Research Journal of English Language and Literature (RJELAL)
A Peer Reviewed (Refereed) International Journal
Impact Factor 6.8992 (ICI) http://www.rjelal.com;
Email:editorrjelal@gmail.com; ISSN:2395-2636 (P); 2321-3108(O)

Vol.13.Issue 3. 2025
(July-Sept.)

RESEARCH ARTICLE

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA
2395-2636 (Print):2321-3108 (online)

Subtitle Translation of *MAO ZEDONG 1949* by Multimodal Discourse Analysis Framework and Intersemiotic Shift

Li Sun[1] & Ao Lin[2]
[1]Associate Professor and MA supervisor, School of Foreign Languages, North China Electric Power University, Beijing, China. Email：sunli@ncepu.edu.cn
[2]MA Candidate, School of Foreign Languages, North China Electric Power University, Beijing, China.

DOI: 10.33329/rjelal.13.3.406

Article info
Article Received: 20/08/2025
Article Accepted: 08/09/2025
Published online: 12/09/2025

Abstract

Multimodal translation requires the effective integration of various modes. However, the precise methods for achieving this cooperation remain a key challenge in the translation process. This paper addresses this issue by proposing and applying an integrated framework that combines Zhang Delu's multimodal discourse analysis with Feng Dezheng's intersemiotic shifts. Using the subtitle translation of *MAO ZEDONG 1949* as a case study, this analysis aims to identify effective strategies for future film translation activities. The findings demonstrate that the multimodal discourse analysis framework can serve as a valuable guide for the entire translation process, providing a justifiable rationale for use specific choices, while the intersemiotic shift fills the choice category. When translating the cultural loaded factors, the Compensation method or Compensation with Addition tends to help the translator to convey the meaning with the help of other modes, and the cultural loaded factor shall also be explained on the side of the image or activity; when translating the ST involving the particles, the Omission+ Addition can be useful. And when dealing with the visual language like flag language, the Typographic Transformation is adoptable.

Keywords: multimodal translation; multimodal discourse analysis framework; intersemiotic shift

1.Introduction

For years, the principles of effective communication and international dissemination have been a subject of extensive discussion, spanning disciplines from communication studies and sociology to translation theory. In recent years, as China's global influence has grown, the nation has increasingly prioritized its "going global" strategy. In this context of global cultural exchange, the translation of prominent works is a crucial method for promoting Chinese culture abroad. Film, as a significant medium, plays a vital role in showcasing China's unique character and profound spiritual heritage. Therefore, careful attention to all aspects of film-making is essential, with subtitle translation emerging as a key area of study. As a sub-field of audiovisual translation, subtitle translation is defined by Díaz Cintas and Remael (2007) as consisting of the presentation of a written text, generally on the lower part of the screen, that endeavors to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, inserts, graffiti, inscriptions, placards and the like), and the information that is contained on the soundtrack.

The burgeoning body of eye-tracking research on the behaviour of viewers' gaze shows that the on-screen mobilisation of subtitles always draws attention from viewers, even when audience members are able to understand the dialogue they are presented with (d'Ydewalle and De Bruycker 2007) This demonstrates the importance of subtitles, which implies that meticulous attention must be paid during the film translation process.

Film subtitle translation serves not only as a linguistic conversion but also as a crucial bridge for international communication. The quality of these translations directly shapes foreign audiences' understanding and perception of Chinese films and, by extension, Chinese culture. Consequently, conducting systematic and in-depth research on the subtitle translation of Chinese films is of profound practical and academic significance for cultural dissemination.

This paper aims to analyze the subtitle translation of *MAO ZEDONG 1949* through a multimodal approach. By applying an integrated framework that combines Zhang Delu's multimodal discourse analysis and Feng Dezhen's intersemiotic shifts, this study argues that the film's subtitle translation successfully utilizes specific strategies to effectively convey cultural nuances. This research will conclude with practical suggestions to inform future film translation endeavors.

1.1 MAO ZEDONG 1949

MAO ZEDONG 1949 is a film that is approximately 2 hours and 20 minutes in length. It vividly recreates the historical events of 1949, when the Central Committee of the Communist Party of China relocated to Xiangshan and actively promoted peace talks with the Kuomintang. Despite the eventual failure of these talks, the Central Committee made a swift and decisive choice to launch the Crossing-the-Yangtze Campaign, a pivotal event that paved the way for the founding of the People's Republic of China.

Through its nuanced character portrayals and compelling plot, the film comprehensively depicts the struggles and triumphs on the eve of a new era for China. This powerful narrative makes the audience feel as if they are experiencing that tumultuous period firsthand. Notably, the film's subtitle translation demonstrates the intricate interaction and synergy between multiple modes, making it a valuable subject for scholarly analysis. To fully explore this, a clear understanding of the theoretical underpinnings of multimodality and multimodal translation is essential.

The film effectively integrates various modes, particularly through the profound synergy of the visual and auditory channels. In the practice of subtitle translation, this interplay among modes is indispensable. The accurate conveyance of many translated examples often relies on the detailed information within the image, which is crucial for transmitting the source language's intended meaning. The fifth part of this article will conduct an in-depth analysis of specific cases to further reveal the internal logic and operational mechanisms of modality interaction within the subtitle translation process.

1.2 Multimodality and Multimodal Translation

Multimodality has become increasingly significant in both social and daily life. Communication frequently involves more than just spoken or written language; it also incorporates gestures, printed text, and even music. Consequently, the translation of this complex type of discourse has emerged as a crucial field of study.

1.2.1 Multimodality and Modes

Kress and van Leeuwen (2001) define multimodality as "We have defined multimodality as the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined". Based on this general idea, many scholars have conducted numerous researches in this field. According to Zhang and Feng (2020) And this "semiotic product or event" is called multimodal discourse. Multimodal studies are concerned with the production, dissemination and reception of multimodal discourse from a wide range of perspectives. Matthew Reynolds also has suggested that the term "multimodality" does not name a multitude of entities called "modes" that are separate and countable. Rather, it points to an ever- varying continuum of resources for meaning- making that can be divided up in different ways. (Boria et al., 2020) From this point of view, a fair opinion can be derived that the modes that work in the multimodal discourse should be regarded as a whole system and every subordinate semiotic elements forms an organic entity. A feasible way to conduct reliable work is to probe into the meaning of the united whole and the individual semiotic components' function and relations.

As it is said above, "modes" serve as a very key concept in it. It seems that modes function as lines and cross each other to print a beautiful work. Thus, it is necessary to understand what the term "mode" means. Kress (2010) believes "Modes are semiotic resources which allow the simultaneous realisation of discourses and types of (inter) action." In the work Kress collaborated he also suggested that"mode as material shaped in the history of social and semiotic work"(Boria et al., 2020) Zhang and Feng (2020) has summarized that There are two different meanings of"mode"that are currently in use: (1) multimodal texts and artefacts combining the use of various semiotic modes such as language, images, gesture, typography, graphics, icons or sound; and (2) semiotic modes that are transmitted via different perceptual modes (i.e. sensory modes), namely visual, auditory, haptic, olfactory and gustatory perception. Based on that, we may easily come up with the idea of mode.

1.2.2.Multimodality Translation

Multimodality is considered a resource and meanwhile a challenge for translation scholars (O'Sullivan 2013).

The term "multimodality translation" is not a completely new concept. The linguist Roman Jakobson famously categorized three types of translation: interlingual (between languages), intralingual (within the same language), and intersemiotic (between different sign systems). The last of these, intersemiotic translation, involves converting a sign from one semiotic system into another. For example, translating words into music or visual signs would be considered an intersemiotic translation.

A multimodal understanding of translation requires that not only language and its functions/functioning, but also other central modalities such as music and image as well as the associated submodalities must be examined with regard to their functions and cultural specificity--and, in their different ways,(Boria et al., 2020) Thus, basic research frame is formed. The analytical work at least integrates the mode, its cultural and social factors(according to Kress's definition), and its function.

Multimodal translation is a kind of cross-mode translation. This will meet some doubts when realize one mode to another. Thus, there must be some kind of compromise or change at various degrees. Klaus Kaindl has offered a case of a German translation of the opera *Carmen* (premiere in Paris in 1875 at the Opéra Comique) . The Court Opera is a totally new environment for the play. There must be some changes. To overcome this problem, the translator, Julius Hopp, modified his text to suit the conventions of the Romantic opera, which was reflected in the musical realisation and the relationship between spoken and sung modes. The orchestra was extended, the spoken passages were deleted, and the singing style was adapted to the conventions of the Romantic opera. (Boria et al., 2020)

Building on the concept of intersemiotic translation, the changes that occur during mode transfer are multifaceted. The cultural and social acceptance of the target audience becomes a critical consideration in this process. Consequently, a new mode may need to be added, a mode may be removed, or the form of a mode may be altered. This raises a new area of inquiry: the rationale behind transferring certain modes to others. Analyzing this question allows for the identification of the specific functions and relationships of these modes.

Klaus Kaindl also pointed out that "A transfer is only possible if there are enough similarities between modes for them to be comparable and, thus, transferable, and if there is a sufficient difference that makes the transfer necessary. Therefore, translation gives us an insight into what modes have in common and what differentiates them."(Boria et al., 2020)

In the past few decades, both China and abroad have conducted several kinds of studies Ketola (2016), worked out a framework for the cognitive dimension of multimodal translation; Boria(2020) has conducted research to offer a possible multimodal translation methodology to deal with different kinds of genres, including literature, music, and dance. Pérez-Gonzále（2014） aims to integrate multimodality into translation and interpreting studies, Baldry and Thibault（2006） come up with a transcription framework for the multimodal analysis, which has inspired both linguistic studies and translation projects.

At present, most of the multimodal translation in China is carried out under the framework of multimodal discourse analysis.

Zhang Delu (2009) in *Exploration of a Comprehensive Theoretical Framework for Multimodal Discourse Analysis* distinguished the modal systems and systematically sorted out the forms and relationship characteristics of multimodal discourse. He proposed a theoretical framework for multimodal analysis, emphasizing that the analysis process should start from four aspects: Culture, Context, Content and Expression. Zhang Delu's analytical framework affirmed the role of culture. As mentioned earlier, many linguists and translation scholars have affirmed the social and cultural factors of mode. Zhang Delu also analyzed the social factors as a very important part. Social factors will be reflected in the interpersonal function in certain situations. For example, when expressing power relations, there will be corresponding reactions in the tenor, field and mode aspects. Therefore, Zhang Delu also included it in the scope of multimodal discourse analysis. Multimodal expression is a combination of verbal and non-verbal semiotic mode. Their relationship is not isolated, and their roles are intertwined, thus forming a certain relationship that involves the Content aspect. The expression of modality requires certain carriers, such as language, body movements, facial expressions, etc. Therefore, these factors must be included in the investigation, and Zhang Delu classified them under the Expression aspect.

Feng Dezhen (2017) in *Discussion on Basic Issues of Multimodal Discourse Analysis* explained why to study multimodal discourse, the methods and scope of study, providing a basis for the legitimacy of multimodal discourse analysis and thus laying the foundation for multimodal translation. In his work (2020) *Multimodal Approaches to Chinese-English Translation and Interpreting*, he has revised the "semiotic shifts" proposed by Delabastita. He has revised and added some parts while changing the "semiotic shifts" to "intersemiotic shift". The previous frame includes three parts, namely, Adiectio, Detractio, and Substitutio. While in his work, there are five types, namely, Addition, Omission, Omission+Addition, compensation Typographic Transformation. In this paper, this frame will be adopted to present the relation between the different mode.

The subtitle translation of *MAO ZEDONG 1949* involves the complex interaction of different modes, which provides a strong foundation for a multimodal translation analysis. The decision to apply this specific analytical approach is further supported by the relative maturity of Zhang Delu's framework and the operational utility of Feng Dezheng's intersemiotic shifts. Therefore, this paper will integrate these two frameworks to conduct a comprehensive study. Specifically, it will adopt the four analytical aspects proposed by Zhang, while incorporating Feng's framework into the detailed explanations to demonstrate the precise mechanisms of how different modes interact.

2.The Multimodal Discourse Analysis Framework

Zhang Delu (2009) posited that systemic functional grammar is a highly effective framework for exploring multimodal discourse without modification. He adopted and integrated Martin's five levels, proposing a comprehensive analytical framework that is structured around four distinct but interconnected aspects: Culture, Context, Content, and Expression (see Figure 1). While each level is realized in its own way, they are not separate; rather, their interconnectedness forms the basis of his proposed framework, as detailed below.

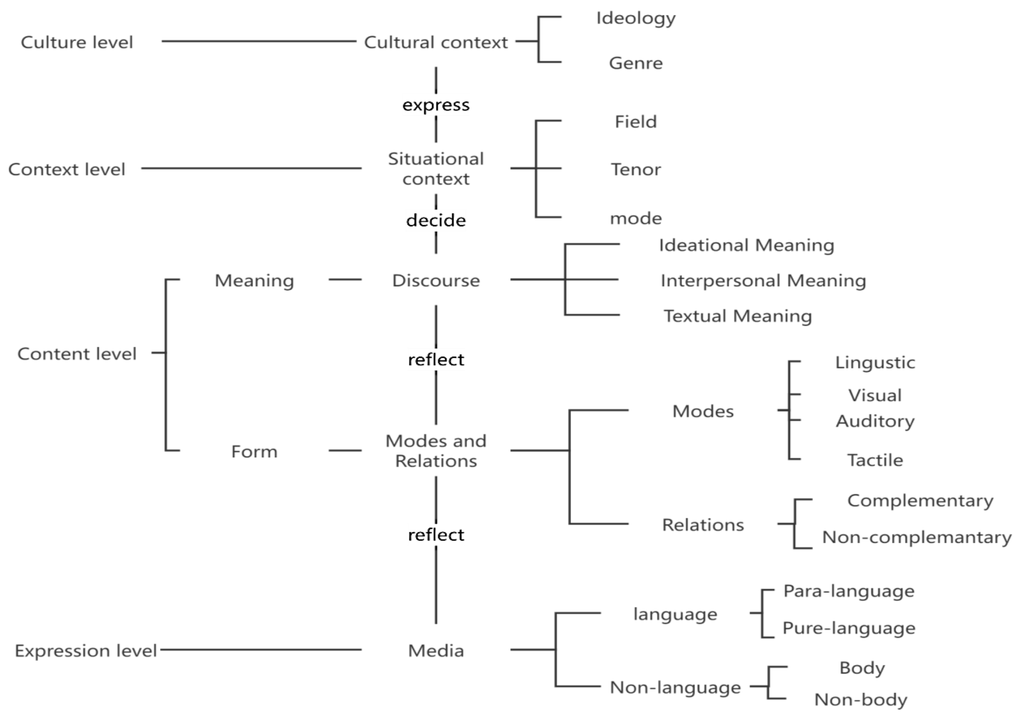


Figure 1. Multimodal Discourse Analysis Framework (Zhang, 2009)

2.1 Culture Level

The Culture level forms the foundational stratum of multimodal communication, realized through ideology and genre. As Zhang Delu (2009) establishes, this level governs communicative traditions, formats, and techniques while enabling contextual interpretation. It encompasses ideology – comprising cognitive patterns, behavioral norms, social conventions, and implicit societal codes – alongside genre, which manifests as communicative structures and procedural potentials that operationalize ideological frameworks. Crucially, this dimension constitutes the essential basis for generating and interpreting situational meaning. Without cultural anchoring, contextual understanding risks ambiguity or misrepresentation, potentially obscuring a discourse's profound significances.

It is evident that the cultural level serves as the fundamental basis for the generation and interpretation of meaning within situational contexts. Without the support of a cultural framework, the meaning of specific situations may become ambiguous, risking misinterpretation and making it difficult to grasp the deeper significance of the discourse.

When translating for audiences from diverse cultural backgrounds, translators must handle relevant content with greater caution. Their goal should be to accurately convey the deeper meaning and ensure that the target audience can fully comprehend and connect with the cultural connotations, thereby achieving effective cross-cultural communication and emotional resonance.

2.2 Context Level

Context level can be regarded as the situational context that realizes the cultural context. Zhang(2009) believes that in the specific context, communication will be restricted by the context factors which includes field, mode, and tenor.(based on the systematic functional grammar) Liao (1999) explains that field means the happening social activities which influences the choice of words and syntax; mode represents the social and role relationship between the participants which affects the choice of sentence structures and tone of communication; and tenor refers to the function of speech. It can be verbal or written style. It can be homiletic or an inferential manner. The tenor impacts the cohesion of the discourse.

The situational context lays the foundation and sets the basic features for communication while plays a decisive role in the meaning conveyed by the speech.

2.3 Content Level

The Content level is bifurcated into a meaning level and a form level (as depicted in Figure 1). The meaning level conveys the discourse's meaning, but it is not an isolated component. Instead, it is realized through specific forms and the relationships between them, which constitute the core of the form level.

In terms of form, Zhang considers multimodality to involve linguistic, visual, auditory, and tactile forms. However, current multimodal scholarship often focuses on the visual, verbal, and auditory aspects due to various constraints. These limitations may be material-based; for instance, analyzing a book cover or a poster is unlikely to involve auditory or tactile factors. Technical limitations may also restrict research, as current films can rarely convey tactile sensations to the audience. Therefore, current research must provide sound records and explanations of the remaining modes to offer necessary resources for studying their interaction. Regarding the relationships between modes, Zhang (2009) identifies two clear types: complementary and non-complementary. A complementary relationship exists when one mode is insufficient to convey the full meaning and requires other modes to fill the gap. All other relationships are considered non-complementary. These two types of relations can be further analyzed in detail.

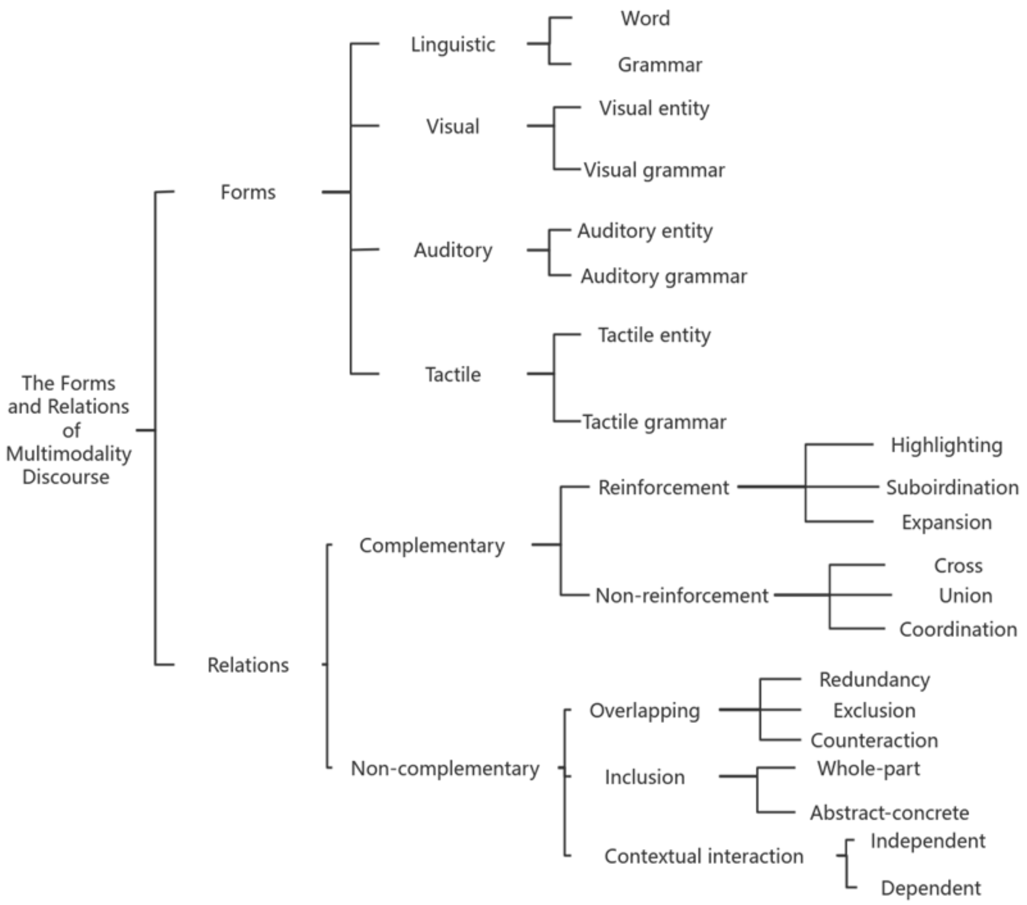


Figure 2. The Forms and Relations of Multimodal Discourse (zhang, 2009)

2.4 Expression Level

According to Zhang's framework, the Expression level is achieved by media, while the media encompasses of language and non-language media, which can be analyzed into a more detailed version as follow:

In multimodal discourse, both language and non-language media play a significant role in conveying meaning. Therefore, the collaborative intervention of elements such as music and body movements is crucial, as they can either reinforce or contradict semantic transmission. Consequently, in relevant research and translation practices, the functions of these media must be included in the scope of analysis.

In film subtitle translation, the complexity lies not only in the interaction among multiple modes but also in the multi-level and variable contextual factors that increase the difficulty of the task.

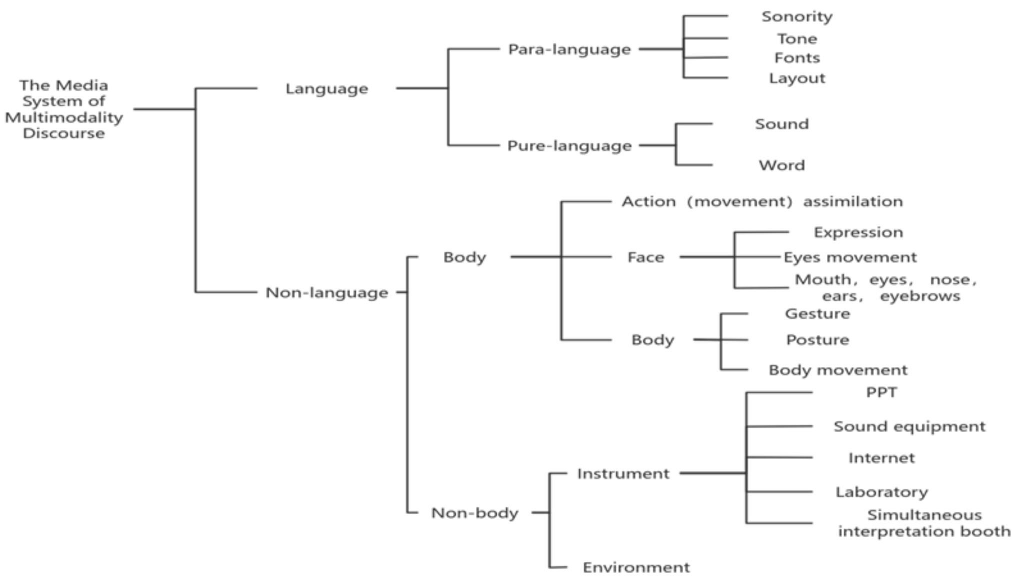


Figure 3. The media system of Multimodal Discourse

Based on these considerations, this paper will use Zhang Delu's multimodal analysis framework as its theoretical foundation. It will systematically analyze the film through the four core aspects of Culture, Context, Content, and Expression to deeply explore the intricate relationships among various modes in the subtitle translation process and, in doing so, reveal its internal mechanisms and characteristics.

3. Intersemiotic Shift

Intersemiotic shift is a more direct methodology for achieving cross-mode communication. It aims to integrate semiotic resources to facilitate the meaning-making process. This methodology is based on Delabastita's semiotic shift, but it has been revised by Feng Dezheng to better suit the needs of multimodal analysis.

3.1 Semiotic Shift

Delabastita is one of the first very few scholars who tries to deal with the various forms of translation taking place in the audio-visual communication. In his studies(1990), he found that "Translation acts like a'gatekeeper'and, accordingly, the shifts introduced by the dubbing process in the imported film material can be studied as evidence of the differences between the respective *Symbolmilieus* of source and target culture...Film translation is therefore not just a matter of language conversion, and the actual reality of film translation is conditioned to a large extent by the functional needs of the receiving culture and not, or not just, by the demands made by the source films." For this expression, a fair conclusion can be brought up that he has realized that translation is subtle a tailoring work. It needs to fit all kinds of restrictions and try to make the target language audience to accept the translation in overall aspects, including cultural, formal, social, and political factors.

When discussing the sound-image synchronicity in dubbing. He has noticed that "The pronunciation of different languages obviously has a different visual impact. These differences are caused not only by strictly phonological features, but also by the divergence between various paralinguistic and gestural patterns, such as facial expressions and body language."(Delabastita, 1990) This also suggests that he believes the para-language and other potential influencing semiotic factors should be carefully encompassed in the question of film translation.

When comes to discuss the semiotic nature of the signs in film, listed three relative features:

1. Film communication usually proceeds through two channels: the visual channel (light waves) and the acoustic channel (air vibrations).
2. There is a multitude of codes that shape a film into a complex meaningful sign: the verbal code (an aggregate of various linguistic and paralinguistic sub-codes), narrative codes, vestimentary codes, moral codes, cinematic codes.
3. As a product of mass communication, the transmission and manipulation of films and fragments of films is subject to certain technical restrictions. (Delabastita, 1990)

Then he put forwards several types of operation based on the classical rhetoric, namely *repetitio, adiectio, detractio, substitutio, and transmutatio.*

Delabastita(1990) defines that *Detractio*: visual and/or acoustic and verbal and/or non-verbal signs have been deleted (cuts); *Repetitio*: the film has been reproduced with all of its original material features; (in strict linguistic terms this would be a case of non-translation); *Adiectio*: new images, sounds, dialogue or spoken comments have been introduced; substitutio is the change of acoustic/verbal signs in dubbing. But he do not give a very accurate and specific definition to *transmutatio.*

3.2 Intersemiotic Shift

Feng(2020) has summarized Delabastita's mode consisting of three kinds of intersemiotic shifts, which are *adiectio, detractio, substitutio.* The reason for removing the *transmutatio* perhaps can be that even Delabastita himself hasn't offer a detailed and operable definition. As for the deletion of *Repetitio,* the reason can be relative obvious that a film is, after all, a kind of discourse. When translating a discourse, loyal to the original meaning can be regarded as a very key and universally acknowledged issue for the translators to obey. Thus, he summarized Delabastita's work to three types. "We argue that the model proposed by Delabastita, which consists of three kinds of intersemiotic shifts, is not sufficient to describe the translation process involving different kinds of modes."(Feng, 2020)

On the basis of Delabastita's model and finding the limitation, Feng revised a new model consisting of five types, namely, Addition, Omission, Omission+Addition, Compensation, and Typographic Transformation. Feng (2020)suggests that the *adiectio,* and *detractio* are less frequently used in translation study than Addition and Omission. And Substitutio is left out because no typical cases complying with the description provided by Delabastita could be identified.

The following shows the specific definition of the model of Feng(2020)

1. Addition means adding either verbal or non-verbal means in the Target Text (hereinafter TT). We proposed two kinds of Addition. The first one is to add punctuation marks in the TT to transmit the meaning contained in the speech para-verbal modes.

2. Omission means to omit the verbal modes of the ST and resort to non-verbal means which are usually the accompanying settings for transmitting the omitted meanings. The omitted parts are mostly expressions with cultural connotations and addressing terms.

3. Omission + Addition, which involves three different kinds of shifts. The first kind is to omit verbal modes such as modal particles in the translation and add non-verbal modes like punctuation marks to transmit the original meaning of the ST. The second kind is similar to the first, as modal particles are omitted in the

translation, but the difference lies in the added part: content words such as " please" are added in the translation.....In the third kind of shift, the omitted part is the verbal modes; non-verbal modes like punctuation marks are added. A very typical example is the omission of repetition and the addition of non-verbal modes such as punctuation marks.

4. Compensation, the meaning is twofold in our study; both usually happen when the ST contains culture-specific expressions. On the one hand, the translation in the TT is usually an explanation of the ST due to the untranslatability of the source text so as to compensate for the translational loss. On the other hand, the other medial variants in the accompanying setting can help to compensate for the loss of the cultural connotation when they could not be explained explicitly in the TT.

   The difference between Compensation and Omission is that in Compensation, the literal meaning of the expression in ST is verbally transmitted and the cultural conno tations are supplemented by the accompanying scenes; in Omission, the whole expressions in ST are completely omitted and the IMAGE can function to reveal or supplement the meanings of the original text.

5. Typographic Transformation, which refers to the change of typographic forms of the language when translating from ST into TT. The forms here mainly refer to the font, size, and color of the words, which may be shifted in the translation process resulting from the need to transmit the meaning of non verbal modes.

As for the feasibility of this model, Feng has studied the Chinese costume drama-- *Zhenhuan Zhuan*--collecting 408 translation examples, and offers a detailed explanation of some samples. Thus, it is fair to say the model of the Feng is operable and feasible.

4. The Possibility to Integrate Intersemiotic Shift with Multimodal Discourse Analysis Framework

Feng(2009) has quoted Chuang's words "'different semiotic modes contribute different kinds of meanings to the film text'(2006, 381)" and praises Chuang's work as a notable study, which has affirmed that Feng also believes meaning is not a result of a single semiotic resource, but the universal collaboration of different semiotic modes. Zhang(2009) also admits that the pure-language perspective of discourse analysis is no longer a "panacea" to give a whole and in-depth study. The meaning is also achieved through the non-language part, including para-language sectors, body sectors, environment and so on. Both of the scholars acknowledge the importance of non-language semiotic factors in the meaning making process, which is also the very basis of multimodality analysis.

When discussing the definition of multimodality study, Feng had adopted the definition of Kress that "the use of several semiotic modes in the design of a semiotic product or event" (Kress and van Leeuwen 2001, 20), while Zhang use "the phenomenon of using various senses such as hearing, vision, and touch to communicate through language, images, sounds, movements, and other means and symbolic resources." Kress' definition is from a more macro-level, while Zhang has made the semiotic resources more detailed and specific examples.

Still, Feng has adopted the definition of mode proposed by Kress as "a socially shaped and culturally given resource for making meaning"(Kress 2009, 54). In Zhang's work he doesn't give a accurate definition for mode, but we can grasp the term from his words like "However, when you are discussing with your friends about how it is snowing, the size of the snow, the way it is falling, etc., the process, manner and current state of the snowfall will all directly participate in the overall meaning of your expression, and your words will show a strong situational dependence. Therefore, the communication of words with strong situational dependence has the characteristics of multimodality... the above example has the multimodal feature of jointly realizing the communicative meaning through auditory and visual modes."(Zhang, 2009) The figure 2. in the previous part also demonstrates that Zhang regard modes has four types: visual, auditory, linguistic, and tactile modes. To find a better understanding of the common ground between the Zhang and Feng through mode, we can probe into the specific examples made by Feng. For example,

"She shouts, "You would not have known I am such a heartless soul!" Her emotion is extremely sorrowful and anguished, which can also be seen from her facial expression, and her cry can be well received by the Chinese audience. However, as she speaks Chinese, whose pronunciation and intonation are totally different from that of English, it is hard for the acoustic effect to function the same way for the English-speaking audience. In this case, the addition of an exclamation mark can be regarded as the most suitable way to compensate for the acoustic loss."(Feng 2020)

The success of a translation often relies on the visual and auditory modes, such as a character's facial expression (e.g., Zhenhuan's) and their intonation, to achieve successful communication. Both of these are crucial semiotic resources for understanding meaning and generating potential interpretations.

Zhang has proposed several types of modal interaction (see Figure 2). Feng's model is also based on the relationships between modes, but Zhang's typology is more detailed. Furthermore, Zhang's framework includes potential expressive methods at the expression level that should be considered during the meaning-making process, such as sonority, tone, fonts, and layout—elements also analyzed in Feng's work. The following examples demonstrate this.

"Marquess Guo's internal monologue is read out as a kind of voice-over, so the ST audience can immediately understand what is in his mind. However, if this internal monologue is directly translated into English as ordinary dialogues, the TT audience might misunderstand that he is talking with Zhenhuan. Moreover, it will be confusing if the TT audience think he is talking, as in the scene his mouth is closed. Therefore, changing the style of the font can be a possible way to solve this problem, reminding the TT audience this part is different from ordinary dialogues."(Feng, 2020)

With the explanation in the third part of this paper, we can understand that Feng's model is more tend to be the specific translation methodology. It teaches the translator how to translate in the situation of multimodal discourse. While with the idea of the second part of this paper, it may easily to say Zhang's framework is prone to a general framework that guild the translator to think how to translate in the multimodal discourse. Thus, the integration of their studies is tend to be practicable.

5. Analysis

In this part, an analytical study integrating multimodal analysis framework and intersemiotic shift will be presented. This case study will serve as the tool to link the actual discourse and the methodology.

5.1 Example Studies Based on the Multimodal Discourse Analysis Framework and Intersemiotic Shift

In order to describe the multimodal discourse in a well-rounded way, a description of the potential semiotic modes is needed. To achieve that end, the semiotic modes shall be described at the same time. Baldry and Thibault (2006) offer a detailed transcription methodology. However, their work is to detailed to study, but it can be summarized as in a relative easy way that there are four different vertical columns from left to right. The first vertical column is used to give a number for each lateral columns. The second vertical column is used to insert the image that need to describe with an aim to give the researchers and the readers an overall view of the described work. In the third vertical column, the image will be described in detail, telling the readers what is happening in the scene, including the facial expression, gestures, and so on. The last can be used to describe the sound-relevant features, including speech properties and background music.

In this part, the paper will integrate Zhang's multimodal discourse analysis framework with Feng's intersemiotic shift to carry out some practical research.

Example 1

Source Text (ST)1: 两军对垒

Target Text (TT)1:two armies are opposing each other

ST2：剑拔弩张

TT2：on the verge of breaking out a fight

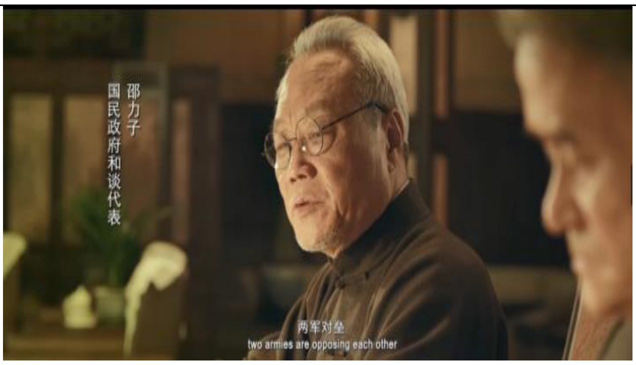
Number	Image	Visual description	Auditory description
1	 邵力子 国民政府和谈代表 两军对垒 two armies are opposing each other	Shao Lizi solemnly expression his statement	**Speech:** The tone of Shao is solemn, negotiating **Background music:** Low orchestra
2	邵力子 国民政府和谈代表 两军对垒 two armies are opposing each other	Same as above	Same as above

Table 1. Example 1

The culture level has been mentioned in the previous text, which involves the ideology composed of people's thinking patterns, life philosophies, living habits, and all the unspoken rules of society (Zhang Delu, 2009). Specifically, in the context of China, subtitle translation should effectively present various cultural elements with Chinese characteristics, including ideas, behaviors, and language expressions. In this film, the communication form and content are mostly presented through character dialogues. Based on this, the cultural level is analyzed in this paper.

The Chinese idiom "两军对垒，剑拔弩张" (liǎng jūn duì lěi, jiàn bá nǔ zhāng) vividly conveys a tense situation on the brink of war.

While this phrase is highly evocative for a Chinese audience, its literal translation, "with swords drawn and bows bent," may not fully capture the same sense of urgency and gravity for a foreign audience.

Instead of a literal translation, the subtitle adopts the Compensation method to make the meaning more explicit and concrete. The translator conveys the original intent by emphasizing that a conflict is imminent and the two armies are in an extremely tense state. This choice successfully bridges the cultural gap and ensures the target audience fully grasps the serious, high-stakes nature of the scene.

From the perspective of Content, the most significant visual elements are the close-up shot of the character in the middle and the subtitles at the bottom. The central character, Shao Lizhi, directs his gaze toward the Chinese peace negotiation representative. This, combined with his facial expression, creates a serious and tense atmosphere that sets the tone for the subsequent dialogue.

Sound is an indispensable mode in films, and music is an important component of the sound modality. In narrative films, music can be divided into subjective music and objective music. The former refers to the music composed by the composer to shape the character's personality, express the character's inner emotions, or create an atmosphere, which is added to the film during post-production； the latter refers to the music that the characters can perceive, reflecting the real objective world. (Wu Jianguo & Li Yujing, 2024)

From the perspective of sound， the sound sources here include character dialogue and background music. The character dialogue is characterized by its solemnity. The background music is a low and subjective music， and the addition of subjective music greatly enhances the emotional color, together with the character dialogue, it determines the expression style and forms a complementary - reinforcement - expansion relationship, which also provides a practical basis for the use of the Compensation method. In the Longman Dictionary of Contemporary English， "oppose" means "to fight or compete against another person or group in a battle" (Wang Lidi & Li Ruilin, 2011c: 1317). In the Oxford Dictionary, "on the verge of" means "very near to the moment when sb. does sth. or sth. happens" (Hornby et al.， 2010: 2237)， and "break out" means "of war, fighting or other unpleasant events to start suddenly" (Hornby et al., 2010: 233). The choice of words such as "opposing"， "on the verge of", and "break out" all reflect a strong sense of power and crisis， highlighting the combined effect of auditory and visual perception.

From the perspective of context, the field of this example is the exchange of ceasefire opinions between representatives of the Kuomintang and the Communist Party. The mode is formal spoken language, and the tenor is the speech of the Kuomintang representative in the peace talks. The relationship between the two sides is equal. Integrating the relevant information in the comprehensive context, the subtitle translation should pay attention to the realization of seriousness, formality, and equality. In subtitle translation, considering the effect of seriousness and urgency conveyed by the above aspects and the formal characteristics of the scene, the phrase "两军对垒" is translated directly to "two armies facing each other", accurately restoring the historical background of the tense standoff across the river between the Kuomintang and the Communist Party at that time, allowing the target language audience to intuitively understand the hostile situation in this context. As for "剑拔弩张", it is translated as "on the verge of breaking out a fight" by Compensation, which not only conveys the meaning of an impending conflict in the source language, but also conforms to the language habits and acceptance methods of English expression. At the same time, it echoes the tense and low-pitched orchestral music in the auditory aspect, highlighting the urgent atmosphere of an impending battle.

But this translation also seems to have room to be improved. The TT does not possesses punctuation. If the second TT were added with an exclamation, the serious atmosphere and the tensive condition between the two armies may be better achieved.

Example 2

ST:如果还是不听 肆意横行

TT:If they keep moving like mad animals,

ST:可以开炮

TT cannoneers get ready to fire!

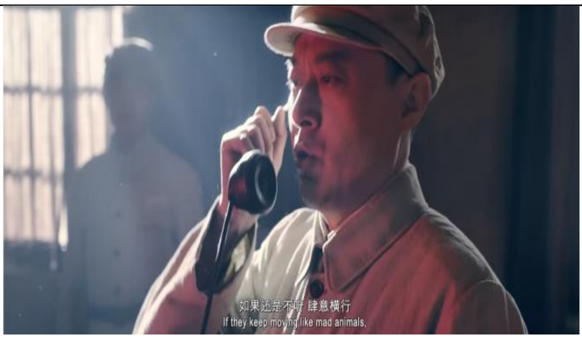
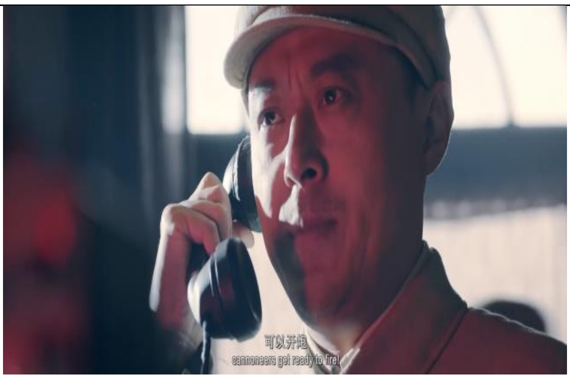
Number	Image	Visual description	Auditory description
1	 如果还是不听 肆意横行 If they keep moving like mad animals,	Chinese commander is sending the order.	**Speech:** The tone of is solemn and powerful; the sonority is high. **Background music:** Tensive orchestra
2		Sending the order with a determined mood.	Same as above

Table 2. Example2

At the culture level, the phrase "肆意横行" in the "Modern Chinese Dictionary" means "acting recklessly and according to one's own will" (Institute of Linguistics, Chinese Academy of Social Sciences, 2016: 1242), and "横行" means "acting tyrannically and doing evil deeds by taking advantage of one's power" (Institute of Linguistics, Chinese Academy of Social Sciences, 2016: 537). Therefore, "肆意横行" refers to doing evil deeds by relying on one's own power and acting according to one's own will. However, if the original meaning is directly translated into English, it would be too long for the subtitle. Therefore, the original meaning should be combined with the specific context and content levels for expression.

From the perspective of Context, this example involves the British warship HMS Amethyst's navigation in the Yangtze River despite a warning and the subsequent response of the Chinese military. The mode is an oral command, with the tenor being a battle order issued by a Chinese general to repel the British fleet. This establishes an underlying power dynamic where China is portrayed as superior to the UK. Therefore, the translation must be attentive to the implied meaning conveyed by these contextual factors.

From the content level, in terms of music, it is composed of a loud and serious character voice and tense orchestra music. This music generates a sense of crisis, while the commander's loud voice conveys his fearlessness and firm resolve in the face of danger. Together, these auditory elements significantly enhance the overall narrative tension and highlight the impending conflict.

In this example, the visual aspect is a close-up shot of the Chinese commander. The second lateral column adopts a perspective change from eye level to a low angle, zooming in on the character's image. According to Baldry and Thibault (2006) Vertical perspective is concerned with the power, status, and solidarity relations between the viewer and the depicted world. view the depicted world from below such that the viewer is placed in a position of inferiority. Thus, the viewer may unconsciously take the commander as a high status position and assuming the Chinese army possesses a stronger ability. This technique subtly elevates the commander's status in the viewer's perception, suggesting that the Chinese army possesses superior strength. The portrayal of the UK army's invasion is thus framed as a mere display of assumed power. This handling of the shots not only foreshadows the upcoming shift in the offensive and defensive dynamic but also contributes to the overall meaning-making process across the visual, linguistic, and auditory modes. These three modes work in concert to form a complementary- non-reinforcement- coordination multimodal relationship.

The information at the expression level is mainly conveyed through the commander's serious face and resolute eyes. The non-verbal body language formed by the two also enhances the emotional expression.

Therefore, by integrating the information from the four levels,the translation "If they keep moving like mad animals, cannoneers get ready to fire!" uses the Compensation method to translate "肆意横行" as "mad animals" to represent the British army, reflecting the power relationship and the commander's unyielding spirit in the face of danger. The figurative metaphor not only intensifies the wild and out-of-control behavior of the British warship but also implicitly highlights the legitimacy and reasonableness of the Chinese military's self-defense actions. This concrete expression， in turn, further enhances the synergy between the auditory and visual aspects. In the comprehensive context constructed in this paragraph, the translation is fully supported by the context and is consistent with the scene.

There are no punctuation marks in the original text, but in the translation, commas are added, which achieves a function to separate the two pieces of information, allowing the exclamation mark in the second part to gain an emphasis function and effectively make up for the deficiency in emotional expression in the ST. This treatment not only enhances the expressiveness of the translation but also echoes the tense atmosphere jointly created by the auditory and visual aspects as a whole.

Therefore, to effectively promote China's image, film subtitle translation must integrate information from all four levels of analysis—Culture, Context, Content, and Expression—while also considering the dynamic coordination among multiple modes.

Translators should select flexible translation methods, such as the Compensation method used in this example. The choice to translate "肆意横行" as "mad animals" accurately reflects China's determination to resist a powerful force and fight bravely. Furthermore, the flexible use of punctuation marks, such as the addition of an exclamation mark through the Addition shift, effectively emphasizes the auditory information, adding to the dramatic tension of the scene.

Example 3

ST: 大意失荆州，关公走麦城

TT: Self ego leads to the complete failure.

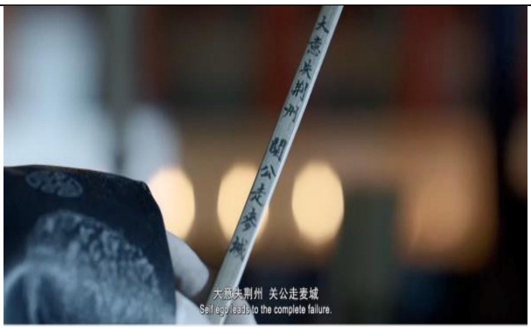
Number	Image	Visual description	Auditory description
1	 大意失荆州 关公走麦城 Self ego leads to the complete failure.	A bamboo bar with “大意失荆州，关公走麦城” on it	**Background music:** Low-pitch orchestra

Table 3. Example3

The source subtitle constitutes a Chinese proverb imbued with historical allusions, specifically referencing the narrative of "suffering a major defeat due to negligence and facing one's downfall." This historical allusion derives from the Three Kingdoms period(220-280 A.D), wherein General Guan Yu, tasked with defending Jingzhou, opted to campaign against Cao Cao. Despite intelligence reports warning of Sun Quan's impending assault on Jingzhou, Guan Yu dismissed them, ultimately resulting in the loss of the strategic commandery and his own capture at Maicheng. The original phrase carries profound and implicit semantic connotations, which are hard to understand for foreigners without a deep understanding of Chinese culture and history

In its translation, however, a literal rendering was deliberately avoided. Instead, a Compensation strategy was employed to concretize and explicate the historical allusion of "走麦城"(meaning one’s failure). This approach serves to articulate the metaphorical resonance of fate embedded within the idiom, thereby enhancing cross-cultural comprehensibility for international viewership.

On the semiotic level, the visual mode features a close-up shot focusing intently on the inscribed bamboo slip, ensuring the clear transmission of this pivotal visual information to the audience. The auditory mode is characterized solely by somber orchestral music, devoid of any dialogue, thereby fostering a solemn and tragic atmosphere that foreshadows the protagonist's inevitable outcome. These multimodal elements operate in a relationship of complement-reinforcement, and emphasis, collectively foregrounding the translated subtitle's message of "the complete failure."

Contextually, the field of discourse involves Chiang Kai-shek’s seeking instruction from heaven to help him knowing his fate and future prospects. In traditional Chinese culture, this form of divination is a significant practice used to resolve existential uncertainties or guide future actions. Chinese people use such bamboo bar to ask the heaven for the methods to resolve existential uncertainties or to orient future actions. Consequently, the translated content must adopt an interpretive rather than literal approach, mitigating potential cognitive load imposed by unfamiliar cultural aspects.

However, it is still acceptable that an additional subtitle can be added on the side of the screen to illustrate the fortune-telling activity with different fonts or layout.

Example 4

ST:「贵船应立即停下」

TT:*Lima- "You should stop your vessel instantly".*

ST:「请终止一切企图」

TT*Xray - "Stop carrying out your intentions".*

ST「可能会遇到危险」

TT: *Uniform - "You are running into danger".*

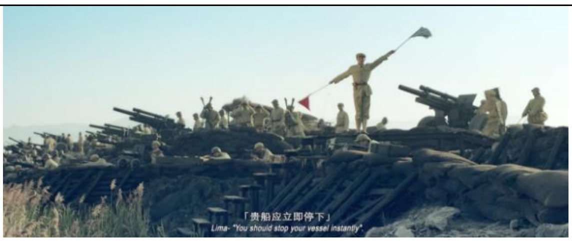
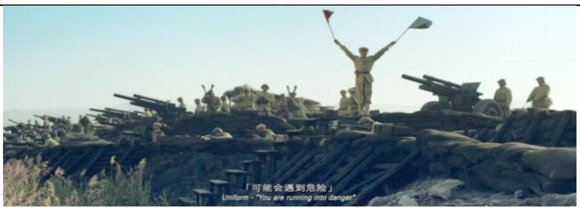
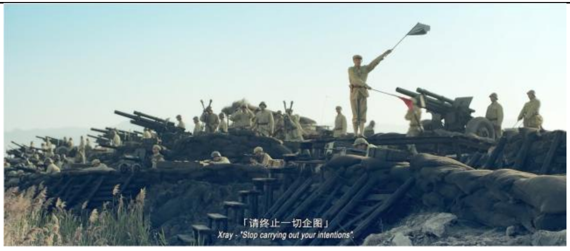
Number	Image	Visual description	Auditory description
1	 「贵船应立即停下」 Lima- "You should stop your vessel instantly".	A soldier standing on the battle field to sign with a flag	**Background music:** Tensive orchestra
2	 「可能会遇到危险」 Uniform - "You are running into danger".	Same as above	Same as above
3	 「请终止一切企图」 Xray - "Stop carrying out your intentions".	Same as above	Same as above

Table 4. Example 4

In this Example, the major difference is the font of ST and TT. According to Zhang, non-linguistic media involves the distinction between body and non-body. The non-body media encompasses the tools and instruments. The flag is used as one of the major media to convey meaning since the lack of speech. Here in the ST, the meaning of flag language is put into the "「」" to show the difference between speech or dialogue. And the TT uses the italic font to show the difference. Both the ST and TT can be explained by the Typographic Transformation. The perception of flag language is not achieved by the auditory mode but by the visual. Thus, the subtitle translation cannot just make the acoustic signs to the written form. It is necessary to distinguish this phenomenon, which leads to the change of font.

Example 5

ST: 主席啊

TT: Mr. Chairman,

ST: 清华园到了

TT: We are reaching the Tsinghua Station.

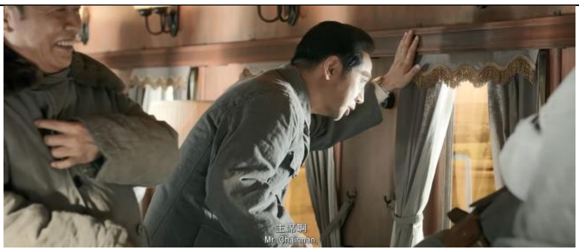
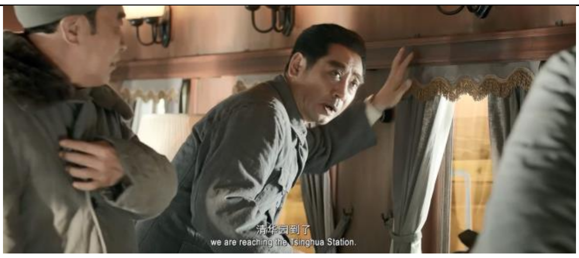
Number	Image	Visual description	Auditory description
1		Zhou Enlai is seeing outside from the train	**Speech:** Narrative tone **Background music:** Mild and lyric orchestra
2		Zhou Enlai is telling.	Same as above

Table 5. Example 5

This Example doesn't involve an important cultural sector. It is just a casual conversation. From the Content level, the modes involved are auditory and visual. The visual mode tells that Zhou has observed something from the train, which may be the station board. The auditory aspect is creating a comfortable and narrative atmosphere. From the context level, this conversation's filed is to state the fact, reminding the train is approaching the destination. The tenor shows the power relationship. One is subordinate and the other is superordinate. And the mode is verbal communication. This raises the basis to use Omission+Addition.

In the first subtitle, the Chinese particle "啊" (a), which has no specific meaning, is omitted. In its place, the English translation adds "Mr."and a comma, which not only reflects the subordinate's respect but also maintains the conversational tone. In the second subtitle, the phrase "清华园到了"(Qīnghuáyuán dào le), which means "some place called Tsinghua yuan is here"is a concise statement of fact. The translator, however, combines this with the visual information (the movement Zhou observing the station board) and the auditory narrative tone to produce a more fluid translation. The particle "了" (le) is omitted, and the subject "we"and the verb "are reaching" are added, resulting in a more complete and natural-sounding English sentence.

5.2 Summary of the Examples

From the Example studies, a relatively profound conclusion can be drawn that in the translation processes, the ST with cultural information will unavoidably pose challenges to the translation. One of the many problems is that subtitles only possess a little room to express information. Thus, it will technically limit the words used. Secondly, the time also challenges the translators. The ST and auditory mode will only last for a few seconds. If the translator cannot make his or her translation to fit the time scope. There will raise another problem: sound-image synchronicity. Thirdly, the cultural factors sometimes may be easy to translate, and sometimes we can even use Omission to omit them like particles. However, sometimes the concise ST may load with lots of information. For instance, "大意失荆州， 关公走麦城".

The above Examples, which carry lots of cultural information, are mostly solved with Compensation. It translates the potential meaning with the help of other modes. It is effective to convey the main idea of the communication. But in some places it can be improved. For example, when dealing with "大意失荆州， 关公走麦城", it has been written on the bamboo slip. This represents a traditional Chinese fortune-telling activity. One uses a cup or some container to stole a lot of bamboo slips with some proverbs on it. The one shakes it until a bamboo slip drop out or one chooses one slip out, which will be the guidance to solve the upcoming problems or telling one's fate. This explanation could be added on the side of the image. If this kind of information is also added to other movies with this kind of activity, the Chinese cultuel could be well disseminated.

In the film, the use of punctuation can be barely found in ST, However, in the TT, the punctuation can be found almost everywhere. The reason may lie in that the previous Chinese doesn't possess punctuation marks, but in English, this is necessary. Another reason could be that the ST audience can understand the emotion expressed by the semiotic sign itself. The word has its own meaning and emotional potential, and the ST audience may easily understand them. However, the TT audience doesn't possess the shared social semiotic knowledge. Thus, the Addition tends to help the TT audience to construct the meaning.

6. Conclusion

In the translation activity, the multimodal translation shall be carefully noticed, especially in the film or audiovisual discourse. During the translation processes, the translator could use the multimodal discourse analysis framework as the guidance to consider the whole process and to gain the reasons to translate in a certain way. In the actual translation, Addition, Omission, Addition+Omission, compensation, and Typographic Transformation could serve as the method to deal with it, because the two models get their common ground, but multimodal translation may not always successfully be achieved. There must exist the common ground to make modes transferable.

The subtitle translation is not only to translate the speech the speaker produced. The involved images or hidden cultural information may also shall be translated, which can be put on the side of the image or activity.

References

Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. David Brown Book Company.

Boria, M., Carreres, Á., Noriega-Sánchez, M., & Tomalin, M. (2020). *Translation and multimodality: Beyond words*. Routledge.

Chuang, Y.-T. (2006). Studying subtitle translation from a multi-modal approach. *Babel, 52*(4), 372–383.

Delabastita, D. (1990). Translation and the mass media. In S. Bassnett & A. Lefevere (Eds.), *Translation, history and culture* (pp. 97–109). Pinter Publishers.

Díaz Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St. Jerome.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist, 12*(3), 196–205.

Ketola, A. (2016). Towards a multimodally oriented theory of translation: A cognitive framework for the translation of illustrated technical texts. *Translation Studies, 9*(1), 67–81. https://doi.org/10.1080/14781700.2015.1124899

Kress, G. (2009). What is a mode? In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 54–67). Routledge.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Edward Arnold.

Pérez-González, L. (2014). Multimodality in translation and interpreting studies. In S. Bermann & C. Porter (Eds.), *A companion to translation studies* (pp. 119–131). Wiley-Blackwell.

Zhang, M., & Feng, D. (2020). *Multimodal approaches to Chinese-English translation and interpreting*. Routledge.

廖巧云. (1999). 功能语法理论在文体分析中的应用——语篇分析范例 [The application of systemic functional grammar in stylistic analysis: A discourse analysis case study]. *外语与外语教学* [Foreign Languages and Their Teaching], (08), 14–17.

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 [Multimodal contextual reconstruction and the dissemination of Chinese film and television culture: A case study of subtitle translation in *My People, My Country*]. *山东外语教学* [Shandong Foreign Language Teaching], 45(02), 11–21.

王立弟, & 李瑞林. (2011). *朗文当代英语大辞典* [Longman contemporary English dictionary]. 商务印书馆.

中国社会科学院语言研究所词典编辑室. (2016). *现代汉语词典* (第 7 版) [Modern Chinese dictionary] (7th ed.). 商务印书馆.

冯德正. (2017). 多模态语篇分析的基本问题探讨 [Discussion on fundamental issues in multimodal discourse analysis]. *北京第二外国语学院学报* [Journal of Beijing International Studies University], (03), 1–11, 132.