



A Multimodal Discourse Analysis of *Planet Earth II* from the Perspective of Visual Grammar

Wenyi Cui, Lin Zheng*

China University of Petroleum, Shandong, China.

How to cite this paper: Wenyi Cui, Lin Zheng. (2023) A Multimodal Discourse Analysis of *Planet Earth II* from the Perspective of Visual Grammar. *Journal of Humanities, Arts and Social Science*, 7(12), 2455-2459.
DOI: 10.26855/jhass.2023.12.010

Received: November 18, 2023

Accepted: December 15, 2023

Published: January 9, 2024

*Corresponding author: Lin Zheng, China University of Petroleum, Shandong, China.

Abstract

As society develops by leaps and bounds, the expression of information and the transmission of discourse is no longer dependent on a single language symbol, but on the interaction and presentation of multiple symbols. Therefore, traditional research on a single text can no longer comprehensively and deeply analyze discourse. Multimodal discourse analysis came into being and gradually became a hot topic in the field of linguistics in recent years. Based on visual grammar, this paper takes the BBC natural documentary *Planet Earth II* as the research corpus. It selects a large number of images from it to analyze the uniqueness of the documentary of the representative meaning, interactive meaning, and compositional meaning. This paper aims to discuss how the documentary uses modal resources such as text, sound, picture, and layout to jointly construct the overall meaning of the discourse, hoping to bring beneficial creative inspiration and innovative ideas to the producers of the documentary industry and summarize some valuable experience for the development of the domestic natural documentary market.

Keywords

Multimodal, Visual Grammar, *Planet Earth II*, Image

1. Introduction

With the development of mass media and digital information dissemination, multimodal discourse has become one of the most expressive expressions after written discourse. Multimodal discourse refers to the use of hearing, vision, touch, and other senses to communicate through language, images, sounds, actions, and other means and symbolic resources (Zhang Delu, 2009). In recent years, more and more attention has been paid to multimodal discourse analysis in the domestic and overseas, but its application research is rare in nature documentaries. Nature documentary is a kind of television art form that integrates multiple modes, the narrative way of film storytelling and the true reproduction of documentary. It uses reality to arouse people's thinking, which helps the audience intuitively understand the magical scene of nature. BBC nature documentaries often use the combination of ultra-high-speed photography and time reduction to present natural time and space in the form of dramatic time and space, which is good material for corpus research. The return of *Planet Earth II* after ten years has aroused a lot of attention, calling on people to focus on the impact and changes of human behavior on the whole ecological environment. From the perspective of visual grammar, this paper takes the BBC Natural documentary *Planet Earth II* as the research object to systematically explore how multimodal resources work together to construct discourse, to convey its central idea change to the audience.

2. Visual Grammar

Multimodal discourse analysis rose in the 1990s. It adopted the social semiotics and meta-functionalism of the

systemic functional theory, and its research belonged to the innovative field of linguistics. Based on Halliday's language metafunction theory, Kress and Van Leeuwen expanded their concepts with the reference points of conceptual function, interpersonal function, and textual function. They regarded image elements as a social symbol and created a multimodal discourse visual grammar system for analyzing images. Just as the grammatical description words of the language are combined into sentences and discourses, visual grammar will describe the characters, places, events, and other elements described to form a visual narrative of different complexity (Kress G & Van Leeuwen T, 2001). The visual grammar system is a dynamic organic whole with three cores, namely, representative meaning, interactive meaning, and compositional meaning. This analysis method has important guiding significance for further understanding the social semiotic characteristics of language and the meaning of multimodal discourse.

At the beginning of the study of discourse analysis, people did not study it as a mode of meaning expression. Since the multimedia discourse emerged, people studied it as an independent academic field. In the early 20th century, Bransford and Johnson emphasized the importance of visual signs in discourse comprehension (Dai Shulan, 2013). R. Barthes explored the interplay between images and language in expressive meaning in his 1977 paper. Royce emphasized the complementarity of different symbols in multimodal discourse, it also discussed the synergy between symbols in the context of second-language teaching. Although the research on multimodal discourse in China is still in its early stages, the field and scope of study have been progressively expanding. For example, Zhang Delu established a comprehensive framework of multimodal discourse analysis based on systemic functional linguistics and explored the relationship among various modes (Zhang Delu, 2009). Feng Dezheng combined metaphor theory with multimodal discourse analysis. In his paper, the metafunctional resources of visual images are regarded as metaphorical potential, and the construction of metaphors from three perspectives is analyzed (Feng Dezheng, 2011). At present, most of the research focuses on images, advertisements, emblems, promo, and other forms of hypertext from the perspective of visual grammar, while documentary research is less. Therefore, this thesis focuses on BBC *Planet Earth II* and explores the discourse construction and meaning of nature documentaries from the perspective of visual grammar.

3. A Multimodal Discourse Analysis of *Planet Earth II*

Planet Earth II was released on November 6th, 2016. It has six episodes, including islands, mountains, jungles, deserts, grasslands, and cities. Each episode is about 58 minutes long and is interpreted by Sir David Attenborough. It is the first time that the BBC has used the innovative ultra-high definition 4K camera to shoot the whole process, using UAV and remote control shooting technology. The 42 photographers in the team have traveled 117 times and across 40 countries, with a production cycle of four years. This season focuses more on the connection between humans and nature and reconstructs human understanding of nature. It explains the impact of human behavior on the living environment of wild animals and plants, and how animals and plants respond to this change. Meanwhile, it calls on human beings to pay attention to the change in today's natural ecological environment. The author selects some classic clips and pictures in the documentary and makes a specific analysis according to the visual grammar.

3.1 Representative Meaning

The representative meaning of images consists of narrative representation and conceptual representation. Narrative representation mainly includes action, environment, reaction, and so on. The people, things, and objects in the image can generate meaning through structure. Vector is the symbol of narrative representation. The narrative vector is the intermediary of the relationship between constituent elements. Conceptual representation expresses the participating objects of category, structure, or meaning, which is not constrained by time. There is no vector in the conceptual representation. The narrative representation in *Planet Earth II* is mostly embodied in the contact between animals. For example, in the picture of five lions besieging a buffalo at 9:56 in the fifth episode, the lion is the originator of the action. The contact between the two species constitutes a vector, and the splash of water during the fight is also an element. Combined with the picture interpretation and background music, the audience's visual and auditory senses interact with each other through these elements together and naturally enter the narrative context. But the reason why the lions do this in the picture is for their children. If they cannot catch the prey as soon as possible, the young lion may die. This is also a feature of the documentary. The positive and negative roles in the traditional nature documentary are reversed, so that the viewer can see the difficulty and vulnerability of predators when sympathizing with prey, breaking people's prejudice against some animals, and guiding people to pay attention to the warm side

of cold-blooded animals.

3.2 Interactive Meaning

The interactive meaning in visual grammar mainly studies the interactive behavior among image producers, viewers, and various modes in the image, which is mainly manifested in four dimensions: contact, social distance, attitude, and modality.

3.2.1 Contact

Contact is a kind of communication between the participants and the audience in the image, which refers to a kind of meaningful relationship between them indirectly constructed through eye contact. When the participants' eyes point to the viewer, the two will have a contact relationship. Visually, participants ask for image effects from the viewer. Kress and Van Leeuwen defined this composition as asking for the image (Kress & Van Leeuwen, 1996). If participants and viewers have no eye contact, and the image only shows some information, this composition is defined as providing the image. For example, in the opening shot of the first episode, there are no animals, only mountains, jungles, deserts, and other scenery. Although there is no change in the action in the picture, the shot has been moving forward slowly. This kind of image has no moving vector, which is to provide class images, directly reflecting the theme of the sixth episode of this season. But from 39 seconds, the movement of the camera is no longer the ecological environment scene at the beginning, but a dynamic picture of various animals, such as penguins rolling in the waves in 55 seconds, two young cheetahs fighting and biting in 58 seconds, and the migration of flocks of animals in 1 minute and 8 seconds. At this time, the providing image at the beginning is changed to the requested image. Although there is no direct eye contact, the picture is asking for interaction from the audience. When carrying out this symbolic interaction, people should not only pay attention to the visual factors but also focus on the auditory factors, such as the sound of the waves lapping on the rocks and the cheetah biting. These elements together construct the information transmission of the text to the audience convey the idea of loving and respecting nature to the audience, and let the audience feel the vitality and vigor of nature and animals.

3.2.2 Social Distance

According to the theory of visual grammar, the distance of the image reflects the distance of interpersonal relationships. The framework can be divided into four types in line with its size: intimate distance, personal distance, social distance, and public distance. Intimate distance refers to the close-up of the head and shoulders of the character. Personal distance refers to the lens above the waist. Social distance is reflected in the whole body of the characters. Public distance refers to a character's whole body and background environment (Liu Dan, 2020). The second episode, from 20 minutes 20 seconds to 20 minutes 54 seconds, films a brown bear following the smell to find a tree that he likes and rubbing against the trunk, which is the way of hair removal of brown bears after the cold winter. In these 30 seconds, the shots between public distance and intimate distance constantly switch, and the process of brown bear hair removal photographs from a multi-dimensional perspective. Several of them are overhead shots, which directly shoot the head of the brown bear. The tight shot makes it easier to arouse the audience's emotions. Moreover, the eye features of various animals often appear in *Planet Earth II*, allowing the audience to intuitively touch and feel the animals. Brown bears are hard for humans to handle, therefore, such a shooting technique shortens the social distance between bears and audiences, so that the audience can understand the living habits of brown bears. In 20 minutes and 34 seconds, lively background music suddenly rings out in the picture, and the brown bear holds the branch with two hands and twists with the music. This presentation mode shortens the psychological distance by combining the multimodal resources of the picture and sound, making it easier for the audience to produce empathy, and changing people's terrible psychological cognition of the brown bear in a relaxed and humorous atmosphere.

3.2.3 Attitude

The reference standard of attitude is perspective. Kress believes that the shooting perspective can show the subjective and objective attitudes of the participants in the picture, and reflect the relationship between the representative participants and the interactive participants (Kress G & Van Leeuwen T, 2001). Images can be divided into horizontal perspective and vertical perspective. The horizontal perspective integrates the audience into the picture situation, while the vertical perspective represents the status relationship between the audience and the picture participants.

In this documentary, the horizontal angle of view and vertical angle of view are constantly switched and perfectly combined. For example, at 10 minutes and 12 seconds in the first episode, the picture of two giant lizards wrestling

together is taken from a horizontal perspective. In the dragon society, size determines everything. The calm of the female lizard is in sharp contrast to the mania of the male lizard. The eye-level shooting can make the audience form empathy, emphasizing that the audience is in an equal relationship with the giant lizard in the image.

In the sixth episode, the male langur jumps at 1 minute and 17 seconds, where the camera angle is a vertically low angle, and then in 2 minutes and 41 seconds, the camera angle changes to a vertically high angle, so that the audience changes from the weak side to the strong side. At this time, the background music also changes from the previous gentle rhythm to tension. The music style is highly consistent with the scene of groups of langurs running in the picture, creating a dramatic effect, making the audience feel the aggression of langurs, and bringing the audience an excellent sense of audio-visual pleasure.

3.2.4 Modality

Modality refers to the channels and media of communication, including language, technology, image, color, music, and other symbol systems (Zhu Yongsheng, 2007). The composition, proportion, and color of the picture in *Planet Earth II* are refined in every detail. For example, at 18 minutes and 13 seconds in the second episode, the image processing here is very ingenious. It only takes one second to show the process of the Rocky Mountains' transition from winter to spring. The picture is very natural and smooth from white snow to green, from static to dynamic, from shallow to deep, from dark to bright. The color contrast diagram is prominent, and the color discrimination and saturation are very high, showing the dramatic seasonal changes in the Rocky Mountains. Simultaneously, the sound of background music is in line with the emotion that the documentary itself wants to express. The audience stays between the mountains and experiences the changes of the four seasons, feeling the beauty of nature and reaching spiritual resonance through the gurgling water and the chirping birds in the picture.

3.3 Compositional Meaning

The compositional meaning reflects the integrity of multimodal discourse construction, mainly through three dimensions: information value, saliency, and framework. The information value is reflected in the position and order of the composition picture. Saliency refers to the degree of attraction of image elements to the audience. Viewfinder refers to connecting or cutting elements in an image by separating lines or frameworks (Zhang Jingyuan & Jia Peipei, 2012). At the beginning of each episode of the documentary, the overall background of the picture is the surface of the earth. With the sound of music slowly ringing, the proportion of the earth in the picture gradually shrinks. A sun slowly appears in the lower left corner of the picture, and then the sun gradually moves to the center. Then the documentary title Planet appears in the middle of the picture. The sun continues to move right behind the letters, lighting up letters. At 21 seconds, the halo of the sun suddenly becomes larger and flickers and a small halo reflected by the sun appears on the left side of the picture. Then the sun begins to shrink and slowly merges with the earth. The two disappear into a light. Finally, only Planet Earth and the light on the right side remain in the picture.

At the beginning of the picture, the earth and the sun are not in the horizontal left and right directions, but in an inclined angle, which visually gives the impression that they are in the lower left and upper right respectively. Normally, the volume of the sun is 1.3 million times that of the Earth, but the picture is just the opposite. The Earth accounts for most of the picture, indicating the theme of the documentary "Earth". The central position of the title fully attracts the attention of the audience. The text is placed in the foreground, covering the sun but not the earth, reflecting the primacy of the information. The fuzziness and clarity of the text meet people's visual focus representation, reminding viewers to pay attention to the generation of important information. In the later stage, the shape change of the solar halo from large to small increases the instant sense of the audience's vision, and the brightness of the color gives people the expectation of brightness and hope. The final light in the picture is not a single color, but a gradual change from yellow in the middle to blue on both sides, from a warm tone to a cold tone. The color transition is very smooth, which improves the overall softness of the picture.

4. Conclusion

Planet Earth II is a very successful nature documentary. Its beautiful pictures and excellent production are examples of similar documentaries. Based on visual grammar, this paper analyzes the visual and auditory scenes in the promotional film in detail from representative meaning, interactive meaning, and compositional meaning, and discusses how the image, color, sound, text, and other symbol resources interact with each other, so as to express the meaning of the documentary and increase the audience's sense of identity. Through the analysis of the corpus, it is

found that the storyline and narration of nature documentaries can arouse the empathy of the audience, help the audience reflect on the impact of their behavior on the ecology, and improve people's awareness of ecological protection.

References

- Dai Shulan. The Origin and Progress of Multimodal Discourse Research [J]. *Journal of Foreign Languages*, 2013(2): 17-23.
- Feng Dezheng. Construction and Classification of Multimodal Metaphors from the Perspective of Systemic Functional [J]. *Foreign Language Research*, 2011(1): 24-29.
- Kress G & Van Leeuwen T. *Multimodal Discourse: The Modes and Media of Contemporary Communication* [M]. London: Edward Arnold, 2001.
- Kress. G. & T. Van Leeuwen. *Reading Images: The Grammar of Visual Design* [M]. London: Routledge, 1996.
- Liu Dan. Multimodal Metaphor Construction of Vertical Screen Micro Film Advertising from The Perspective of Visual Grammar—A Case Study of Huawei Advertising Wukong [J]. *Journal of Huaqiao University*, 2020(1): 154-160.
- Zhang Delu. Exploration on the Comprehensive Theoretical Framework of Multimodal Discourse Analysis [J]. *China Foreign Languages*, 2009(1): 24-30.
- Zhang Jingyuan & Jia Peipei. Reflections on Visual Grammar [J]. *Contemporary Foreign Language Research*, 2012(3): 38-42.
- Zhu Yongsheng. Theoretical Basis and Research Methods of Multimodal Discourse Analysis [J]. *Journal of Foreign Languages*, 2007(5): 82-86.