

Multi-modal Discourse Analysis of the Movie the Tea House

Jianbin Zhu*, Linling Huang^a

College of Foreign Studies, Guilin University of Technology, Guilin, China

*2017135@glut.edu.cn, ^a573626834@qq.com

Abstract

In the 90s of the 20th century, the multi-modal discourse analysis that emerged in the West gradually became one of the hot topics in linguistics, and multi-modal discourse refers to the phenomenon of using hearing, vision, touch and other senses to communicate through language, images, sounds, actions and other means and symbolic resources. Usually people discuss film from psychology, aesthetics and other macroscopic perspectives, and less linguistic discourse analysis methods to discuss, multi-modal discourse analysis theory for film researchers to provide a set of simultaneous analysis of film images, music and text compatible methods, so this thesis aims to analyze the film discourse Tea House based on functional linguistic theory and visual grammar for multi-modal discourse analysis.

Keywords

Multi-modal Discourse Analysis; Functional Grammar; Visual Grammar.

1. Introduction

In this part, the writer expounds the background and purpose of the research, as well as the theoretical and practical significance of the research, and puts forward the research methods and basic framework.

1.1. Research Background

Since the early 50s of the 20th century, American and Chinese speech scientist Harris proposed the research direction of discourse analysis, linguists around the world have proposed a variety of discourse analysis theories and research methods, through the analysis of a large number of examples, to analyze the internal laws of discourse activity and the relationship between discourse activity and ideology and cognitive mode [1]. However, the current research on discourse analysis is limited to the influence of language activities on discourse and communication, and ignores the auxiliary role of other factors, such as images, sounds, colors, movements, and so on. However, a large part of discursive communication, whether written or oral, is manifested through non-verbal factors. The mode of communication that works together by this combination of non-verbal and linguistic factors is multi-modal discourse communication.

Multi-modal discourse refers to the phenomenon of using hearing, vision, touch and other senses to communicate through language, images, sounds, actions and other means and symbolic resources. People have not paid attention to multi-modal discourse, and in modern linguistic research, some scholars have studied it from the perspective of non-verbal characteristics and accompanying language characteristics, and with the emergence of multimedia, in recent years, many scholars have gradually realized its importance and began to study multi-modal discourse analysis[2]. Since film discourse contains various symbolic resources, such as visual images, sounds, language, etc., multi-modal discourse analysis provides a compatible term for film discourse to analyze film discourse. In view of this, film discourse belongs to the category of multimedia discourse, and we can conduct multi-modal

discourse analysis on the film *Tea House* to study the feasibility and operability of multi-modal discourse analysis in film discourse analysis.

1.2. An Introduction of Tea House

Tea House is a drama created by modern writer Lao She in 1956, the movie *Tea House* was released in 1982, the film is adapted from the drama of the same name written by Lao She, is a feature film produced by Beijing Film Studio, directed by Xie Tian, and starred by Yu Shizhi, Zheng Rong, Lan Tianye, Huang Zongluo, Ying Ruocheng and so on. Set against the backdrop of the rise and fall of the Yutai Tea House in old Beijing, it tells the dark and corrupt story of Chinese society for nearly half a century from the early 20th century to the eve of liberation, as well as the inescapable painful struggle and tragic story of ordinary Chinese people in order to survive, seek justice and seek development. It reveals the social contradictions of different eras, expresses several representative Chinese, and expresses the limitations of the existing era and the theme of transcendent value. The film won the 3rd Chinese Film Golden Rooster Award Special Award in 1983 and the Special Award of the Ministry of Culture in 1982 Excellent Film Award.

1.3. Purposes and Significance of the Study

The rise of multi-modal discourse analysis in the 90s of the 20th century, which proved to be an effective non-verbal modal analysis method in the West. But the study of multi-modal discourse analysis in China still lags far behind the Western world, and there was no analysis before 2003. In order to meet this need, this paper hopes to further enlighten the research of multi-modal discourse analysis. The purpose of this study is to expand the application field of multi-modal discourse analysis, theoretically verify the feasibility of applying systematic functional linguistic theory to multi-modal film discourse, and deepen people's understanding of multi-modal discourse analysis and improve people's appreciation of the movie *Tea house*. It also provides a new perspective for the appreciation of film works, thereby improving the audience's film appreciation level.

1.4. Basic Framework of the Paper

Chapter 1 introduces the origin of the selected topic, the background of the study, the theoretical basis of the study, the purpose and significance of the study, and the structure of the thesis.

Chapter 2 provides a review of film discourse and multi-modal discourse analysis. The overview of film discourse includes the semiotic approach to the study of film language and the framework of film discourse analysis. The current status and development of domestic and international research on multi-modal discourse analysis.

Chapter 3 presents a theoretical framework derived from functional grammar, i.e., visual grammar. The functional grammar used in the analysis of language contains three meta-functions: conceptual, interpersonal and discourse functions. In visual grammar, which is suitable for analyzing other modalities such as images, these three meta-functions are renamed representational meaning, interactive meaning, and compositional meaning.

Taking the film discourse *Tea House* as an example, Chapter 4 illustrates how representational meaning, interactive meaning and compositional meaning are achieved through a detailed analysis of various choices of behavior, social distance, attitude, modality system, information value, salience and framing.

Chapter 5 classifies the different modalities and illustrates the relationships between them. In addition, a comprehensive analysis of the selected modalities is presented with the aim of exploring how these modalities contribute to the overall meaning conveyed by the film.

Chapter 6 presents conclusions, limitations, and recommendations for further research.

2. Literature Review

This chapter mainly introduces the origin and research status of multi-modal discourse analysis and view.

2.1. Studies on Multi-modal Discourse Analysis

2.1.1. Studies on Multi-modal Discourse Analysis Aboard

R. Barthes, one of the earliest researchers in multi-modal discourse analysis, explored the interplay between images and language in expressive sense in his 1977 paper *Rhetoric of the language*. O'Toole (1994) analyzes images using multi-modal discourse theory; Kress & Van Leeuwen (1996; 2001) studied the relationship between modality and media, and specifically analyzed the phenomenon of how multi-modal discourse expresses meaning, including visual images, color grammar, and the layout design of newspapers and the role of different media. O'Halloran (2004) analyzes the film discourse.

2.1.2. Studies on Multi-modal Discourse Analysis at Home

Domestic scholar Li Zhanzi (2003) introduced the theory of multi-modal discourse analysis for the first time. Hu Zhuanglin (2007), Zhu Yongsheng (2007), and Zhang Delu (2009) have studied the theory of multi-modal discourse analysis. Hu Zhuanglin (2006), Ye Qichang (2006), Zeng Lei (2006) carried out multi-modal research in the fields of PowerPoint presentation text, hypertext, international conferences, speeches, etc., and Li Miaoqing (2007) applied multi-modal discourse analysis theory to film discourse analysis.

2.2. Study on Filmic Discourse

2.2.1. Researches on Film Language

A clear definition of film discourse is essential and should precede an overview of its development. It is well known that texts containing written and spoken texts are representations of discourse in the traditional sense. And the analysis of discourse is necessarily an analysis of the language in use. [3] However, with the development of technology, the scope of discourse extends to other non-verbal forms that also contribute to the production of meaning. In the context of cinema, discourse refers to what we call the "language" of film-making in action, the way in which the film acts on the viewer. Simply put, film discourse is a way of referring to this action, to the ideology of cinema. It refers to the way in which the story positions the audience.

2.2.2. Framework Adopted for Filmic Discourse Analysis

This section presents two models used in the analysis of film discourse, especially visual analysis, whose core theory is based on functional grammar. The first model was proposed by O'Toole, who named the three functions as reproductive, affective and constitutive, corresponding to Halliday's conceptual, interpersonal and textual functions, respectively. By modifying Halliday's linguistic hierarchy, he divided the hierarchy of visual images into works, episodes, characters, and members. Based on Halliday's theory of systemic function and O'Toole's theory of visual analysis, L O'Halloran moves forward in the study of the meaning of dynamic visual texts in film. His mode of analysis of cinematic discourse also focuses on visual form.

3. Theoretical Framework

Chapter three mainly provides the theoretical framework of this thesis. Then, it briefly introduces Halliday's systemic functional grammar, and visual grammar.

3.1. Functional Grammar

The systemic function grammar was originally constructed by Halliday, who argued that modal semiotic resources are determined by how people use them to produce meaning-the social functions they are given. In the systemic function grammar, there are three main meanings expressed in language, which is often called three meta-functions. These three generalized functions include conceptual functions, interpersonal functions, and discourse functions. Each of these three meta-functions is associated with a specific linguistic subsystem.

Conceptual meta-function: In language, it is mainly reflected in the selection of physical systems, and images and languages can reflect conceptual functions and present human experience. The interaction of conceptual meaning between symbols involves identifying visual message elements, i.e. participants, processes, environments, and attributes in images. These components have semantic characteristics reflected in various visual technologies.

Interpersonal function involves the relationship between the producer and the reader, and the verbal function of an image often requires textual support to determine. In written discourse, interpersonal functions are mainly reflected in tone and modality, but in multi-modal discourse, we need to observe the interpersonal complementarity and tone between text components and other symbols, and observe the way propositions are presented to the reader. The theory of modal systems of language also applies to other semiotic systems.

The discourse feature deals with how speakers construct their messages from a distance, allowing them to smoothly integrate into linguistic events. In addition to interacting with the audience and saying something to them about the world, the speaker constantly signals to them how the present part of their message fits in with the others. [4] Simply put, the discourse function is the function of language, which indicates how the author organizes the discourse. It is mainly manifested in the structure of topics and information structures.

3.2. Visual Grammar

With the rapid development of technology and society, discourse has shown diversity in form. How to effectively integrate modalities and media in order to interpret the discursive meaning of texts more comprehensively has always been an important aspect of multi-modal research. There are certain differences in research methods at home and abroad, mainly reflected in multi-modal analysis Chinese the function of words and images: Barthes (1968, 1977) believes that the interpretation of images must be combined with the interpretation of words, while Kress & Van Leeuwen (2006) believes that images are related to words, but do not depend on words, and the image itself has both organization and structure. Through the analysis of a large number of static texts, Kress & Van Leeuwen (2006) proposed that the analysis of multi-modal discourse with visual grammar can be divided into representational meaning, interactive meaning, and compositional meaning.

3.2.1. Representational Meaning

Representational of meaning is the connection between different modalities through expressing modalities such as the character's reaction, action, language and other as well as their external, which explains the relationship between the participants of meaning and the process, divided into two kinds of narrative and conceptual structure: Narrative includes action process, reaction process, speech process, and mental process; where action process and reaction process can be further divided into interactive and non-interactive modes. The elements in the diagram form a vector by forming a diagonal line. Vectors exist in narrative images, while conceptual images do not have vectors. Narrative images show developing actions and events, processes of change, spatial arrangements of moments; while conceptual images represent more stable meanings, such as in terms of categories, structures or meanings, which include categorizing processes, analysis processes, and symbolic processes.

3.2.2. Interactive Meaning

Interactive meaning refers to the fact that each action constitutes an interactive function, which expresses not only the orientation of the language user to the receiver, but also the social relationship between the two. The factors that influence the interactive meaning are contact, social distance, attitudinal relationship, social power and modality. Contact means that the participant in the picture makes eye contact with the viewer and has an emotional appeal to the viewer, forming a demanding. Conversely, information is giving. Social distance can be indicated by the size of the represented component, that is, the distance between the represented component and the reader, e.g., a close-up can suggest a close relationship, and a long shot suggests an impersonal relationship. Attitudinal relationship is judged by the horizontal position of the observer in relation to the constituent, which is classified as frontal, lateral, marginal, etc. The power relationship is reflected by the vertical relationship between the observer and the object: if the object is more powerful than the viewer, the viewer needs to look up to it, the object gives a feeling of subordination and oppression; otherwise, the viewer and the object are equal in status, can be level, the viewer's status is higher than the object perspective can be overlooking. Mood refers to a certain means of pictorial expression such as color, reproduction of details, depth, tone, etc., the degree of use divided into three categories: high, medium and low.

3.2.3. Compositional Meaning

The compositional meaning breaks the original boundary between text and image, linking the representational and interactive meanings through information valuation, framing and salience, i.e. the influence of the compositional features of things inside the frame on the understanding of the image. Information value is achieved by the position of the person or thing placed in the image. According to Kress & Van Leeuwen (2006), factors placed above express ideal, with idealized or generalized substance, while factors below express real, with specific information, such as details, etc. The information in the center is important. The information placed on the left is known information, while the information on the right is new information. Salience is achieved by the thickness of the lines and the intensity of the colors in the image. Thick lines and strong colors have high salience, and vice versa, low salience. Framing is achieved by many visual framing strategies, such as corner framing visual strategy, upper left corner framing strategy, lower right corner framing strategy, etc. In conclusion, changes in information value, salience, and framing components cause changes in the overall layout of the discourse, thus affecting the realization of the meaning of the composition and the meaning of the whole discourse.

This paper presents a multi-modal discourse analysis of the classic scene of the movie Tea House from these three dimensions of meaning.

4. A Case Study of Tea House

The purpose of this chapter is to conduct an actual multi-modal discourse analysis of the movie Tea House and to make a specific scenario analysis of the realization of the three meanings.

4.1. Scene 1: Chang Siye Clash with Er Dezi

The beginning of Chang Siye discussing that the mangler does not fight the foreigners but is here to be powerful, the next shot to Er Dezi is walking inside, Er Dezi turned back and glared at Chang Siye. Here pinched off the ambient sound, suddenly quiet for a few seconds, it is not because the people in the tea house feel that something is going to happen to stop talking, because at this time only we the audience to notice this person, pinched off the sound of conversation makes the atmosphere all of a sudden tense up, as shown.



Figure 1. Er Dezi Glares at Chang Siye **Figure 2.** Er Dezi Fights with Chang Siye

Analyzed from the level of representational meaning, both in Figure 1 and Figure 2 can be categorized as action processes. In the right picture there are two participants, Er Dezi is the initiator of the action, and Chang Siye is the target, and the target is the person to whom the vector refers, who is the target in terms of frame size, color saturation, and psychological expression of facial expression. In the picture on the right, the two frames are each half of the frame, and Chang Siye's expression is easily recognizable and full of courage to dare to fight against the evil forces. When two participants are connected by vectors, they are called interactors. There are two vectors in the picture, one is the diagonal line formed by the hands pulled by the two when they fight, and the other is the diagonal line formed by the eyes of the two when they look at each other.

Analyzed from the level of interactive meaning, the represented person in the right picture has eye and body contact with the participant, connecting the two in a direct way, which is a demanding picture. Chang Siye's hand is pressed against Er Dezi's hand, slightly prevailing over him, as if he is telling the audience to never bow down to the black forces and fight to the end.

Analyzed from the level of meaning of the composition, the left picture Er Dezi is located in the left side of the composition indicates new information, just after the Chang Siye discusses the mangler does not fight foreigners but everywhere, Er Dezi as new information appears on the left side of the screen, telling the audience that Er Dezi is the mangler that Chang Siye just told. In the right picture, Er Dezi is dissatisfied with Chang Siye, rushed to fight, the two have been in the middle of the screen, belongs to the important information.

4.2. Scene 2: The Qing Dynasty Collapsed and the People Run Away

The palace was surrounded by the revolutionary party, and soon it was fired. The people in it took the baggage and fled for their lives in all directions, in a panic, even if the gold and silver jewelry fell, they couldn't care.

Analyzed from the level of representational meaning, the shot of Figure 3 shows the viewers the turmoil of the time and witnessed the downfall of the Qing Dynasty through the panic of sound modality and the contrasting color modality of light and dark. In Figure 3, people run out of the room with each other's help, which is the process of interactive action. The scene is a narrative representation, and the characters run into the scene from far to near, their anxious expressions revealing the tension of the situation. The falling of the jewelry box in Figure 3 is a non-interactive action process.

Analyzed at the level of interactive meaning, the participants in Figure 3 have eye contact with the audience, creating emotional communication with the audience and making them curious and worried about where they are about to flee, belonging to picture of demanding. The shot in Figure 3 shows the audience a flat frontal view, implying that the participants and the audience are equal. The scattered jewelry boxes in Figure 4 represent the participant as an object that is

only displayed or offered in the viewer's eyes, with no relationship to the viewer. The shot is a close-up, making the object stand out.

Analyzed at the level of compositional meaning, the house on the left side of the background in Figure 3 is known information, and the flames in the lower right corner are new information. The salience of this image lies in the growing fire, which contrasts with the dimly lit wall and gradually takes over the frame. Figure 4 is taken in close-up with the brightest light, and the jewelry box is placed in the middle of the picture, indicating important information and emphasizing the crowd's panic.



Figure 3. Running Away From the Palace **Figure 4.** Scattered Jewelry

4.3. Scene 3: Sacrifice

Wang Lifa, Qin Erye and Chang Siye give their own tribute, for fear that in the future there will be no one to burn joss paper for themselves! joss paper flying down in the black space, in the slow-motion push slightly mournful beauty, metaphorically in the chaotic world people are nothing more than like floating weeds without support, like mole crickets as humble and insignificant.



Figure 5. Sacrificing Himself

Figure 6. Flying Joss Paper

Analyzed from the level of representational meaning, In Figure 5 participant throws the joss paper into the air, the joss paper in the air is the target, and a vector is formed between the participant and the target, so it is a narrative representation. Figure 6 is a non-interactive action process because the action is initiated by the object, and the action itself is not directed at a target.

Analyzed at the level of interactive meaning, the participants in Figures 5 and 6 do not communicate with the viewers in any way, so they belong to the same category of information-providing pictures, merely showing them to the viewers and provoking them to think about their meaning of doing so. Figure 5 is the only overhead shot of the whole film: several people have to throw joss paper around the circle to pay tribute to themselves, and the audience experiences a sense of depression. Figure 5 is a close-up of the joss paper in the air, which flies down in the black space, slightly mournful and beautiful under the slow-motion pushing, metaphorically showing that in the chaotic world people have no support and are as humble and small as ants. Figure 5 and Figure 6 are both information-providing pictures.

Analyzed at the level of compositional meaning, Figure 5 participants are located at the bottom of the picture to indicate reality, and the joss paper are located at the top is the ideal message to express the wish that someone will burn joss paper for them. Figure 6 is a close-up shot and the color of the paper money is highlighted, echoing the black space and using prominence to emphasize their last wish. The atmosphere of sadness is set.

5. A Synthetic Analysis

In this part, the data collected will be analyzed to provide answers to the questions put forward in the Methodology chapter. Firstly, the distribution and language features of attitudinal resources in Planet Earth II are described respectively. Secondly, it discusses how the Planet Earth II conveys the ecological thoughts through language.

5.1. Relationship among Different Modalities in the Film

Film discourse is a combination of different modalities, which includes not only language or verbal text, but also visual images, sound, music and so on. Generally speaking, it can be divided into two categories, namely visual modality and auditory modality. The main modality used in movies is the visual modality, which includes images and text. The main elements in images are brightness, color and movement, while the main elements in text are size, color and text type. The accompanying modalities throughout the film are the film's sound and a small amount of music. Since film is a dynamic multi-modal discourse, it not only involves multiple sign systems, but also presents the dynamics of modality to the audience visually and aurally. When only one modality fails to express the communicator's meaning, other modalities play the role of reinforcement, complement and coordination. The relationship between different modalities can be divided into complementary and non-complementary relationships. Complementary relationships are those in which other modalities complement and reinforce the original modalities. Non-complementary relationship refers to any relationship other than complementary relationship. In the film discourse Tea House, the relationship between various modalities belongs to complementary relationship, which can be further divided into reinforcing and non-reinforcing relationship. Reinforcing relationship means that one modality appears as the dominant modality, and the other modalities play a reinforcing role to the dominant modality. The non-reinforcing relationship refers to the indispensable presence of various modalities, such as the combination of visual and auditory modalities in movies. The visual and auditory modalities are integrated together and together contribute to the integrated meaning of the cinematic shot.

5.2. The Integration of Different Modalities in the Film

There are two modes of integration in the film discourse "Tea House". One is the integration between picture and sound, including the film's soundtrack. The most comprehensive model is the integration between picture, word and sound. It is described in detail in the case study in Chapter 4.

6. Conclusion

The conclusion section is the most important topic of this chapter and consists of two sections, namely, conclusion, limitations of this paper, and recommendations for further research. A detailed description of them is given below.

6.1. Conclusion

Based on the functional linguistics and visual grammar, this paper presents a multi-modal analysis of the film discourse Tea House in terms of the three meta-meanings of visual grammar. Firstly, as it is found that other modalities other than film language are considered, such as

image and sound, based on the scarcity of such research on film, this paper tries to provide new ideas for film analysis by focusing on modalities other than language in film; secondly, Kress and Van Leeuwen's visual grammar is limited to static images, while film contains multiple dynamic modalities including visual forms and auditory forms. Therefore, this paper not only considers static images but also describes the dynamic presentation of images; third, this paper is a comprehensive approach that considers all the meta-meanings of visual grammar; fourth, in previous work, visual modalities or auditory modalities have been studied without discussing the relationship or integration between them. This paper attempts to fill this gap by providing a comprehensive understanding of film discourse.

Specifically, the paper first provides an overview of previous work on film discourse and multi-modal discourse analysis. Then based on the theoretical framework of functional grammar and visual grammar, the paper proposes a framework for multi-modal analysis of the film discourse of *Tea House*. The analysis is conducted in terms of three meta-meanings, including the representational meaning, the interactive meaning and the compositional meaning in visual grammar, each of which is elaborated through the use of images from the film *Tea House*.

6.2. Limitations and Suggestions for Further Research

The limitations of this paper are mainly as follows: Firstly, although this paper attempts to provide a comprehensive analysis of the film in terms of the three meanings of visual grammar, it is impossible to give equal importance to each meaning due to the limited content. Secondly, although this paper illustrates the relationship and integration between visual and auditory modalities, there is less research on this aspect because the film *Tea House* is an old film with less sound and music material and is difficult to describe in words.

Acknowledgments

This work is supported by the 2021 Teaching Reform Project of Guangxi Undergraduate Education (2021)GB212).

References

- [1] Zhang Chuting. General Philosophy of Higher Education [M]. Beijing: Higher Education Press, 2009.
- [2] Zhang Delu. Exploring an Integrated Theoretical Framework for Multi-modal Discourse Analysis [J]. Chinese Foreign Languages 2009(01).
- [3] Metz, C. Film Language: A Semiotics of Cinema [M]. New York: Oxford UP, 1974.
- [4] Thompson, G. Introducing Functional Grammar [M]. London: Edward Arnold, 2000.
- [5] Brown, G. & G. Yule. Discourse Analysis [M]. Cambridge: Cambridge University Press, 1983.
- [6] Metz, C. Film Language: A Semiotics of Cinema [M]. New York: Oxford UP, 1974.
- [7] Eco, Umberto. Articulations of the cinematic code [A]. In Bill (ed.). Movies and Methods [C]. Nichols. Berkeley: University of California Press, 1976.
- [8] O'Toole, M. The Language of Displayed Art [M]. London: Leicester University Press, 1994.
- [9] Kress, G. & T. Van Leeuwen. 2006. Reading Images [M]. London: Routledge.
- [10] O'Halloran, K. Visual semiosis in Film [C]. Multimodal Discourse Analysis. London ; Continuum . 2004.
- [11] Hu Zhuanglin. Multi-modality in Social Semiotics Research [J]. Language Teaching and Research, 2007(1):1-10.
- [12] Zhu Yongsheng. Fifty Years of Discourse Analysis: Retrospect and Prospect [J]. Foreign Language, 2003(3):43-50.