

The Image of Sichuan from Multimodal Perspective: A Case Study of Bilingual Promotional Videos

Tingting Gou*, Jing Chen

School of Foreign Languages, Southwest Petroleum University, Chengdu, Sichuan, China

*Corresponding Author.

Abstract: In the process of accelerating the implementation of China's cultural "going global" strategy, bilingual image videos have become an effective way of promotion, showcasing distinctive regional features with the help of texts, images and sounds. From the multimodal perspective, this paper taking the bilingual promotional videos of Sichuan in 2022 and 2023, explores the significance of the verbal and visual modalities, analyzes how the two interact with each other, and interprets the image constructed by the promotional videos. It is found that multimodal semiotic resources are utilized to jointly construct the image of Sichuan as one with a long history, good ecology, technological progress and diversity. The multimodal perspective can provide viewers with enhanced means to appreciate promotional videos, and offer creators valuable references for the translation and production of bilingual branding contents.

Keywords: Multimodal Discourse Analysis; Bilingual Promotional Videos; Verbal Modality; Visual Modality; Image Construction

1. Introduction

As the typical dynamic multimodal discourse, bilingual promotional videos integrate multiple semiotic resources and convey the core meaning by efficiently integrating various modalities. With the purpose of cultural dissemination and enhancing influence, such branding videos provide a comprehensive display of local economy, culture, history, ecology and social life, etc. From 2022 to 2023, the Sichuan International Communication Center (SICC) produced and released two bilingual promotional videos *Symphony of Seasons on the Land of Abundance* and *Spectacular Sichuan on the Silk Road*. These

videos serve as important windows for the whole world to understand Sichuan, and have emerged as key promotional tools. The former was released in 2022, with the duration of 6 minutes. The video uses the four seasons as the narrative thread to connect the typical symbols of Sichuan, showing the magnificent mountains and rivers, rich cultural heritage and flourishing social scene of the province. The latter released in 2023 is 6 minutes and 13 seconds in length. Starting with silk, the video showcases Sichuan's culture and modernization achievements. The two bilingual promotional videos vividly depicted Sichuan's economy, ecology, culture and history, telling the unique story of Sichuan.

This paper aims to discuss and analyze the meaning construction process of and the interrelationship between the verbal and visual modalities from the perspective of multimodal discourse analysis, taking two bilingual promotional videos of Sichuan as the research subjects, with a view to providing some reference for the production and translation of bilingual promotional videos in cross-cultural communication. This study attempts to answer the following questions: 1) what meanings are expressed by the verbal modality in the two bilingual image videos; 2) what meanings are conveyed by the visual modality in them; and 3) how do the two modalities interact with each other and what kind of image of Sichuan is constructed in the two image videos.

2. A Brief Overview of Multimodal Discourse Analysis

Multimodal Discourse Analysis (MDA) emerged in the 1990s, with its theoretical basis of Systemic Functional Linguistics (SFL) founded by M. A. K. Halliday. MDA accepts the view from SFL that language is social semiotic and has meaning potential, and considers other semiotic systems as source of meaning. Based on Halliday's SFL, Kress and

Leeuwen analyzed the meaning construction of visual symbols and proposed Visual Grammar (VG) to reveal how visual elements construct meaning.

Internationally, MDA has been applied across various fields such as education, media studies, and digital communication, providing insights into how different modalities work together to shape understanding and influence perception. The multimodal discourse researches in China are mainly reflected in the two aspects: theoretical exploration and applied research. The introduction of MDA framework into China can be traced back to Li Zhanzi, who was among the first to apply this framework in his research in 2003 [1]. Subsequently, scholars such as Hu Zhuanglin, Zhu Yongsheng, and Zhang Delu also conducted a series of studies in this field. Zhang Delu explored a comprehensive theoretical framework for MDA, and Zhu Yongsheng discussed the theoretical foundation and research methodology of MDA, pointing out the direction for the development of discourse analysis in China [2]. MDA integrates language and other semiotic resources for research, and has been applied in many fields, such as communication, education, art and so on. In the field of communication, the objects of MDA used to be posters, advertisements and movies. For example, Wang Hongyang discusses the construction of interactive meaning in multimodal advertisement discourse by adopting VG [3], Peng Yi and Yang Haiyan explore the overall meaning construction of multimodal print commercial advertisement discourse [4], and Lu Jian and Wu Wenzhi studied the subtitle translation of a movie from the multimodal perspective [5]. Existing studies on promotional videos from the multimodal perspective have primarily focused on English translation studies, image analysis, and national image, urban image or character image construction [6-8]. However, the research on multimodal image construction of promotional videos has yet to fully explore the image construction of provincial image. The image of Sichuan is particularly understudied. Therefore, this paper intends to examine the image construction of Sichuan in two bilingual image videos from a multimodal perspective.

3. Theoretical Foundations of Multimodal

Discourse Analysis of Bilingual Promotional Videos

Bilingual promotional video is the typical multimodal discourse which involves multiple symbols such as text, image and sound. Systemic Functional Grammar (SFG) grounded in the theory of social semiotics provides a framework to decipher the discourse meaning of verbal modality; VG offers tools to analyze the discourse meaning of visual modality, and the multimodal synergistic relationship proposed by Zhang Delu helps to reveal the relationship between modalities.

3.1 Systemic Functional Grammar and Visual Grammar

According to Halliday, the founder of SFG, language is a system of meanings, which mainly has three metafunctions (meanings) i.e. ideational metafunction, interpersonal metafunction, and textual metafunction [9]. Ideational metafunction refers to the function of reflecting the occurrence and process of things and events in the objective world as well as people's inner world; interpersonal metafunction means that language has the function of establishing, maintaining and reflecting interpersonal relationships; textual metafunction is about the function of forming sentences and organizing and transmitting information according to the context. Ideational metafunction mainly involves transitivity, voice and ergativity; interpersonal metafunction includes mood, modality and key; and textual metafunction mainly covers thematic structure, information structure and cohesion [10].

Based on SFG, Kress and van Leeuwen analyzed the semantic relations of visual symbols and proposed VG [11], which has three levels of meaning: representational meaning, interactive meaning and compositional meaning. Representational meaning is mainly reflected in the representation of the relationship between people, events and places through images. It can be categorized as narrative representation and conceptual one, with the main difference being the existence or absence of vector. Interactive meaning refers to the interaction between the audience and the image, which is realized by contact, social distance, perspective and modality. Compositional

meaning regards the meaning constructed through image space and location, including information value, framing and salience.

3.2 Multimodal Synergistic Relationships

Zhang Delu proposed a comprehensive framework for MDA based on SFG, believing that different modality forms of discourse actually serve to reflect the overall meaning of the speaker. He distinguishes the relationship between different modalities: complementary relationship and non-complementary relationship [12]. Complementary relationship refers to the fact that one modality cannot fully express the meaning, and needs to be supplemented by other modalities. It can be further categorized into reinforcement and non-reinforcement. Reinforcement relationship indicates one modality is the dominant form of communication and other modalities highlight it. Non-reinforcement means that multiple modalities jointly contribute to meaning making and none of them can be omitted. It can be realized by coordination, coalition, and intersection. Additionally, non-complementary indicates that there is no direct connection between different modalities which are relatively independent. This relationship can be reflected in blending, inclusion or embedding, and contextual interaction. Blending refers to the simultaneous appearance of multiple modalities that do not reinforce each other, but simply convey consistent information through different semiotic systems. Inclusion or embedding happens when a modality does not contribute something new to the expression of the overall meaning, but gives more specific information to other modalities. It can be divided into part-whole relationship and specific-abstract relationship. Contextual interaction consists of context-independent relationship and context-dependent one. The former means that the multimodal discourse functions autonomously without direct connections to the context. In contrast, context-dependent interactions imply that the discourse relies on the context, which in turn provides support and background for the discourse.

3.3 Image Construction

Multimodal promotional videos are among the most effective ways to introduce and promote a country, a region, a city or an enterprise. The

image in this context does not refer to an objective representation, but rather to the perceptions and interpretations constructed by the audience. The image of a country is the recognition and evaluation of the country's political, economic, social, cultural and geographic conditions by the public at home and abroad [13]. Regional image should also include many aspects of a region, such as economic image, political image, social image and cultural image. Regional image is a symbolic and stable stereotype formed by the audience for a region [14]. Shaping a good regional image is conducive to enhancing regional competitiveness and promoting regional economic and social development. In recent years, many scholars have paid attention to the discourse construction of China's regional image presented by domestic and foreign mainstream media, such as the image of Shanghai in *The Times* and the construction of Guangxi's image in the *Vietnam News* [15-16]. Huang Dongqun utilized the theory of inter-semiotic complementarity to study the image construction of Fujian [17]. Currently, studies on regional image construction from the multimodal perspective focus on metaphorical cognition, discourse interaction paths, and discourse synergy analysis [18-20]. Previous studies have demonstrated the effectiveness of multimodal analysis in understanding the construction of national and urban images. However, research on provincial image construction remains relatively underexplored. Against this backdrop, this paper examines the bilingual image videos of Sichuan through a multimodal lens to explore the economic, cultural and social images.

4. The Image of Sichuan and the Meaning Construction

The two bilingual promotional videos *Symphony of Seasons on the Land of Abundance* and *Spectacular Sichuan on the Silk Road* make use of images, texts and sounds to transform abstract concepts such as economic development, history and culture, city symbols and natural sceneries into concrete scenes. In the two videos, the verbal and visual modalities are particularly prominent, working together to present a multifaceted image of Sichuan to the global audience. To explore the significance of these

two modalities and their synergistic relationship, selected cases from the videos are analyzed. Additionally, the constructed image of Sichuan is interpreted through the lens of these modalities. A total of seven examples are chosen to illustrate the diverse characteristics of Sichuan across different domains.

4.1 The Ideational Meaning in Verbal Modality and Representational Meaning in Visual Modality

Ideational metafunction, according to SFG, is to define the world. According to Kress & van Leeuwen, the representational meaning indicates that images can record and represent the external world. The ideational meaning of verbal modality thus corresponds to the representational meaning of visual modality. For the ideational meaning, the two promotional videos reflect the objective world and people's inner world mainly through the transitivity system and voice. The transitivity system utilizes different types of processes to express people's experiences, such as material, mental, and relational processes [21]. A process consists of three components: process, participant, and circumstance. As for representational meaning, the two videos employ a combination of conceptual and narrative processes, in which the conceptual process serves as the foundation, while the narrative process as the supplementary, to show Sichuan's cultural symbols, natural beauty and social life. Thus for better illustration, typical screenshots of the two image videos are taken and analyzed to explore the ideational meaning and the representational meaning. Hereinafter ST refers to the source text of Chinese subtitles or scripts in the video screenshot, and TT the target text and English translations. The original Chinese characters in ST is transformed into Chinese Pinyin in this paper. Additionally, promotional video 1 refers to *Symphony of Seasons on the Land of Abundance*, and promotional video 2 *Spectacular Sichuan on the Silk Road*. Every example includes the subtitle and the image.

Example 1:

ST: Zai zhe li, tian gao di kuo, xin chi shen yuan.

TT: On this land, our dream is big and the sky is the limit.

Table 1. Transitivity of ST and TT in Example 1

Transitivity Analysis	
ST	Zai zhe li (circumstance , tian (carrier) gao (attribute) di (carrier) kuo (attribute), xin (beholder) chi (behavioral process) shen (carrier) yuan (attribute).
TT	On this land(circumstance), our dream(carrier) is (relational process) big (attribute) and the sky(carrier) is(relational process) the limit (attribute).

Example 1 is taken from 2 minutes and 58 seconds of *Symphony of Seasons on the Land of Abundance*. It includes Chinese and English subtitles and a image that shows the volleyball match hosted by Pidu district, Chengdu.

At the verbal level, Table 1 shows the transitivity analysis of the ST and the TT in Example 1. In the ST, "tian gao di kuo" means the sky is high and the earth is vast, indicating a very wide space; "xin chi shen yuan" means the mind is focused on the yearning for things or places. According to the transitivity system in ideational meaning, the ST employs relational process and behavioral process. In the ST, the attributive relational process is used, which consists of a carrier and an attribute. These processes reflect the relationship between entities. Specifically, there are three instances of attributive relational processes that illustrate the nature of the relationship between things. The behavioral process, on the other hand, refers to physical actions and involves a participant known as the beholder. Given the concise and condensed nature of language in image films, the word "shi" (which symbolizes the relational process) is often omitted between the two Chinese characters in these relational processes. The main information of the ST is that the earth is vast and people are bold enough to pursue their dreams. In order to reflect this ideational meaning in translation, the translator employed a unified relational process based on the Chinese interpretation. "tian gao di kuo" is translated as "the sky is the limit", and "xin chi shen yuan" is translated as "our dreams are big". The two translations are highly condensed with a striking contrast provided by "limit" and "big" to express the core meaning of the ST. Formally, the translation is as concise as the ST. On the whole, the TT achieves functional equivalence with the ST in terms of ideational

meaning.

At the visual level, according to the representational meaning, there are two participants in this image: one is the volleyball, the other is the two groups of volleyball players. Representational meaning indicates the representation of things and principles in the real world, which includes narrative process and conceptual process. The image with a vector is the narrative image, and the image without is the conceptual one. The image shows the narrative representation and the vector is the action of the volleyball rolling down the ground. It also narrates the event of volleyball players fighting for the volleyball through a behavioral process in which the athletes are the actors and the volleyball is the goal. The energy and drive of the young generation are demonstrated in this image.

Overall, verbal modality and visual modality in Example 1 interact effectively with each other and form context-independent relationship. The image is not directly related to the subtitle. The verbal modality conveys the message of boldly pursuing a dream, while the visual modality presents a picture of many athletes scrambling for the volleyball in the volleyball match, which creates a discourse environment for the text. Additionally, the English translation adds the word “our” to get viewers close to the video and help them to better embrace the meaning. On the whole, the subtitle is impressive and the image shows the vitality of the young generation, constructing the image of Sichuan with vigor and high morale.

Example 2:

ST: San xing dui yi zhi zhong fa xian da liang si chou can liu wu.

TT: At Sanxingdui, a large amount of silk residues were uncovered.

Example 2 is taken from 4 seconds of *Spectacular Sichuan on the Silk Road*. It presents silk residues from the Sanxingdui site. At the verbal level, ST in Example 2 adopts the active voice to express ideational meaning. As shown in the Table 2, the ST, consisting of circumstance, process and goal, conveys the information objectively through a non-subject sentence that demonstrates the formal feature of bilingual promotional videos. Given that the relationship between goal and process is the core of the ST, active voice is employed to emphasize this relationship. In order to keep

the same effect as the ST and make the translation concise and objective in TT, the translator did not add actor but changed the linguistic form by converting the active voice to passive voice to realize the ideational meaning of language.

Table 2. Voice Analysis of ST and TT in Example 2

	Description	Voice Analysis
ST	San xing dui yi zhi zhong (circumstance) fa xian (process) da liang si chou can liu wu (goal).	Actor: / Processes: fa xian Goal: da liang si chou can liu wu Type of non-middle voice: active
TT	At Sanxingdui (circumstance), a large amount of silk residues (goal) were uncovered (process).	Actor: / Processes: uncovered Goal: silk residues Type of non-middle voice: passive

At the visual level, there is no vector in Example 2 which shows the symbolic process of conceptual representation. Conceptual representation mainly describes the properties and qualities of static objects. And symbolic process indicates the participants [22]. As shown in the Example 2, the image presents the silk residues of Sanxingdui site. As silk residue is a part of the cultural relics and symbolizes the history and culture, therefore, the silk residues of Sanxingdui in the image symbolize that Sichuan is a historical participant and witness of the ancient Silk Road, demonstrating the integration and connection between Sichuan and the world.

Overall, the subtitles are translated appropriately, with a blending relationship between verbal modality and visual modality in Example 2. The verbal modality presents the fact that there are a large amount of silk residues at Sanxingdui, and the visual modality shows the audience the corresponding image. The two modalities appear simultaneously, conveying the consistent information. When audiences watch the video, they can have deeper understanding of the silk residues in Sanxingdui with the help of subtitles. On the whole, the two modalities collaborate with each other, vividly conveying meaning to the audience and constructing the image of Sichuan that has spanned thousands of years in history.

4.2 The Interpersonal Meaning in Verbal

Modality and Interactive Meaning in Visual Modality

For interpersonal meaning, the two image videos mainly reflect interpersonal relationship and conduct social communication through mood and modality. For the interactive meaning, the promotional videos reflect the interaction and communication between the creator of the image, the image and the viewer through contact, social distance, perspective and modality. Contact refers to the eye contact between the viewer and the participants in the video. The image with eye contact is called “demand”; and the image without falls into “offer”. Social distance means the social relationship between the image participants and viewers, which is distinguished by the choice of long shot, medium shot and close shot. Perspective (also attitude) refers to the angle of the shot, which is categorized as horizontal level and vertical level. Horizontal level include frontal angle and oblique angle. In addition, modality are mainly determined by the color saturation of the image. It is further divided into three values: high sensory modality, medium sensory modality and low sensory modality to express credibility or authenticity. The interpersonal meaning of language and the interactive meaning of images will be analyzed in the following examples.

Example 3:

ST: Zai zhe li, cao mu shan he, liu tang zi ran de feng zi.

TT: On this land, the beauty and charm of nature overflow.

Example 3 is taken from 1 minute and 23 minutes of *Symphony of Seasons on the Land of Abundance*. The image presents a scene of a marmot appearing on a prairie.

SFG suggests that mood could help realize interpersonal meaning. Mood refers to the speaker’s attitude towards the situation or thing, which shows different communicative purposes. Four mood systems are generally distinguished in English: declarative mood, interrogative mood, imperative mood and exclamative mood. At the verbal level, the ST in Example 3 adopts declarative mood to express the declarative function of language and present the fact. When translating, the translator retains the declarative mood of the ST to achieve the same expressive effect and realize the interpersonal metafunction of

language. Given that “grass, trees, mountains and rivers” fall into the category of “nature”, they are omitted in TT to make the translation more concise. The verbal modality constructs the image of a place with beautiful natural scenery.

At the visual level, Example 3 is offering image as the marmot is not looking at viewer. This image adopts close shot, with the marmot showing its upper body. It represents close social distance between marmot and the viewer. The horizontal angle illustrates the equal relationship between the marmot and the viewer, showing that humans and animals coexist harmoniously and are in the same ecosystem. For modality, VG believes that colors with high sensory modality have the effect of guiding the viewer’s emotional experience. Example 3 presents a stark contrast between high sensory modality and low sensory modality through the green prairie and the brown marmot because of the different saturation. The bright green with high saturation is a vivid illustration of beautiful ecological environment in Sichuan and enables audience to gain valuable experience in terms of emotion.

In general, the subtitle translation appropriately expresses the meaning of the ST in Example 3. Verbal and visual modalities are related to each other with an inclusion relationship. To be more precise, it is the part-whole relationship. The text introduces the “whole”, i.e. nature, while the image showcases the “part”, i.e. prairie and marmot. There is an obvious interaction between text and image. Two modalities complement each other, showcasing Sichuan’s good ecological environment and constructing an image of the beautiful Sichuan with excellent natural scenery.

Example 4 and Example 5:

ST: “Hua long yi hao” yao zou chu guo men, jiu bi xu yong you ju you wan quan zi zhu zhi shi chan quan de zheng qi fa sheng qi.

TT: The “Going Global” of Hualong One demands for complete intellectual property right of the vapor generator.

Example 4 and Example 5 are at 3 minutes and 38 seconds and 3 minutes 42 seconds respectively of *Spectacular Sichuan on the Silk Road*, showing the scene where the researcher introduces Hualong One. In the two examples, the subtitle is not the voiceover, but narrated

by the researcher Zhang Fuyuan. Hualong One is a third-generation nuclear power technology with fully independent intellectual property rights, and is the world's first nuclear power unit in commercial operation.

At the verbal level, the modality system expresses the speaker's judgment about the success and validity of the proposition he is making, the obligation he is asking the other person to assume in a command, or the personal will he is expressing in a proposal. Modality can be expressed by finite modal verbs, or by modal adverbs expressing probability or frequency. Chinese also has corresponding volitive verbs to express modality. The ST uses the high-value volitive verb "must" to illustrate the necessary conditions for realizing "Going Global" of Hualong One. In Example 4 and Example 5, Zhang Fuyuan, the chief designer of Hualong One reactor's vapor generator, is the speaker. The volitive verb "bi xu" indicates his judgment of the possibility of the proposition he is talking about. The high-value volitive verb in ST is omitted by the translator in TT without the corresponding English modal verb "must". The verb "demand" without modality is employed to make the translation more objective and to reduce the personal judgment brought by the high volitive verb in the ST. Given that bilingual promotional videos are oriented to the world, the translator takes into account the level of understanding and acceptance of the English-speaking audience and harness modality to achieve cultural communication.

At the visual level, there is no eye contact between the participant and the viewer, so Example 4 and Example 5 are both offering category. The two images adopt close shot with the researchers showing the upper body, showing the close social distance. Additionally, the horizontal frontal angle gives the viewer a stronger sense of presence and involvement, which indicates the photographer's hope that the viewer can gain a deeper understanding of the development of China's nuclear power technology. As for modality, black, white, and grey are employed in the two examples. The low sensory modality and low saturation enhance the mystery of scientific research. The black clothes of the researcher contrast with the faint yellow machine in Example 4, reflecting the rigor and professionalism of the

researcher.

Overall, there is a reinforcement relationship between verbal modality and visual modality in Example 4 and Example 5, introducing the fact about Hualong One. Images reinforce the texts by presenting two pictures of vapor generator. With the help of images, the text can be better understood by the audience. The two modalities work together to reflect the seriousness and determination of the researchers, and to construct the image of Sichuan, which attaches great importance to scientific research and technological development.

4.3 The Textual Meaning in Verbal Modality and Compositional Meaning in Visual Modality

The two promotional videos mainly articulate the elements within the text through cohesion. Cohesion is categorized into grammatical cohesion and lexical cohesion. Grammatical cohesion specifically refers to reference, substitution, omission and conjunction. For the compositional meaning, the two videos convey the overall meaning of the image through information value, salience and framing. Information value is determined by the position of each element in the image (left or right, bottom or top, center or margin), and different positions result in different information values, i.e. different significance. Salience refers to the attractiveness of an element in an image to the viewer. It can be analyzed in terms of size, image focus, and the relationship between the foreground element and the background. Framing indicates the visual frame lines that disconnect and connect elements of the image[23].

Example 6

ST: Fu yao zhi shang, zhe li xia zhi bai shi xing.

TT: And we resolve to rise to the top.

Example 6 is at 3 minutes and 14 seconds of *Symphony of Seasons on the Land of Abundance*. This image shows a scene of the Zigong Lantern Festival. The Zigong Lantern Festival is the local traditional folklore in Zigong, Sichuan Province, which was listed as a national intangible cultural heritage in 2008.

At the verbal level, The ST is condensed by using the four-letter phrase. This text belongs to the Summer Chapter of *Symphony of Seasons on the Land of Abundance*. The word

“xia zhi” makes a pun. It refers to one of the twenty-four solar terms, and implies that Sichuan is at the boom period. On the whole, the TT is slightly inadequate. “Rise to the top” only conveys the meaning of “Fu yao zhi shang”, which implies Sichuan is unstoppable and thriving. The translation of the sentence fails to convey the connotation of “bai shi xing”, therefore fails to realize the textual meaning of the ST.

At the visual level, the compositional significance of Example 6 is realized by the information value, framing and salience. The sky is in the center of the image and the other buildings are on the margin, representing significant and secondary information respectively. The tower at the center of the picture is bright, distinguishing it from the light of the surrounding buildings. Additionally, the golden tower covers a large area in the whole picture, which is especially conspicuous and remarkable against the dark night. All this shows high salience. For framing, the visual frame lines of the image connect all elements through the distribution of color, with different modules having different colors. The layout of Example 6 allows viewers to appreciate Zigong Lantern Festival from the macro level, and also triggers the interest of viewers who want to experience it in person, which in turn promotes the development and popularity of tourism in Zigong, and even the whole of Sichuan province.

In this example, the Chinese subtitle is not is not literally represented in translation, instead the English subtitle is rather echoing the framing of the image, offering a bird’s eye view to the viewers. The verbal modality and visual modality have the contextual interaction relationship. To be more precise, it is the context-independent relationship. There is no direct connection between two modalities. The verbal modality describes “rise to the top”, while the visual modality shows the scene of the Lantern Festival held in Zigong, Sichuan province, which provides an indirect discourse environment for the visual modality. With the help of the image, the audience can understand the double meaning of “xia zhi”. The two modalities work together to achieve communication, presenting an image of Sichuan that is modernized, economically prosperous, and technologically advanced.

Example 7:

ST: Jing wei jiao cuo, zhe shi da mei si chuan de hua juan.

TT: The threads are interwoven into a splendid scroll of Sichuan.

Example 7 is from 3 seconds and 34 seconds of *Spectacular Sichuan on the Silk Road*, showing the beautiful natural scenery in Sichuan province.

At the verbal level, the ST uses reference in grammatical cohesion to realize textual meaning in Example 7. The word “zhe” stands for “Jing wei jiao cuo”, which makes the ST more concise. In the ST, the words “Jing wei” and “hua juan” are important information. The former refers to the lines of latitude and longitude in geography, aiming to illustrate the vastness of Sichuan and its infinite scenery. Here, the translator employed the word “threads” to express the implied meaning of “Jing wei”. The TT “threads are interwoven” implied that there is a large span of latitude and longitude in Sichuan province. In order to realize the textual metafunction of the ST, “zhe” is omitted in the TT appropriately according to the context. The above translation choice not only makes the TT accurately convey the meaning of the ST, but also construct the image of Sichuan, which has a vast territory and picturesque scenery.

At the visual level, mountains at the top of Example 7 are inverted and fall into the ideal information; the mountains, clouds, green hills and lakes at the bottom are real information. The Chinese and English subtitles in the center of the figure are especially prominent against the background of the sky and mountains, attracting viewers’ attention. For framing, there is an obvious framing strategy between the visual elements of green mountains, rivers, blue sky and clouds in this image. First, there are obvious framing lines between visual elements, such as the framing line between the green mountains and the water and that between the green mountains and the blue sky. Additionally, the visual elements are regarded as separate entities to express meaning. In Example 7, the green mountain indicates high vegetation coverage; the green water represents clean environment; the blue sky refers to good air quality. The combination of these individual elements constitutes a picture of ecological harmony and beautiful environment, reflecting the image film maker’s ecological concepts of green,

low-carbon and sustainable development.

On the whole, the verbal modality and visual modality effectively interact with each other, and the relationship between them is inclusion. To be more precise, it is the specific-abstract relationship. The verbal modality talks about the abstract object “scroll”, and the visual modality presents the concrete scene: the beautiful mountains and rivers of Sichuan, like a scroll. The two modalities interact with each other and work together to convey meaning to viewers, and build up the image of Sichuan with a vast area and beautiful geographical environment.

The two promotional videos *Symphony of Seasons on the Land of Abundance* and *Spectacular Sichuan on the Silk Road* show the long history, beautiful scenery and diversified society of Sichuan province in all-round respects through verbal modality and visual modality. Some of the English translations in the videos are not rendered literally and some connotations of the Chinese are lost in translation. Additionally, some of the images do not fit well with the texts. However, generally, the two bilingual videos effectively present Sichuan’s multifaceted identity. The first video begins with the theme of the four seasons, while the second starts with silk. The selected figures cover a range of fields, including economy, culture, society, ecology, and sports. The ST and TT, though with some minor losses in meaning, are generally clear and concise, effectively serving for the purpose of foreign publicity. Verbal modality and visual modality cooperate with each other and constitute a variety of relationships, such as blending, inclusion or embedding, reinforcement, and contextual interaction, to jointly construct the image of Sichuan, which is vigorous, technologically advanced, ecologically harmonious, full of history and modernity, as well as diversified.

5. Conclusion

Bilingual promotional videos can be used as important business cards for countries, regions and cities to enhance their popularity and construct unique images. This paper analyzes two bilingual promotional videos of Sichuan in 2022 and 2023 from a multimodal perspective, exploring in depth the meaning construction and the interaction between verbal modality and visual modality. The synergistic effect of

the two modalities makes the two bilingual image videos successfully present the real image of Sichuan to the world.

At the verbal level, the Chinese subtitles in the image videos are concise and condensed, and most of the English subtitles successfully achieve metafunction equivalence with the ST. Overall, the two bilingual videos realize ideational, interpersonal and textual meanings through verbal modality to construct the image of Sichuan where history and modernization are intertwined. For the part of the translation that could be optimized, some Chinese source texts are difficult to translate effectively. It is therefore advisable to consider conveying their meanings through visual imagery. In such cases, text and images can work together in a non-complementary relationship to jointly construct the intended meaning of the Chinese source text. At the visual level, the promotional videos show the audience the vigorous Sichuan through representational meaning, interactive meaning and compositional meaning. However, the construction of a region’s image does not only come from the surface material carriers such as regional landscape and infrastructure, but also includes non-material carriers such as regional reputation and service. For the production of regional promotional videos, many emphases should be placed on the following parts when choosing images: first, to enhance the relevance between images and text; second, to increase the number of images of non-material carriers, such as those reflecting people’s life. On the whole, the two bilingual promotional videos of Sichuan province have good interaction between verbal modality and visual modality which have multiple relationships, realizing the construction of the provincial image in the two videos, and shaping the image of Sichuan with good ecological environment, flourishing development of science and technology, continuation of history and culture, and ongoing economic take-off, which is conducive to enhancing the soft power of regional culture and promoting the overall development of the region.

Acknowledgments

This paper is supported by the project of Southwest Petroleum University Translation Research Center (The Image Construction in

Promotional Translations from Multimodal Perspective: Taking Sichuan Promotional Videos as Examples) (No. TRC202310).

References

- [1] Z. Li. Social Semiotic Analysis of Multimodal Discourse. Foreign Languages Research, 2003, (05):1-8.
- [2] Y. Zhu. Theory and Methodology of Multimodal Discourse Analysis. Foreign Language Research, 2007, (05):82-86.
- [3] H. Wang. The Construal of the Interactive Meaning of Multimodal Advertising Discourse. Journal of Sichuan International Studies University, 2007, (06):31-34.
- [4] Y. Peng, H. Yang. The Construal of the Integrative Meaning of Multimodal Commercial Advertising Discourse. Journal of University of Science and Technology Beijing(Social Sciences Edition), 2010, 26(04):1-6.
- [5] J. Lu, W. Wu. Subtitle Translation Study of The Flowers of War from the Perspective of Multimodal Discourse Analysis. Shanghai Journal of Translators, 2012, (04):36-38.
- [6] W. Ji. The Multimodal Discourse Restructuring in Culture-introducing Video Translation. Foreign Language Education, 2021, 42(05):82-86.
- [7] T. Zhu, L. Fan. Multimodal Discourse Analysis of Riyadh's Bid for Expo 2030 Promotional Video. Media Forum, 2024, 7(15):40-43.
- [8] F. Chen, J. Zha. The Image Construction of National Leaders Foreign Affairs Promotion Context from Multimodal Cognitive Perspective. Foreign Language Learning Theory and Practice, 2017, (03):42-49.
- [9] M. A. K. HALLIDAY. An Introduction to Functional Grammar. London: Arnold, 1985.
- [10] G. Huang. Systemic Functional Linguistics: Forty Years on. Foreign Language Teaching and Research, 2000(01):15-21+78.
- [11] G. Kress, T. van Leeuwen. Reading Images: The Grammar of Visual Design. London: Routledge, 2006.
- [12] D. Zhang. On a Synthetic Theoretical Framework for Multimodal Discourse Analysis. Foreign Languages in China, 2009, 6(01):24-30.
- [13] Y. Lin, X. Miao. Discourse Construction of China's National Image in American Mainstream Media: A Diachronic Study Based on Text Mining. Journal of PLA University of Foreign Languages, 2023, 46(03):87-95.
- [14] Y. Deng. Regional Image: Western Media Coverage and Our Constructive Strategies: The Case of Henan Province. Journal of Zhenzhou University, 2017, 50(02):154-157.
- [15] X. Fu, J. Xu. A Study of the Image of Shanghai in The New York Times and The Times (2010-2017). Journalism and Mass Communication, 2020, (02):80-87.
- [16] Y. Guo, X. Tan. The Construction Paths of China's Regional International Image Discursive System: Based on the Construction of Guangxi Image in Vietnam News. Journal of Anhui University of Science and Technology(Social Science) , 2023, 25(06):64-72.
- [17] D. Huang. On the Fujian Image Construction from the Perspective of Multimodal Discourse Analysis. Journal of Zhangzhou Institute of Technology, 2022, 24(03):64-71+81.
- [18] S. Liu. The Cognitive Construction of Multimodal Metaphors in Regional Image Publicity Films. Journal of Wuyi University(Social Science Edition) , 2022, 24(04):74-79+92.
- [19] J. Li, Y. Su. Multimodal Discursive Interaction Approach to Regional Image Construction: A case study of Fujian image publicity MV. Foreign Language and Literature Studies, 2022, 39(04):41-52+134.
- [20] Q. Qiu, Y. Zhang. Study on the Construction of National Economic Image from the Perspective of Multi-modal Discourse Analysis. Journal of Nanchang Normal University, 2023, 44(04):9-14.
- [21] D. Zhang. Syntax in Systemic Functional Linguistics. Journal of Tongji University(Social Science Edition), 2012, 23(01):89-98.
- [22] X. Miao. Interpersonal Meaning and the Construction of Discourse. Shangdong Foreign Language Teaching Journal, 2004, (01):5-11.
- [23] X. Yuan, R. Nai. Research on Multimodal

Discourse Meaning Construction in
International Communication of “Cultural

China”. Foreign Language Education,
2022, 43(05):23-29