Multi-Modal Discourse Analysis of Public Crisis Anti-Epidemic Documentaries—Take *The City of Heroes* as an Example

Min Zhang

Chongqing Normal University, Chongqing, China

Abstract—With the rapid development of multi-modal discourse analysis theory, discourse research is no longer limited to text analysis. The documentary brings together multiple modalities such as thematic text, subtitles, images, background music, etc., providing researchers with new ideas. This article takes the documentary *The City of Heroes* launched by Xinhua News Agency as an example, analyzes the discourse of visual, voice and image modality, and explores the meaning conveyed by the documentary in a public crisis, which is conducive to expanding the multi-modality in the documentary. By discourse analysis, we can understand the way it realizes the meaning of the text.

Index Terms—public crisis, documentary, multi-modal, *The City of Heroes*

I. INTRODUCTION

Multimodality refers to communication through multiple senses such as hearing, sight, touch, etc., with the help of language, image, sound and other means and symbol resources. That is, the transmission and communication of information is not the result of a single modality, but the result of multi-modal interaction (Zhang, 2018). Multimodal discourse analysis emerged in the 1990s, and its theoretical foundation is derived from systemic functional linguistics represented by Halliday. Functional linguistics theory regards language as a social symbol, and studies language from the interactive relationship between language and social and cultural situations. It believes that language has three meta-functions: conceptual function, interpersonal function, and textual function (Li, 2003). Kress & Leeuwen（1996） extended it to represent meaning, interaction meaning and composition meaning to establish a grammatical analysis framework of visual images, which corresponds to the three major functions of the system(Kress, Leeuwen, 1996). Li Zhanzi (2003) first introduced it to my country (Li, 2003). Other scholars in China have all carried out rich research on multi-modality.

As a cross-cultural, cross-temporal media form, documentary has an irreplaceable cultural mission (Zhang, 2020). The new pneumonia that broke out in 2019 has been positioned as a major public health emergency with the fastest spread and the most difficult prevention and control in my country since the founding of the People's Republic of China. In order to commemorate this epidemic prevention, Xinhua News Agency released the first panoramic documentary *The City of Heroes*, which recorded major news events from the complete closure of Wuhan to the release of the closure. By analyzing the three modalities of its multi-modal discourse media system and conducting discourse analysis, the author explores the meaning conveyed by documentaries during public crises, which is conducive to expanding the discourse analysis of multi-modality in documentaries and understanding its realization language and the way to the meaning of it.

II. TEXTUAL MODAL DISCOURSE ANALYSIS

A. Subject Textual Analysis

The subject text is the title of a film, a summary word for a certain segment. In *The City of Heroes*, with time as the axis and the development of character stories as clues, it is divided into six chapters in total. Connecting these six thematic texts together is the complete summary of the entire Wuhan closure to unblocking event. The text of each topic and the content of the scene are shown in Table 1.

TABLE 1
SUBJECT TEXT AND SCENE

Number	Subject text	Scene
No.1	Closing off outbound channels	The epidemic broke out and Wuhan was closed down.
No.2	Saving Lives	Airport: Military medical team and national medical staff support Wuhan; Hospital: Emergency treatment of lives.
No.3	Crucial Movement	Construction site: Construction of Huoshenshan Hospital and Leishenshan Hospital.
No.4	Everyone is a Hero	Qingheju Community: Temperature measurement, isolation, closed management, local resources to support Wuhan.
No.5	Great Love	Huoshenshan Hospital: 26-year-old granddaughter takes care of 89-year-old grandmother.
No.6	Spring Arrives After Winter	Wuhan: Patients are cured, new lives come, Wuhan regains life.

The subject textual text is the title of a film, a summary word for a certain segment. In *The City of Heroes,* with time as the axis and the development of character stories as clues, it is divided into six chapters in total. Connecting these six thematic texts together is the complete summary of the entire Wuhan closure to unblocking event. The text of each topic and the content of the scene are shown in Table 1.

The subject text is a theme of the scene. As shown in Fig.1, the four characters of the film name *The City of Heroes* are Xingkai font in Chinese character, which are always kept in the lower right corner of the entire screen. The purpose is to highlight the theme of the entire film and praise the heroic image of Wuhan. The picture is set with the silent Yangtze River Bridge and the dim Yangtze River as the background. In the middle of the whole picture, it says "Closing off outbound channels" which is also in Xingkai font, occupying the center of the whole picture, highlighting its main position. The viewer is told that the main message of this chapter is the suspension of various outbound channels. At the same time, it echoes the title of the movie *The City of Heroes* in the lower right corner. This indicates that the fight against pneumonia "Wuhan War" really started.

Figure 1

It can be seen from this that the function of the subject text is to summarize a certain segment, inform the viewer in advance of the main content of the chapter, so that they can interpret the subsequent picture with the known information. At the same time, it can also make up for the understanding deviation caused by incomplete or incorrect information acquisition caused by the excessively fast information transmission of the image.

B. Subtitle Discourse Analysis

Subtitles can allow viewers to accurately obtain the information that the producer wants to convey, and directly present the subtitles to the viewers with the simultaneous sound, so as to compensate for the information that may be overlooked by the image or sound.

Example I:

Many years from now, when people look back

at the battle against the novel coronavirus,

they won't forget a day in early 2020,

when a lockdown was imposed on Wuhan.

A lockdown that triggered off

a people's war against the epidemic.

And Wuhan is the eye of the storm in this war.

At the beginning of *The City of Heroes*, as in the above example, subtitles appear in the first 30 seconds. Through these subtitles, tell the viewer the time of the event occurrence "At the beginning of 2020", the location "Wuhan" and the event "the battle against the novel coronavirus". This subtitle is supplemented by dubbing, and the background is completely black. The viewer seems to enter Wuhan along with this, and return to February 2020 that unforgettable a people's war.

In summary, another embodiment of the text modal is subtitle. Its main function is to convey the information to the viewer most straightforwardly to make up for the information that may be overlooked in the picture. In the subtitles of the documentary *The City of Heroes*, a large number of specific digital descriptions are used, such as 346 medical teams, over 42 thousand medical workers, more than 4 thousand medical experts, 1 billion Chinese people. These numbers can

directly convey accurate information and most truly reflect the specific situation of the Wuhan epidemic. At the same time, it also reflects the determination and perseverance of the whole people to fight the epidemic and defeat the virus together in support of Wuhan from different power of the whole country.

III. IMAGE MODAL DISCOURSE ANALYSIS

Based on the Visual Grammar Theory of social semiotics, this chapter analyzes the discourse of image modalities from the meaning of representation, interaction and composition.

A. The Meaning of Representation

The meaning of representation corresponds to the conceptual function of systemic functional linguistics, which means that visual images can reflect the empirical events of the subjective and objective world outside of their own symbol system (Ding, 2020). According to the characteristics of visual images, Kress and Leeuwen (2006) divided the meaning of representation into narrative representation and conceptual representation (Kress & Leeuwen, 2006).

Narrative representation can be subdivided into action process, reaction process, and speech process (Ding, 2020). Fig. 2 is the action process in the narrative reproduction. The sender of the action is the takeaway deliveryman, and the receiving target is this electric car full of living supplies. It is a typical transitive behavior process, which recurrences the scene where the deliveryman is working at the risk of being infected by the virus when the closure is about to be closed. The deliveryman carefully prepared to start the electric car, and there were so many items on his electric car that there was nowhere to put his foot. It shows the panic of the citizens under the virus outbreak, and also shows the dedication of these unknown people who risk their lives to work.

Figure 2

Conceptual representation can be further divided into three types: classification process, analysis process and symbolic process (Pan, 2019). In addition to the action process (medical staffs swearing oaths), Fig. 3 also includes a classification process: five different teams volunteering in supporting Wuhan from various hospitals in Guangxi Province just represent five new forces. It can be seen that Guangxi Province medical staffs have confidence and determination in supporting Hubei Province; the symbolic process: the different color flags of each representative team symbolize the new power from different regions, and the red costumes represent the unified China, the one-hearted China. The gathering of different forces will definitely restore Wuhan to its former vitality and enable China to overcome this war without gunpowder.

Figure 3

In summary, the meaning of representation is the true recurrence of a certain scene, allowing viewers to feel the actions and expressions of the characters intuitively, and it is the enhancement of language. When the text and image describe a certain thing together, it can be maximized and to convey to the viewer the intention to express and the reproduction of the real scene.

B. The Meaning of Interaction

The description of video multimodal discourse mainly draws on the hierarchical classification of film and television works. According to the length of time, film and television works generally include 6 levels: frame, shot, scene, sequence, stage, and type (Zhang, Guan, 2013). *The City of Heroes* has a total duration of 45 minutes and 10 seconds, using the screenshot application PR to capture a total of 65,040 frames. Excluding duplicate screenshots, a total of 590

effective screenshots were obtained. Refer to the summary table of the realization of the interactive meaning of Kress and Van Leeuwen to get Table 2. The following will conduct an in-depth analysis of the documentary *The City of Heroes* from the four subsystems of the interactive meaning.

TABLE 2
THE INTERACTION MEANING STATISTICS IN *THE CITY OF HEROES*

Image behavior	Method to realize	Number of using times
Claim	Staring at the audience	60
Supplement	No gaze	128
Intimate/personal relationship	Close-up lens	72
Social relationship	Medium shot	138
Impersonality	Telephoto lens	75
Participation	Frontal perspective	69
Detachment	Oblique viewing angle	135
Viewer's rights	Overhead angle	63
Equal relationship	Flat angle	174
The power of the characters in the image	Upward angle	20
The emotions of the characters in the image	Close-up	26

1. Contact

Contact is an interactive relationship in which the film producer establishes an ideal contact interaction through the gaze "contact" between the characters in the image and the viewer. The way of "gazing at the audience" of the characters constitutes the claim image. On the contrary, it constitutes the provision of the image. The image is constructed through the shooting location--the reader relationship can adjust the symbolic power relationship, interactive relationship and intimacy between the reader and the image, so as to guide the reader to the "ideal interpretation of the images" (Pan, 2019, p.82).

Figure 4

The protagonist in Fig.4 is a doctor working in an intensive care unit. Outside the isolation ward, he communicated with his wife who came to visit by writing. Through the isolation door, the viewer can see the doctor's haggard eyes and the simple blessing to the four-month-old child in his hand. By "gazing at the audience", the doctor not only communicated with his wife, but also with the eyes of millions of viewers, as if there were countless family members worried about doctors and patients standing outside the door of the ward. In this way, an intimate relationship is established, which triggers empathy between the viewer and the image character.

It can be seen from Table 2 that there are 128 frames that is belong to supplement type. The image characters do not have direct eye contact with the viewer, but only provide information to the viewer. There is no relationship between the viewer and the viewer to establish a certain character emotion to achieve emotional resonance. A total of 60 frames are claim type, and a certain relationship is established through gaze, which not only provides information but also spreads the emotions of the characters to the audience outside the lens. At the same time, Fig.4 also belongs to a straight angle, placing the viewer and the image character on the same horizontal position, expressing the equal relationship between the two parties, and also expressing respect for the devotee.

2. Distance

Distance refers to the social and interpersonal relationship established through lens framing that is different from that of the audience. Different social distances will produce different viewing effects. Normally, the close shot is to narrow the distance between the viewer and the image participant and then form an intimate relationship. As shown in Fig. 4, all belong to the close shot; the middle shot is to reflect the social relationship, such as Fig.3, which reflects the new type of social relationship formed by the support team; but the long-term view is just the opposite of the close-up view. The main highlight is the social environment or the social relationship between non-individuals.

It can be seen from Table 2 that the middle shot used in *The City of Heroes* consist of 138 frames, which are more than close shots or long shots. The middle shot shows a group of doctors busy in the ward, several construction workers rushing to work overnight under night lights, or volunteers walking around in the community. The middle shot reflects the busy, selfless and fearless images. It is a kind of supplement images and does not establish a certain relationship

with the audience, but only conveys the real anti-epidemic scene. In the close shot, there are the doctor's face deformed by the goggles, and the feet of Dean Zhang Dingyu that are not flexible due to frostbite. The long shots usually explain the time, place and environment of events, such as the empty Yangtze River Bridge, the empty Hankou Railway Station, and thousands of workers who are constructing temporary hospitals.

The entire documentary uses the alternate shooting technique of long-range and short-range shots, allowing the audience's sight to continuously change scenes and viewing distances with the lens, as if from an empty Yangtze River bridge to a hospital full of disinfectant water, forming a strong visual contrast.

3. Attitude

Attitude refers to the establishment of an equal or unequal social relationship between image participants and viewers through the perspective of shooting. It is mainly divided into vertical viewing angle and horizontal viewing angle. The vertical viewing angle includes two angles: bottom-up and top-down angle. It can be seen from Table 2 that the documentary uses a total of 174 horizontal viewing, that is, the audience's line of sight is at the same level as the participants. The purpose of it is to establish an equal social relationship.

The purpose of bottom-up angle is to let the audience experience the tallness of the scenery or the heroic image of the characters. It can be seen from Table 2 that there are 20 vertical shots to praise the heroic image of the characters, to reflect the great and sacred images of the characters such as doctors, volunteers, and workers, and to embody the power of the characters in the image from a perspective of looking-up angle.

Figure 5

The top-down angle is mainly used to reflect the broadness of the environment, emphasizing the environment, space, and the position of the characters in it. It has a macroscopic expression and is often used as the opening and closing shots to attract the audience. As shown in Fig.5, taken from a long range top-down angle, it shows a magnificent picture of a large number of excavators simultaneously starting to operate. It was emphasized that in an emergency, a large number of workers did not fear the invasion of the virus, and were ordered in danger, in order to race against time, hurrying to work. It only took less than two weeks for the two hospitals to rise from the ground, and made great contributions and assistance to the creation of a good admission place for suspected patients. By top-down angle, all people and things become very small, just like weak humans fighting the virus alone. But when many hands make light work, we must have confidence in defeating the virus.

In summary, through the horizontal perspective, an equal or unequal social relationship between participants and viewers can be formed. At the same time, the communication between the two parties from the same perspective can also reflect the emotion of respect for the image participants. The vertical angle of view is mainly to reflect the broadness and magnificence of the environment or the building, forming a strong perspective impact, in order to infect the viewer.

4. Modality

Modality is an important manifestation of the meaning of interaction. According to Kress's (2006) visual grammar theory, modality can be classified into three categories: high modality, medium modality, and low modality (Kress, Leeuwen, 2006). The color saturation of an image reflects its modality value, which will affect the audience's acceptance and credibility of the information. As shown in Fig.1, the background is the calm and waveless Yangtze River, and the Yangtze River Bridge appears lifeless under the cloudy sky. The whole picture is a low-sensory modality, conveyed to the audience the cold scenes on the first day of Wuhan's lockdown, It also shows that this people's war has really started. Fig.3 is a high-sensory modality. The various red ones represent new hope and energy injection against the epidemic. It also indicates that the battle against the epidemic is not smooth, but a confrontation between life and death.

It can be seen from Table 2 that this documentary also uses 26 close-up shots, such as the doctor's face deformed by the goggles, the foggy goggles that have been worn for a long time, and so on. Through these close-up shots, a wave of Chinese power that has risen up in this ordeal is shown, and the emotions reach people's hearts. It also stimulates the audience's national pride, enhances cohesion and arouses strong resonance among the audience.

C. The Meaning of Composition

The meaning of composition corresponds to the meaning of text in systemic functional linguistics, which can be

analyzed from three aspects: information value, significance and framing (Feng, Xing, 2011). The information value is realized by the spatial layout of each constituent element in the figure, and its role in the whole is determined by its position. Significance refers to the degree to which the constituent elements attract the audience's attention. The framing reveals the subordination of the constituent elements, as shown in Fig.6.

Figure 6

Fig.6 is a scene taken in the intensive care unit. The protagonist of the picture is the critically ill patient with equipment all over his body. He is in the left half of the entire image and belongs to the background view. The surroundings of the old man are blurred, which is not only the shooting environment is too intense, but also the image of the old man's vision is blurred due to the disease. The right half of the image is a medical monitoring instrument. The value on the instrument is the information value. These heart rate, blood pressure that deviate greatly from the normal value just verify that the patient condition has not been well improved, which is belong to the prospect view. The significance of the instrument is obviously high. The audience's first attention is completely attracted by the value on the brightly lit instrument, which enables them to truly experience the seriousness of the illness and the strong infectivity of the new crown epidemic. Through analysis, it can be seen that the spatial layout of various elements, font size, light intensity, etc. are all very important information, and then it contributes to the construction of the overall textual meaning perfectly.

It can be seen that through the three factors of information value, significance and framing, images of different meanings can be constructed to convey different information to the viewer. In discourse analysis, it is necessary to combine the two, and at the same time link the textual mode and the image mode to perceive the meaning of its composition as a whole.

IV. VOICE MODAL DISCOURSE ANALYSIS

As an important sensory receiver of hearing, voice mode also plays an important role. The voice modality and image modality in *The City of Heroes* constitute a complementary relationship, not a strengthening relationship. Only when sound and images are combined can static text become dynamic pictures more effectively. The voice mode in this documentary includes two types: simultaneous voice and background music. In terms of dubbing style, it chooses a quiet female voice with a personal touch instead of the traditional male dubbing, and tells the incident in a soft tone. It's as if everyone is looking back at the past experience personally, to feel the truth, plain but full of meaning.

The City of Heroes uses a lot of simultaneous voice to create a sense of reality and atmosphere on the scene, so as to attract and infect the audience. Thus, the theme of "Hero City" is fully reflected. For example, when Dr. Zhong Ming said, "But after I came here, I found many of my ways to treat critically ill patients didn't work. I saw many of my patients dying." The audience can't help but feel the severity and terrifying degree of pneumonia. When the news broadcast from CCTV sounded, "Chinese President Xi Jinping, general secretary of the Communist Party of China Central Committee and chairman of the Central Military Commission, stressed putting safety and health as the top priority and making resolute efforts to curb the spread of the virus." You will feel the importance of the country, the strength of the motherland, and the commitment of the party, so that everyone has confidence in defeating the epidemic.

Good background music with images can highlight and strengthen the theme content and touch the hearts of the audience. As the scene changes in the documentary, the music melody and rhythm also change. For example, when the broadcast announced that Wuhan had been pressed the pause button, the Yangtze River Bridge in the video was still and there was no background sound, thus creating an atmosphere of tension and danger. When the two-month lockdown of Wuhan was finally announced, the ending music *Listen to me say thanks* sounded. The warm singing of the little girl washed the city of Wuhan for two months, and also awakened the gratitude and awe of every Chinese under the gloom to the medical and nursing groups, patients, volunteers and other civilian heroes.

The commentary did not clearly point out the "hero", but the simultaneous voices all pointed out the theme. Xinhua News Agency's first-line interview reporter Xu Yang said "Wuhan is a city with heroes everywhere"; Zhong Nanshan said: "Wuhan was really a city. A very heroic city"; Chinese President Xi Jinping pointed out that "Wuhan is a heroic city, and the people of Wuhan are a heroic people." Such an ordinary statement highlights its extraordinary greatness and also echoes the theme of the documentary-the city of heroes.

V. CONCLUSION

Through multi-modal analysis, it can be concluded that *The City of Heroes* has two meanings: Firstly, it is called "the City of Heroes" because people from all over the country support Wuhan gathering here; secondly, the city and people in Wuhan paid a huge price. They sacrificed, endured, helped each other and never give up, so it is also called the "hero" city.

The City of Heroes uses three different modalities to outline the whole picture of the epidemic war from the perspective of civilians. It has a broad perspective and arouses people's respect and gratitude for the heroes. Through analysis, it can be seen that the textual mode is the main mode of presentation of the theme, the image mode is the enhancement of the language, and the voice mode is the supplement to the image rather than the enhancement. The documentary is narrated in six chapters. There are more than 40 civilian heroes in the whole story. Through the use of a large number of squint angles, medium shots and horizontal viewing angle, on the one hand, it shows the pictures of these heroes who persisted in fighting in the face of difficulties when the epidemic broke out; on the other hand, it also shows that the Chinese power that has risen in the ordeal stimulates national pride, strengthens cohesion, and arouses strong resonance from the audience.

The documentary successfully conveyed its discourse meaning through a multi-modality, and conveyed everyone's great mission and determination to fight the epidemic. The multi-modal discourse analysis and interpretation of documentaries help us understand the way in which it realizes the meaning of the text.

REFERENCES

[1] Ding Jinshu. (2020). "1.4 Billion We Are China" multimodal discourse analysis of CCTV public service advertisements. *Popular Literature and Art*, *(23),*118-119.
[2] Feng Dezheng, Xing Chunyan. (2011). Spatial metaphor and the construction of multi-modal meaning: A case study of automobile advertising. *Foreign Language*, *34*(3): 56-61.
[3] Kress G & Leeuwen T. (1996). *Reading images*. Brown Prior Anderson.
[4] Kress G & Leeuwen T. (2006). *Reading images: The grammar of visual design*. Routledge, 41.
[5] Li Zhanzi. (2003). A socio-semiotic analysis of multimodal discourse. *Foreign Language Research*, *(5)*: 1-8.
[6] Pan Yanyan. (2019). Analysis of national security discourse from a multimodal perspective: Taking the comparative analysis of Chinese and American police image propaganda films as an example. *Foreign Language,(01)*,78-87.
[7] Zhang Delu. (2009). A comprehensive framework for multimodal discourse analysis. *Chinese Foreign Language Journal, (01)*,24-30.
[8] Zhang Daotong. (2020). The cultural mission of Chinese documentary and its international dissemination. *Art Review, (09)*,8-15.
[9] Zhang Jing, Guan Ling. (2013). *Film and television audiovisual language*. Communication University of China Press.

Min Zhang was born in Chongqing, China in 1997. She will receive her Master degree in English Teaching from Chongqing Normal University in 2023.

She is a second year graduate student now. And she research interests include education, linguistics and psycholinguistics.