed into a solemn commitment to the individual life of Xiaolan. The background music also quietly changes, no longer being the grand melody of idealism, but focusing on the dignity and strength of the continuation of individual life. At this moment, the sunlight is gentle, the snow reflects bright but not dazzling high-key light, and the faces of the characters are enveloped in delicate shadows. This visual composition establishes a high-toned image, making the audience strongly empathize with the narrative of "survival is victory", and completing the visual elevation from the collective mission to the value of individual life.

3.3 Analysis on compositional meaning

Under the visual grammar framework proposed by Kress & Van Leeuwen (1996), the information value, salience, and framing jointly constitute the multimodal semantic system of the image. The following analyzes the symbolic interaction of the title scenes of the *The Shadow*, *The Danger*, and *The Edge* chapters in *Cliff Walkers* from these three dimensions:



Figure 4

3.3.1 Information value

In the three scenes, the title text occupies the visual center and becomes the dominant of the information transmission. The background elements serve as supporting information to provide contextual support.

The Shadow

The background features a close-up of Zhang Xianchen and the torture device. The characters' scars and tattered clothes convey the individual's pain and secrets. The central title *The Shadow* elevates the individual's predicament to a secret game.

Zhang Xianchen's body becomes the carrier of the shadow. The translation of "shadow" essentially involves the cross-linguistic reconfiguration of the image. It abandons the literal translation of "Bottom Card" and instead grasps the core contradiction of concealment and exposure, light and darkness, incorporating the secret game of espionage and the struggle of human nature into the concise metaphor of shadow. This translation strategy enables the language to break through the boundaries of symbols and form a synergy with the visual narrative and atmosphere creation of the film, ultimately outlining in the audience's minds the most lethal game rule in the espionage world, that is, the secrets in the shadow are the decisive cards that determine life and death.

The Danger

In Chinese culture, a chess game can represent confrontation. There is an idiom "chess meets its match", indicating that two people have equally excellent chess skills. A chess game often requires ingenious planning and clever traps. Therefore, the original version *The Danger* can be used to indicate that the action is crucial and the situation is tense, but directly translating it as chess would prevent the audience from directly experiencing the connotation of the Chinese word "棋". To avoid misunderstandings and mistranslations of cultural imagery, the translator directly translated it as *The danger*, summarizing that this plot is tense and the main characters are all in danger. By converting the cultural imagery of the original text, the target language audience can feel the meaning of the imagery.

The background is a city street scene covered with snow, with wires crisscrossing like the lines of a chessboard. The central title integrates the complex environmental information into a metaphor of the situation like a chess game, highlighting the severity of the situation due to the complexity of the background space.

The Edge

The background depicts a car driving in a snowy night, with the dynamic blurred streets and car lights creating a tense atmosphere. The central title *The Edge* condenses the action into an image of an edge-like decision, and the vehicle's movement trajectory becomes a visual extension of the critical point of life and death. From the plot perspective, the agents are already in a dangerous situation, and the identities of each of them are unclear. The outcome of the following task is undetermined. Therefore, the Chinese title "生死 (Life and Death)" was used in this section, as the Chinese idiom often uses phrases like "生死未卜" or "生死难料" to indicate that the outcome of life and death is difficult to determine. The titles of each section, whether in Chinese or English, are symmetrical and profound in meaning. So, to maintain consistency in format and respect the original content, the translator translated it as *The Edge*, ingeniously explaining the meaning of life and death, and enabling the target language audience to receive the information conveyed by the original text.

3.3.2 Salience

Salience refers to the way elements attract the visual attention of the audience through size, color, focus, etc. From the perspective of color, the titles of the three chapters all adopt bright blue, creating a visual contrast against the cold-toned background. The snowy scene in *The Danger*, the dark cell in *The Shadow*, and the dim night lamp in *The Edge*, all with low saturation to highlight the brightness and sharpness of the blue titles, echoing the sense of tension. From the perspective of size and position, the title fonts are large and centered, occupying the visual center of the picture, overwhelming the background elements in terms of area and position, and strengthening the authority of the information.

3.3.3 Framing

The shooting method and elements form a metaphorical spatial narrative.

The chapter title images of *The Danger* adopt wide panoramic shots, with wires crisscrossing like a spider web and buildings forming enclosed spaces. The title is surrounded by wires, suggesting being caught in a network of multiple forces. The wide expanse of the frame enhances the complexity of the situation.

The background of the chapter title of *The Shadow* uses a close-up imprisonment, intensifying the imagery of restraint. The frame focuses on the prisoner and the torture device, with the wooden frame of the torture device firmly binding the person. The title hangs above the figure, seemingly a judgment on the shadow, and the closed nature of the frame echoes the oppressive feeling of being crushed.

The background of the chapter title of *The Edge* is a dynamic scene. The frame captures the side view of a car, with a blurry background and the dynamic blur of the car body. The title implies that the vehicle is heading towards the edge of life and death, and the fluidity of the frame intensifies the sense of urgency.

4. Findings and discussion

This study applies the framework of multimodal discourse analysis, specifically Kress and Van Leeuwen's visual grammar, to examine the subtitle translation of the film *Cliff Walkers*. The analysis reveals that subtitle translation is not merely a linguistic transfer but a complex process of multimodal meaning reconstruction, involving the dynamic interaction among language, image, sound, and other semiotic resources.

The findings demonstrate that the translator of *Cliff Walkers* effectively utilizes multimodal resources to achieve cross-cultural communication. For instance, in the translation of culturally specific terms such as *Utrennya*, the translator adopts a direct translation strategy to preserve the symbolic meaning, which is further reinforced by the visual and auditory elements—snow, darkness, and somber music—to construct a conceptual metaphor of hope amidst peril. Similarly, the translation of chapter titles (*The Shadow*, *The Danger*, *The Edge*) reflects a conscious effort to convey the thematic tension and narrative progression through concise language, while maintaining coherence with the visual composition of each segment.

Moreover, the analysis of interactive meaning highlights how shifts in camera angles, social distance, and character gaze influence the audience's emotional engagement and interpretation of the subtitles. The transformation from indirect "providing" images to direct "demand" images in the two *Utrennya* dialogues exemplifies how multimodal coordination enhances the emotional and ideological impact of the dialogue.

These findings suggest that successful subtitle translation in multimodal contexts requires translators to go beyond textual equivalence and engage deeply with the audiovisual narrative. The synergy between verbal and non-verbal modes not only facilitates comprehension but also enriches the cultural and emotional resonance of the film.

However, this study also acknowledges certain limitations. The analysis is primarily qualitative and focused on selected scenes, which may not fully represent the entire film. Future research could incorporate more quantitative methods, such as audience reception studies, to evaluate the effectiveness of multimodal subtitle translation from the perspective of target viewers.

Conclusion

This paper has explored the subtitle translation of *Cliff Walkers* from a multimodal discourse analysis perspective, using visual grammar as the theoretical framework. The study confirms that subtitle translation is a multimodal practice wherein linguistic choices are intimately connected with visual, auditory, and narrative elements. The translator's adaptive strategies—such as the use of culturally resonant expressions, thematic condensation in chapter titles, and alignment with visual and auditory cues—demonstrate a sophisticated understanding of how meaning is constructed across modes. These strategies not only ensure informational accuracy but also enhance the expressive power and cultural accessibility of the film.

In conclusion, this study contributes to the growing body of research on multimodal subtitle translation by illustrating how visual grammar can serve as an effective analytical tool. It underscores the importance of integrating multimodal awareness into translation training and practice, particularly in the context of globalizing Chinese cinema. As audiovisual media continue to evolve, the role of the translator as a multimodal mediator will become increasingly vital in bridging cultural and semiotic gaps.

Future studies may expand this research by comparing multimodal translation strategies across different genres or cultural contexts, or by employing eye-tracking technology to examine how viewers process multimodal subtitles in real time.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1]. Jewitt, C. (2009). An Introduction to Multimodality. In C. Jewitt (ed.). *The Routledge Handbook of Multimodal Analysis*. London/New York: Routledge.
- [2]. Taylor, C. J. (2004). Multimodal Text Analysis and Subtitling. In E. Ventola, C. Charles & M. Kaltenbacher (eds.). *Perspectives on Multimodality*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- [3]. Chuang, Y. T. (2006). Studying Subtitle Translation from a Multimodal Approach. *Babel*. (4), 372-383.
- [4]. Gottlieb, H. Subtitling: A New University Discipline. (1992). In C. Dollerup and A. Loddegaard (eds.). *Teaching Translation and Interpreting: Training, Talent, Experience*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- [5]. Chaume, F. (2004). Film Studies and Translation Studies: Two Disciplines at Stake in Audiovisual Translation. *Meta*. (1), 12-24.
- [6]. O'Toole, M. (1994). *The Language of Displayed Art*. London: Leicester University Press.
- [7]. Dash, A. K., P. (2016). Patnaik & D. Suar. A multimodal discourse analysis of globalization and cultural identity in three Indian TV commercials. *Discourse & Communication*. 10(3), 209-234.
- [8]. Hosoda, Y. & D. Aline. (2021). Establishing joint attention with multimodal resources in lingua franca guided tours. *Learning, Culture and Social Interaction*. 31(6), 1-20.
- [9]. Jewitt, C., J. Bezemer & K. O'Halloran. (2016). *Introducing Multimodality*. Milton Park: Routledge.
- [10]. Kress, G. & S. Selander. (2012). Multimodal design, learning and cultures of recognition. *The Internet and Higher Education*. 15(4), 265-268.
- [11]. Kress, G. & T. van Leeuwen. (1996). *Reading Images: The Grammar of Visual Design*. London: Routledge.
- [12]. Roohani, A. & M. Hosseini. (2021). Analysis of culture and multimodality in a local English textbook series. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*. 9(2), 171-193.
- [13]. Halliday, M. A. K. (1994). *An Introduction of Functional Grammar (2nd edition)*. London: Edward Arnold.
- [14]. Kress, G. & Van Leeuwen, T. (1996). *Reading Images: The grammar of visual design*. London: Routledge.
- [15]. 董梅 & 袁小陆. (2021). 多模态审美批评话语分析框架建构研究. *外语教学*. (1), 77-82.
- [16]. 黄立鹤. (2015). 语料库4.0:多模态语料库建设及其应用. *解放军外国语学院学报*. (3), 1-7+48+161.
- [17]. 兰杰 & 谢红海 & 王晨乐. (2017). 区域形象视域下新疆多模态外宣体系构建. *新疆社会科学*. (4), 136-141.
- [18]. 李战子 & 陆丹云. (2012). 多模态符号学:理论基础,研究途径与发展前景. *外语研究*. (2), 1-8.
- [19]. 武建国. (2023). *互文性研究*. 北京: 北京大学出版社.
- [20]. 武建国 & 李育静. (2024). 多模态语境重构与中国影视文化的传播——以影片《我和我的祖国》字幕翻译为例. *山东外语教学*, 45(02), 11-21.