Original Paper

Visual Grammar in Multimodal Discourse: A Comparison of British and Chinese Documentaries on Chinese Cuisine

Lihua Wang[1] & Jingxiang Cao[1*]

[1] School of Foreign Languages, Dalian University of Technology, Dalian, China

[*] Jingxiang Cao, School of Foreign Languages, Dalian University of Technology, Dalian, China

Abstract

This study, based on the Visual Grammar (VG), compares two Chinese cuisine documentaries: A Bite of China 1 from CCTV and Exploring China: A Culinary Adventure from BBC. It analyses the visual modes employed in the construction of meaning in each documentary. And we analyze and compare the China's national image presented in the two documentaries. The videos are annotated by ELAN 5.9. We found that (1) A Bite of China (CCTV) is inclined to adopt the conceptual process to construct the representational meaning while Exploring China (BBC) adopts the narrative process to construct the representational meaning. (2)In the construction of interactive meaning, both adopt offer images to realize contact with subtle differences in distribution. The distribution of close-up shots is similar. But A Bite of China (CCTV) has higher proportion of long shot and Exploring China (BBC) has higher proportion of medium shot. In the attitude dimension, both are similar in the horizontal perspective and vertical perspective with subtle differences in distribution. (3)In the construction of compositional meaning, the distribution of center-margin and left-right is similar. Both rarely adopts framing to construct the meaning in visual modes. (4) It is concluded that A Bite of China (CCTV) portrays China as colorful and profound, highlighting the hard-working, intelligent, and life-loving nature of Chinese people, while Exploring China focuses on social change. Despite some shared visual mode characteristics, the documentaries differ in their representational approaches and the representations of China's national image.

Keywords

visual grammar, Chinese cuisine documentary, comparative analysis

1. Introduction

With the rapid development of science and technology, especially the development of multimedia and

the internet, the way people accept information is no longer from a single text. Other modes also play a crucial role in the process of information transmission, such as image, color, sound, etc. Multimodal discourse integrates a variety of social symbols to deepen the visual, auditory, tactile, and other sensory experiences. Documentaries, as a prime example of multimodal discourse, skillfully combine language, images, sound, and other modes to deliver a rich sensory experience, thereby creating a lasting visual memory for the audience. Food Documentaries, in particular, have attracted much attention from society and academia. They provide viewers with a window into the natural features, customs, and cultural nuances associated with food. In May 2012, CCTV released *A Bite of China I*, capturing the audience's interest with its deep dive into Chinese culinary traditions. Later that year, the BBC aired *Exploring China: A Culinary Adventure*, which also aimed to convey the essence of Chinese culture through its culinary lens.

Both *A Bite of China I* and *Exploring China: A Culinary Adventure* focus primarily on Chinese food, using it as a conduit to explore and illustrate the rich tapestry of Chinese culture and tradition. This study wants to explore the differences in visual mode of meaning construction between *A Bite of China I* and *Exploring China: A Culinary Adventure*. And it also compares and analyses the China's national images constructed in documentaries. The research questions are as follows:

(1) What are the similarities and differences in visual modes between British and Chinese documentaries on Chinese cuisine?

(2) What are the differences in the representations of China's national images constructed in British and Chinese documentaries on Chinese cuisine?

2. Literature Review

Multimodal discourse refers to the discourse in which different semiotic modes are combined to convey meanings (Zhang, 2018). Multimodal discourse analysis (MDA) originates from the 1990s (Zhu, 2007). Its theoretical foundation is derived from systemic functional linguistics represented by Halliday. Kress & van (2006) extended it to representational meaning, interactive meaning and compositional meaning to establish a grammatical analysis framework of visual images.

Researchers have paid close attention to the prevalence of multimodal discourse, and as a result, they have conducted a variety of multimodal discourse analysis spanning various disciplines. Researches on MDA focus on multimodal critical discourse analysis (Tian & Pan, 2018), multimodal aesthetic discourse analysis (Dong & Yuan, 2021), multimodal metonymy and image meaning construction (Feng & Zhao, 2017), multimodal stylistic research (Yan et al., 2021), and multimodal translation research (Wu, 2021; Ji, 2021), etc.

As far as research methods are concerned, researchers have begun to use digital technology to annotate and simulate complex multimodal texts establish multimodal corpus and develop multimodal retrieval software (Baldry & Thi-bault, 2008; Gu, 2006). Meanwhile, empirical research on viewers' cognition of multimodal discourse has been gradually rising through questionnaire survey, eye movement

24

experiment and even brain imaging technology (Gidlof et al., 2012; Muller et al., 2012).

In application, relevant researches focus on static media such as print ads, political comics, posters, foreign language teaching materials, photographic images, and dynamic media such as TV advertisements, films and propaganda. However, the research on the comparison of meaning construction and China's national images on documentaries from a multimodal perspective is still insufficient.

3. Methodology

3.1 Materials, Instruments and Procedures

A Bite of China I (A Bite of China for short) delves into the pursuit of food and life, presenting Chinese cuisine to viewers and helping them better understand Chinese cultural traditions, family concepts, life attitudes, and attachment to their hometowns through various aspects related to food in daily life. It also explores the relationship between Chinese people and their food. *Exploring China: A Culinary Adventure* (Exploring China for short), is a four-episode British documentary television series aired on BBC. Ken Hom and Ching-He Huang, both Chinese food specialists, describe their travels through China and the recipes and personal stories they find there. They offer a unique perspective on Chinese food as they embark on a 3000-mile culinary journey across China, revealing not only its food but also its people, history, and culture.

Table 1. The Visual Materials

	A Bite of China (CCTV)	Exploring China (BBC)
Publishing (Year)	2012	2012
Duration (Second)	2966	3551

All episodes of the two documentaries have been downloaded and all the commentary has been collected. After watching these episodes, the first episodes of each documentary, titled "The Gift of Nature" and "Beijing" are selected as the research material. As shown in Figure 1 and Figure 2, both videos were fully annotated with ELAN 5.9. The total duration of A Bite of China's annotation was 2966 seconds and the total duration of Exploring China's annotation was 3551 seconds.
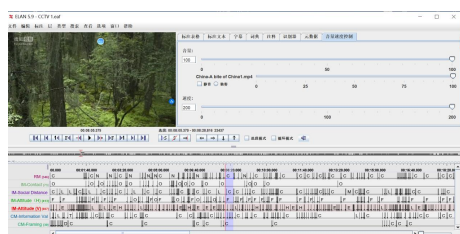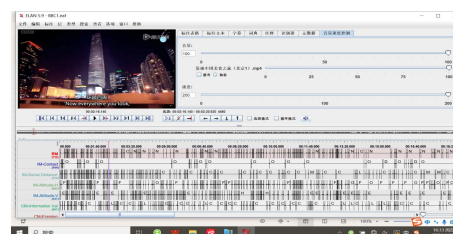

Figure 1. Annotation of A Bite of China

Figure 2. Annotation of Exploring China

This study employs two research instruments, Excel and ELAN 5.9, to answer the research questions through both qualitative and quantitative methods. Based on Visual Grammar, a detailed analysis of visual modes was conducted using ELAN 5.9, allowing for the identification of similarities and differences between the two documentaries in visual modes, as well as the differences in the construction of Chinese images.

3.2 Theoretical Framework

Inspired by Halliday's Systemic Functional Grammar (SFG), (Halliday, 1978), Kress and van (2006) creatively proposed a Visual Grammar (VG) applicable to visual modes in the book *Reading Images The Grammar of Visual Design*. They believe that images also have three meta-functions, namely, representational meaning, interactive meaning and compositional meaning. Representational meaning visually captures the essence of events, participants, and environment in the images. The implementation of representational meaning is divided into narrative process and conceptual process. As Kress and van Leeuwen differentiate, the former serves to illustrate "unfolding actions and events, processes of change, transitory spatial arrangements" while the latter represents participants "in terms of their class, structure or meaning, in other words, in terms of their generalized and more or less stable and timeless essence" (Kress & van, 2006). Notably, narrative process bears a hallmark that the conceptual one is not endowed with: the vector – a set of oblique or diagonal lines formed by depicted images to connect one or more participants.

"The interactive meaning is associated with the social relationship between interactants and the evaluative orientations that participants adopt to each other and towards the represented world of the text" (Kress & van, 2006). The achievement of interactive meaning depends on three elements: contact, social distance and attitude. Contact identifies interactive relations in pictures from which represented participates look directly at the viewer's eyes, or in which this is not the case. To be more exact, images make "demands" in the former case because they "connect the viewer and participant on a formal (and imaginary) level" and "offers" in the latter when "represented participants look away from the viewer" (Bell & Milic, 2002). Social distance suggests different relations between represented participants and viewers by means of the choice between close-up, median and long shot. The closer the shot is, the closer the relationship between the represented participants and viewers. Close-up shot shows personal distance. Medium shot shows close social distance. Long shot shows public distance. Attitude can be realized through the selection of an angle or a point of view, implying the subjective or objective attitude towards represented participants. In the horizontal level, a frontal angle indicates a sort of involvement and an oblique angle a sort of detachment; while in the vertical level, a high angle hints at viewer power, an eye-level angle at equality, and a low angle at represented participant power.

Corresponding to the textual meaning in systemic functional grammar, Kress and van Leeuwen (2006) propose the concept of compositional meaning in visual grammar. Compositional meaning refers to the composition of the whole, that is, "the way in which the representational and interactive elements are made to relate to each other, the way they are integrated into a meaningful whole" (Kress & van, 2006).

26

It is also a way to express the relationship between the participants and the audience of the documentary, and will be analyzed from three aspects: information value, salience and framing. Information value refers to that the placement of different elements gives them with specific information values in visual composition. It is usually achieved by the placement of elements in the image for the various "zones" of the image: left and right, top and bottom, center and margin. The left-right layout refers to given and new information, the central-margin layout to primary and secondary information, and the top-bottom layout to ideal and real information. Salience draws the viewer's attention with the elements of the composition, which can be achieved through color, relative size, angle of view, and foreground or background. Framing refers to the dividing lines formed by solid or dotted lines in an image that connect or cut elements in the image. Salience and framing are closely related. Therefore, visual grammar can be used to analyze the visual mode in documentary.

4. Results and Discussion

4.1 Representational Meaning

The details are shown in Table 2. The proportion of narrative process and conceptual process is different between A Bite of China and Exploring China. Narrative process is occupied more than conceptual process in Exploring China. But it appears the opposite in A Bite of China.

Table 2. Representational Meaning in Documentaries

Process Types	A Bite of China (CCTV)		Exploring China (BBC)	
	Time (s)	%	Time (s)	%
Conceptual	1690.00	56.98	1380.21	38.87
Narrative	1276.00	43.02	2170.79	61.13
Total	2966.00	100	3551.00	100

4.1.1 Narrative Process

A Bite of China introduces six kinds of food acquisition activities, including picking matsutake, digging for bamboo shoots, making Nuodeng ham, digging for lotus roots, Chagan Lake fishing, and deep-sea fishing. These scenes depict people working hard to obtain food sources.

Figure 3. Narrative Process (CCTV)

Figure 3 is the narrative process of A Bite of China. In a and b, individuals are shown working, making them the actors, while the food ingredients, such as matsutake and fish, are the goals. Their arms and the tools in their hands form vectors, illustrating an action process. Additionally, these figures exhibit a reactional process. When a vector is formed by the eyeline of a participant, the narrative process is classified as reactional. Here, the workers are also reactors, and the food ingredients serve as the phenomenon. Their eye-lines create the vector in each figure. Both figures feature more than one person, emphasizing teamwork and collective effort, highlighting the hard work of people in nature. This portrayal suggests that they obtain gifts from nature through their labor.

Figure 3's c and d depict folk activities, illustrating a society that respects tradition and nature. In these images, individuals are both actors and reactors, with the bowl and the signboard for divination acting as the goal and phenomenon, respectively. The participants' arms and eyelines form vectors. "c" shows a blessing ceremony held annually during the fishing season at Chagan Lake, which has evolved into a cultural performance attracting many spectators while retaining its original meaning of hoping for a good harvest. "d" portrays deep-sea fishing, with Captain Lin blessing for a safe return and a good harvest. These scenes reflect the wishes of people, who first pray to nature and then do what they can to achieve their goals.

Figure 4. Narrative Process (BBC)

In Exploring China, the narrative process is prominent due to the presence of many characters, primarily the two hosts and Chinese people. The hosts share their own experiences, engage in

28

conversations with Chinese chefs, and cook Chinese dishes themselves, exchanging views along the way. Through their dialogues and culinary activities, the documentary introduces a variety of Chinese foods, such as noodles, dumplings, street food, and Beijing roast duck. In Figure 4's a, the hosts and Chinese chefs are actors in the narrative process, with the food they are preparing serving as the goal. Their arms and the kitchenware they hold, along with their eyelines, form the vectors. Figure 4's b depicts a woman entering a public restroom, followed by a conversation between the two hosts. This scene features a reactional process, where the two hosts are sensors, and their eyelines form the vector. The woman's surprise at the lack of partitions in the restroom highlights a cultural difference, emphasizing the lack of privacy in this public space.

4.1.2 Conceptual Process

The conceptual process presents relevant information such as history, culture, and context, without the presence of vectors. Given that A Bite of China is a food documentary, it includes numerous introductions to food, from raw ingredients in nature to prepared dishes, as shown in Figure 5's a. The documentary highlights the simplicity of cooking, using basic preparation methods and common ingredients, reflecting the natural and straightforward eating habits of Chinese people. Beyond food, A Bite of China also introduces various natural landscapes, including high mountains, bamboo forests, lakes, seas, and plains, showcasing China's rich natural resources. People living in these environments take from nature and give back to it, demonstrating respect and harmony with the natural world. Figure 5's b exemplifies these natural scenes, underscoring the interconnectedness of people and their environment.
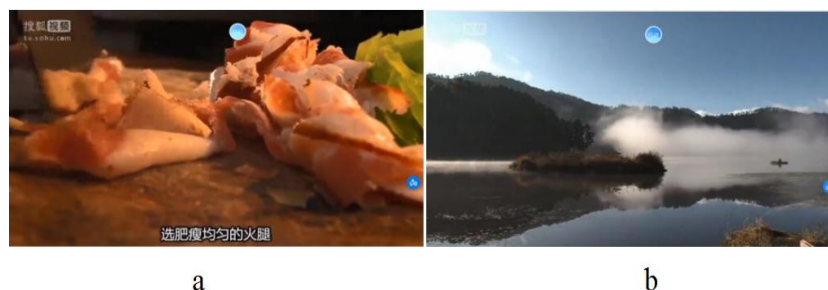

a   b

Figure 5. Conceptual Process (CCTV)

Unlike A Bite of China, Exploring China places less emphasis on the conceptual process. It does not extensively showcase the origins of the food. Instead, it focuses more on presenting Beijing's urban architecture and rural features. Over the past few decades, China's urban landscape has undergone significant changes, with tall buildings rising and many young people moving to big cities for work and life. An implicit theme of the documentary is whether Chinese food culture has been preserved or altered amidst these changes.

Figure 6's a shows food items, such as elbows and pig feet, at a wet market, which the hosts believe are favored by Chinese people. This image lacks participants and vectors, merely conveying a message.

Figure 6's b depicts the abundance of fruits and vegetables in the same market, contrasting with a discussion between the hosts about food shortages during China's natural disasters. This contrast highlights the improvements in food availability over time. Figure 6's c and d show the Forbidden City and Beijing's hutongs, respectively. These images, devoid of participants and vectors, focus on China's historical and cultural context. While food is the explicit theme, the implicit theme is China's urban development.

Figure 6. Conceptual Process (BBC)

The two documentaries differ in their use of representational meaning. The narrative process in Exploring China focuses on what people are doing and their visual communication, while the conceptual process provides context, cultural, and historical information, such as natural landscapes, urban streets, dishes, and raw materials. A Bite of China employs the conceptual process more frequently but also makes significant use of the narrative process. It showcases activities like digging for bamboo shoots, catching fish, and performing folk rituals, highlighting the natural environment and ancient villages. This documentary constructs meaning through people's actions and the introduction of natural scenes, rather than through dialogue, emphasizing the wisdom, traditions, and reverence for nature inherent in Chinese culture. Exploring China favors the narrative process, beginning with the interactions between the hosts and Chinese chefs and their culinary experiences. The documentary transitions from urban kitchens to rural homes, constantly changing settings. While food is the explicit theme, the implicit theme is China's development. The documentary aims to explore China's growth and changes through the lens of food, presenting the evolving society and the future of its food culture to a Western audience.

4.2 Interactive Meaning

Visual modes are crucial in defining the relationship between participants and audiences, enabling the latter to gather information and interpret attitudes, thereby contributing to the creation of interactive meaning. This section employs three elements—contact, social distance, and perspective—to measure

30

interactive meaning.

4.2.1 Contact

Contact in visual communication identifies the interactive relations in pictures based on whether the represented participants look directly at the viewer's eyes (demand contact) or look away (offer contact).

Table 3. Contact in Documentaries

Contact Types	A Bite of China (CCTV)		Exploring China (BBC)	
	Time (s)	%	Time (s)	%
Demand	66.12	2.23	171.15	4.82
Offer	2899.88	97.77	3379.85	95.18
Total	2966.00	100	3551.00	100

Table 3 shows the duration and percentage of demand contact and offer contact. Through the statistic contrasted parallelly, the contact types in the two documentaries of visual modes vary slightly. It proves the documentaries shoulder the tasks of propaganda by applying "offer" images, in which an enormous amount of information about the food and Chinese culture is provided, instead of demanding something from the audiences.

Demand images, although less frequent, play a critical role in establishing a communicative relationship with the audience, thereby attracting more viewers through a sense of social affinity. Figure 7 illustrates demand contact. Figure 7's a, from A Bite of China, shows two participants smiling directly at the camera, inviting the audience to engage with them. Figure 7's b, from Exploring China, features Miss Huang making direct eye contact with the audience, sharing her experiences and reactions to the changes in Chinese villages. These demand images create intimate connections with the audience, encouraging viewer engagement.
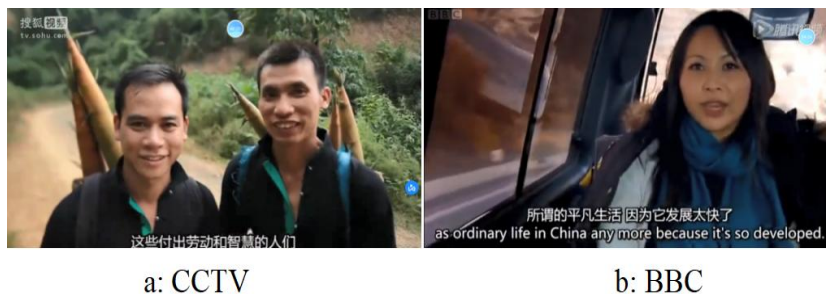

a: CCTV   b: BBC

Figure 7. Demand Contact

Offer images, in contrast, allow the audience to observe the participants without direct engagement. Figure 8 exemplifies offer contact. Figure 8's a, from A Bite of China, shows the process of making

31

Nuodeng hams without direct eye contact with the viewers, focusing on the relationship between the participants and their environment. The offer contact allows viewers to observe the process and the familial bond without direct engagement. Figure 8's b, from Exploring China, depicts the hosts cooking in a modern kitchen, emphasizing the interaction between the hosts and Chinese chefs without direct viewer engagement. The offer contact highlights the communication and collaboration between the participants, allowing viewers to observe the scene without direct interaction.

a: CCTV   b: BBC

Figure 8. Offer Contact

The low percentage of demand contact in both documentaries indicates a focus on presenting information rather than direct viewer engagement. Exploring China uses more demand contact than A Bite of China, reflecting the hosts' direct introductions and personal experiences shared with the audience. In contrast, A Bite of China captures the simplicity and sincerity of Chinese people's lives, emphasizing family and homeland identity.

4.2.2 Social Distance

Social distance suggests different relations between represented participants and viewers by means of the choice between close-up, median and long shot. The closer the shot is, the closer the relationship between the represented participants and viewers. Close-up shot shows personal distance. Medium shot shows close social distance. Long shot shows public distance.

Table 4. Social Distance in Documentaries

Social distance	A Bite of China (CCTV)		Exploring China (BBC)	
	Time (s)	%	Time (s)	%
Close-up Shot	1812.69	61.12	2477.63	69.77
Median Shot	571.56	19.27	779.42	21.95
Long Shot	581.75	19.61	293.95	8.28
Total	2966.00	100	3551.00	100

Table 4 shows the duration and percentage of three kinds of social distance. Close-up shots account for over 60%, with the percentage reaching as high as 69.77% in Exploring China and 61.12% in A Bite of

China. However, the distribution of median and long shots varies, with A Bite of China using more long shots to provide a broader perspective on Chinese food culture, while Exploring China uses more median shots, reflecting its emphasis on personal experiences and interactions.

a: CCTV   b: BBC

Figure 9. Close-up Shot

Figure 9 is close-up shot. It is also the most used shot in the two documentaries which shortens the distance between the viewers. In Figure 9's a, a close-up shot of Huang Shujiang carries brine, showcasing the hard work and traditional practices of making Nuodeng salt, emphasizing the personal and intimate connection with the audience. In Figure 9's b, the two hosts walk down a snack street, encountering crickets, with Huang's expression of surprise and disgust. This close-up shot highlights personal reactions and cultural stereotypes. .

a: CCTV   b: BBC

Figure 10. Median Shot

Figure 10 is median shot. In Figure 10's a, Lao Bao searches for bamboo shoots, showing his familiarity with the land and tools, emphasizing the connection between man and nature. In Figure 10's b, four participants make dumplings, with the hosts carefully learning the process. This median shot highlights the cultural exchange and the hosts' engagement with Chinese cuisine.
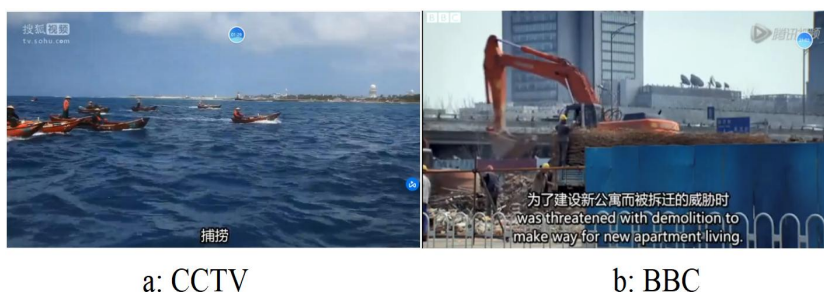
a: CCTV   b: BBC

Figure 11. Long Shot

Figure 11 is long shot. There is more than one participant in the two figures. In Figure 11's a, people go out in boats for ocean fishing, with the vast sky and sea creating a sense of human insignificance in nature, emphasizing respect for the natural world. In Figure 11's b, the demolition of hutongs to make way for new apartments in Beijing, showing the transformation and development of urban China, reflecting the documentary's exploration of traditional and modern aspects. Partly like "offer" images, the long shot also bears the task of conveying the message to viewers. It implies a subtle form of detachment, enabling the viewers to feel close but not related to the participants. At this distance, an invisible barrier lies between the participants and the viewers, serving the public distance.

The prevalence of close-up shots in both documentaries indicates an intention to create a personal connection with the audience. A Bite of China uses long shots more frequently to provide a macro perspective on Chinese cuisine and culture, while Exploring China focuses on personal experiences and interactions, using more median shots.

4.2.3 Attitude

Besides the contact and social distance, the realization of interactive meaning can also be measured by the application of attitude, namely, studying the angle of shots in visual modes.

Table 5. Attitude in Documentaries

Attitude		A Bite of China (CCTV)		Exploring China (BBC)	
		Time (s)	%	Time (s)	%
Horizontal	Frontal	2073.34	69.90	1979.80	55.75
	Oblique	892.66	30.10	1571.20	44.25
	Total	2966.00	100	3551.00	100
Vertical	High	1051.80	35.46	1406.36	39.60
	Eye	1378.46	46.48	1790.27	50.42
	Low	535.74	18.06	354.37	9.98
	Total	2966.00	100	3551.00	100

Table 5 shows the duration and percentage of attitude types in the documentaries. In the horizontal

34

level, the frontal angle is used more than oblique angle, accounting for 69.90% in A Bite of China and 55.75% in Exploration China. Among the vertical angle with the function of balancing the power relations, the eye-level angle leads the percentage up to 46.48% and 50.42% separately in two groups of documentaries. Both documentaries use high and low angles to convey different power relations, but Exploring China uses more high angles, reflecting its focus on the broader context of Chinese development and culture.

Figure 12. Attitude

Figure 12 can be analyzed from the horizontal and vertical level. In "a", the participant of A Bite of China is walking towards the audience, making sour bamboo shoots. In "b", four participants, including the two hosts, face the audience directly. The front angle in these two figures creates an involvement of interactive participants with the viewers, making the audiences feel engaged with the participants. "c and d" are oblique angle. "c" from A Bite of China depicts a picture that fishermen finish fishing in Chagan Lake and are drawing in their nets, showing their body movements as they pull in the net. "d" from Exploring China shows the communication between Ken Hon and Chinese chef Da dong. We can see their side faces, the movement of their hands and the expression of concentration, talking about how to revive the reputation of Chinese food. The viewers possibly feel detached from the participant. In accordance with these examples, the interactive meaning is clearly conveyed by different types of horizontal angles, helping the audience to learn about the unique culture of China. And the figures about vertical angles are discussed below.

In the vertical level, "e and f" are the high angle from A Bite of China and Exploring China respectively. "e" captures a deep-sea fishing scene, portraying participants, the sea, and fishing nets from an elevated

35

viewpoint. This high-angle approach affords viewers a comprehensive view of the fishing process, establishing a hierarchical relationship where viewers hold observational dominance over the participants. Similarly, "f" illustrates the expansive suburbs of Beijing and towering buildings, presenting an aerial view that grants viewers a sense of spectacle and achievement. This perspective shares visual power with viewers, emphasizing the grandeur of the scene. "e" shows the deep-sea fishing in the dark. We could see the participants, the sea and the fishing net. From top to bottom, we can see the whole fishing process. And the viewers' power is higher than the participants. "f" shows the high buildings and the dense suburbs of Beijing. The overall view of Beijing suburbs is presented from a high angle to share the power with viewers, in which the achievement seems spectacular. "b and c" are the low angle from Exploring China and A Bite of China respectively. It contributes to balancing the power relation between the participants and the viewers. "b" depicts hosts engaged in cooking, with the audience positioned lower relative to the hosts, symbolizing a perspective that elevates the hosts' stature in the culinary process. "c" shows the fishing process, highlighting the intricate details of everyday activities and fostering a deeper appreciation for the participants' endeavors. Overall, the eye-level angle is dominant in vertical angles, resulting in the function of sharing the equality between participants and viewers. "a and d" from A Bite of China and Exploring China respectively concentrate on the participants from an eye-level angle. It narrows the distance between the participant and the audience, and painting the scene closer to real life.

On the horizontal level, frontal angle is used more to detach the audience. On the vertical level, the eye level is used at most to show the equality between the participants and the viewers; However, the eye-level angle takes up more than half of Exploring China, and the proportion of high angle will be larger, so as to show Chinese food and cultural traditions. The low angle of A Bite of China focuses on the subtleties of everyday life.

4.3 Compositional Meaning

Compositional meaning is a way to express the relationship between the participants and the audience of the documentary, and will be analyzed from three aspects: information value, salience and framing.

4.3.1 Information Value

In general, the position of each element of an image can effectively convey some specific information. The left-right layout refers to given and new information, the central-margin layout to primary and secondary information, and the top-bottom layout to ideal and real information.

Table 6. Information Value in Documentaries

Information Value	A Bite of China (CCTV)		Exploring China (BBC)	
	Time (s)	%	Time (s)	%
Central-margin	2172.15	73.23	2658.71	74.87
Left-right	528.32	17.81	768.07	21.63

| Top-bottom | 265.53 | 8.95 | 124.22 | 3.50 |
| Total | 2966.00 | 100 | 3551.00 | 100 |

Table 6 shows the duration and percentage of information value types. The central-margin images play a major role in conveying the information, occupying 73.23% in A Bite of China and 74.87% in Exploring China, because the center in the image may show a key religious theme. The frequency of left-right images ranking the second in the whole information value types accounts for 17.81% in A Bite of China and 21.63% in Exploring China. Unlike the two information value types mentioned above, the top-bottom information value type is barely used in documentaries. The following figures featuring the three different information value types are separately analyzed by the placement of specific elements.

First it comes to the most frequently used structure on information value dimension, the central-margin image. Figure 13's a from A Bite of China shows workers in the center with the lake and boat in the margins, highlighting the workers' journey and challenges they face. The participants in the center of the image are the most important part, conveying the positive message of the Chinese people. Similarly, Figure 13's b from Exploring China places the two hosts centrally, engaging with Chinese chefs in the kitchen setting, emphasizing their role in culinary exploration. They believe their work could impress this Chinese chef and they will win the battle.


a: CCTV   b: BBC

Figure 13. Central-margin

The top-bottom structure of the image in documentaries is typically used to present the views of the city. Figure 14 depicts this format, typically showing ideal versus real scenarios or higher versus lower statuses. For instance, Figure 14's a portrays city lights above and urban scenes below, symbolizing aspiration and reality in urban life. The whole view displays a prosperous picture of the city, therefore leaving a positive impression on the audience. While Figure 14's b contrasts sky and mountains with village houses, highlighting rural life's connection to nature and history. The overall color of this image is dark and foggy, displaying a backward and poor picture of rural areas.

a: CCTV   b: BBC
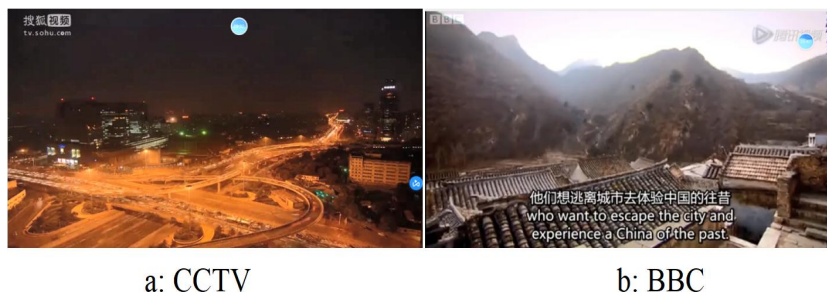
Figure 14. Top-bottom

The information values in Figure 15 distribute left and right. In Figure 15's a from A Bite of China, the comparison happens between the object and participant. It contrasts horses and Mr. Shi, the fisherman, highlighting Mr. Shi's role as a central figure in the image. In Exploring China, Figure 15's b contrasts the hostess and Mr. Han, underscoring Mr. Han's status as a key figure in rural life narration

.   a: CCTV   b: BBC

Figure 15. Left-right

4.3.2 Salience and Framing

In documentaries, salience and framing are closely related. The images always highlight the main characters with color, foreground, size. It is difficult to annotate the salience because they are fused into each other in the screenshot of each frame. Two examples are used here to explain salience. As shown in Figure 14's a, bright lights highlight the bustling and busy city at night. Figure 13's b uses foreground positioning to emphasize the confidence and expectation of the hosts.

Table 7. Framing in Documentaries

Framing	A Bite of China (CCTV)		Exploring China (BBC)	
	Time (s)	%	Time (s)	%
Connection	2878.66	97.06	3506.42	98.74
Disconnection	87.34	2.94	44.58	1.26
Total	2966.00	100	3551.00	100

38

Table 7 shows the duration and percentage of framing. The connection framing plays a major role in conveying the information, occupying 97.06% in A Bite of China and 98.74% in Exploring China. It reveals that the documentaries bear less need for the use of framing and thus rely less on such devices to construct the layout of visual modes because the elements in documentaries blend with each other and belong together.

The disconnection framing separates sky and forest distinctly, reflecting a contrast between natural purity and human impact, while Figure 16's b employs map lines to focus on Beijing's location within China, framing geographical significance.

Both documentaries demonstrate similar patterns in their use of compositional elements. They favor central-margin layouts to highlight key narratives and characters, employ salience techniques to direct viewer attention effectively, and utilize connection framing to maintain visual coherence. These strategies ensure clarity and engagement in storytelling, catering to audience expectations and thematic emphasis. While similarities dominate, nuances in layout preferences and framing techniques do exist. Exploring China tends to slightly favor left-right layouts more than A Bite of China, suggesting a nuanced approach to presenting sequential information. Moreover, A Bite of China utilizes top-bottom layouts more distinctly to contrast ideal and real scenarios, whereas Exploring China employs such layouts less frequently, focusing more on central-margin and left-right arrangements. Compositional meaning in visual grammar, as applied in A Bite of China and Exploring China, underscores the deliberate structuring of visual elements to convey coherent narratives, engage viewers, and highlight thematic messages. These documentaries effectively leverage visual communication to enrich audience' understanding of Chinese culture, cuisine, and societal contexts.

5. Conclusion

Symbolic resources appear increasingly indispensable in human communication activities thereby drawing forth the overwhelming progress of multimodal discourse analysis. Documentaries fall into the category of multimodal discourse and undergo changes, which serves as the research material. This study compares the construction of representational meaning, interactive meaning as well compositional meaning. And it analyses and compares the China's national images constructed in the two documentaries.

In the representational meaning, the difference lies in the fact that A Bite of China favors the conceptual process while Exploring China favors the narrative process. In the interactive meaning, both share the similarity that the "offer" images outnumber the "demand" images. A Bite of China focuses on public distance while Exploring China prefers social distance. As for the attitude, the frontal angle ranks first; The eye-level angle leads a high usage rate in the two documentaries. Three interrelated subsystems are employed to measure the compositional meaning. As to information value, both use many central-margin images with the left-right images ranking second. From the perspective of salience and framing, the similarities indicate that size and position are the most frequently employed

techniques. Both rarely adopts framing to construct the meaning in visual modes. The meaning realization in CCTV and BBC share similarities and differences to some extent, resulting from the different cultures as well as the similar contexts of food documentaries. It reveals meaning construction is a selection process based on the meta-functions on different levels, thus creating a colorful multimodal discourse that can only be understood by those who comprehend the corresponding cultures.

As for the China's national image, the CCTV documentary *A Bite of China I* portrays China as colorful and profound, highlighting the hard-working, intelligent, and life-loving nature of Chinese people. For Chinese people, food culture is an integral part of their cultural tradition (Chen & Chang, 2019). Firstly, it showcases a rich tapestry of Chinese ingredients, cooking techniques, and eating habits. Beyond mere culinary practices, each dish and ingredient is imbued with stories and cultural meanings, providing viewers with insights into Chinese cultural heritage. Secondly, through scenes of fishermen, farmers, chefs, and diners, the documentary portrays Chinese people as diligent, resourceful, and deeply connected to their culinary traditions. It highlights the resilience and optimism of ordinary Chinese people in facing challenges, as well as their profound respect for nature and family. Finally, it presents China's national image and cultural soft power through the unique perspective of food. By showing the charm and cultural connotation of Chinese cuisine, the documentary not only attracts the attention and love of domestic and foreign audiences, but also enhances the international influence.

The BBC documentary *Exploring China: A Culinary Adventure* focuses on social change. It is not only intended to explore the flavor of ancient Chinese food, but also to explore whether traditional Chinese culture will be submerged by the rapid development of modern society, expressing the hidden worry about the demise of traditional culture. Despite some shared visual mode characteristics, the documentaries significantly differ in their representational approaches and the images of China they project. Firstly, instead of tasting roast duck and other authentic Beijing dishes at famous restaurants, the two hosts go to a Hutong restaurant that is nearly demolished. Therefore, the dilapidated alley, the messy dormitory, the bustling but not tidy food market, the kitchen with no modern atmosphere, the city with no sense of order, the desolate village, etc., are "naturally" into the camera. China as "the other" has undoubtedly been deliberately dwarfed. The image of China in BBC documentaries is the "imagination of the other" under the strong discourse of the West, and it is a way of expression of self-examination, self-imagination and writing in the Western society (Zhao, 2019). Of course, the documentary also introduces a group of hard-working and intelligent Chinese people. The master of making noodles has a superb skill, but it is difficult for ordinary people to practice. Chef Liqun, who once had to make a living, has returned to revive the traditional roast duck. It is gratifying that they have preserved the tradition through their own efforts. Here the addition of "the other" perspective also makes the narration of the documentary more objective and real (Chen & Peng, 2020). Due to differences in cultural backgrounds and language barriers, there are inherent challenges in accurately representing the essence and complexity of Chinese culture. This can sometimes lead to

misunderstandings or oversimplifications in cross-cultural documentaries. For any documentary, we should maintain a prudent attitude and distinguish carefully to avoid being misled by false information. At the same time, we should also actively promote the inheritance and development of our own culture, and enhance cultural confidence and cultural consciousness.

This study is to enhance the understanding of multimodal discourse analysis by examining documentaries using linguistic, semiotic, and film and television media knowledge, exploring the importance of continuous visual narrative discourse. But several limitations are unavoidable during the research. Initially, the study is constrained by a relatively small sample size, consisting of only 2 episodes. Additionally, visual images are annotated by a single coder, potentially resulting in some ambiguities. As a result, future research could consider expanding the scope of materials for investigating multimodal discourse further. Moreover, the relation among the multi-modes, especially the multimodal coordination, deserves some profound study.

Acknowledgments

References

Baldry. A., & Thibault, P. (2008). Application of Multimodal Concordances. *Journal of Language and Communication Studies*, *41*, 12-41.

Bell, M. Milic. Goffman. (2002). Gender Advertisements Revisited: Combining Content Analysis With Semiotic Analysis. *Visual Communication*, *1*(2), 203-222. https://doi.org/10.1177/147035720200100205

Chen, W., & Chang, M. (2019). Cross-cultural Narrative and Image Construction of "Cultural China": A case study of BBC Documentaries on Chinese Culture (in Chinese). *Television Research*, (3), 78-80.

Chen, W., & Peng, Z. (2020). The Metaphorical Construction and Multimodal Narrative of "Cultural China" Image (in Chinese). *Television Research*, (2), 83-86.

Dong, M., & Yuan, X. (2021). A Study on the Construction of Discourse Analysis Framework for Multimodal Aesthetic Criticism (in Chinese). *Foreign Language Teaching*, (1), 77-82.

Feng, D., & Zhao, X. (2017). Multimodal Metonymy and Image Discourse Meaning Construction (in Chinese). *Journal of Foreign Languages*, (6), 8-13.

Gidlof, K., Holmberg, N., & Sandberg, H. (2012). The Use of Eye Interviews to Study Teenagers' Exposure to Online Advertising. *Visual Communication*, 11, 329-345. https://doi.org/10.1177/1470357212446412

Gu, Y. (2006). Multimodal Text Analysis: A Corpus Based Approach to Situated Discourse. *Text and Talk*, *26*, 127-167. https://doi.org/10.1515/TEXT.2006.007

Halliday, M.A.K. (1978). *Language as Social Semiotic*. London: Arnold.

Ji, W. (2021). Multimodal Collaborative Reconstruction of Cultural Propaganda Films in External Translation and Communication (in Chinese). *Foreign Language Teaching*, (9), 82-86.

Kress, G., & van Leeuwen. (2006). *Reading Images: The Grammar of Visual Design*. London: Routledge. https://doi.org/10.4324/9780203619728

Muller, M. G., Kappas, A., & Old, B. (2012). Perceiving Press Photography: A New Integrative Model, Combing Iconology with Psycho Physiological and Eye-Tracking Methods. *Visual Communication*, *11*, 307-328. https://doi.org/10.1177/1470357212446410

Tian, H., & Pan, Y. (2018). From Meaning to Intention: New developments in Multimodal Discourse Analysis to Multimodal Critical Discourse Analysis (in Chinese). *Shandong Foreign Language Teaching*, (1), 23-33.

Wu, J., & Li, Z., & Zhang, H. (2021). Research on Multimodal Translation and National Image Construction of Commentary in Beautiful China (in Chinese). *Shandong Foreign Language Teaching*, (5), 31-41.

Yan, Y., Lei, X., & Gao, X. (2021). A Comparative Study on Multimodal Style of Home Pages of Chinese and Foreign University Portals (in Chinese). *Journal of Xi'an International Studies University*, (4), 41-45, 86.

Zhang, D. (2018). Comprehensive Framework of Multimodal Discourse Analysis under the Perspective of System Function Theory (in Chinese). *Modern Foreign Languages*, *41*(06), 731-743.

Zhao, S. (2019). A Study on the Value Construction of Chinese-related documentaries from the Cross-cultural Perspective: A Case Study of BBC Chinese Documentaries (in Chinese). *Audiovisual*, (12), 22-24.

Zhu, Y. (2007). Theory and Methodology of Multimodal Discourse Analysis (in Chinese). *Journal of Foreign Languages*, (5), 82-86.