A Multimodal Discourse Analysis of the Film "Witness for the Prosecution"

Zhang Yuhan
Beijing University of Chemical Technology, Beijing University of Chemical Technology (Changping Campus), No. 29 Nanjian Road, Nankou Town, Changping District, Beijing, 102202, P.R.China
zyh070031@126.com

ABSTRACT

This study conducts a multimodal discourse analysis of the 1957 courtroom suspense film "Witness for the Prosecution", investigating how verbal and visual modalities collaboratively construct and deconstruct the credibility of witness testimony. The analytical framework integrates the Attitude System within Martin's Appraisal System with the interactive meaning components of Visual Grammar proposed by Kress and van Leeuwen, specifically contact, social distance, and shooting perspective. Five representative courtroom clips, selected for their high density of adversarial discourse and multimodal resources, form the dataset. The analysis reveals that Judgement resources overwhelmingly dominate the verbal data, while Offer images and eye-level angle shots establish a baseline of objective narration; in contrast, Demand images, close-ups, and strategically deployed high and low angle shots intervene at pivotal narrative moments to modulate viewer empathy and cognitive judgment. Both prosecution and defense are based on clarifying the facts on judgmental interrogatives, increasing the power of cross-examination on direct camera gaze, and manipulating the authority of speech on perspective shifts. Finally, the multimodal design of this film shows that, finding facts in courtroom is just like performance, and jury-based adjudication is explained in manipulation of rhetorical and visual. Therefore, it provides an analytical framework for legal suspense films.

Keywords: Multimodal Discourse Analysis, "Witness for the Prosecution", Appraisal System, Visual Grammar, Courtroom Discourse

1. INTRODUCTION

In the globalization and diversification of media, human communication transcends the national boundaries of symbols of a language, and uses the resources of languages, images, sounds, body movements and other symbols together, which is the center of creating meaning and transmitting information. Multimodal discourse is not only an important carrier of daily communication and cultural production, its analysis also involves many disciplines such as linguistics, communication studies, sociology, paradigm etc. The perspective of this study has broken through the research boundary of only focusing on one symbol so far, and incorporated symbols from multiple sensory channels such as visual and auditory into the whole analysis framework. By studying the interaction mode and semantic generation mechanism between different modalities, we can explain the deep meaning and social function of discourse more comprehensively and deeply. Thus, it provides a new analytical approach for the interpretation of various text semantics. Through the interaction between verbal and non-verbal symbols, Film, as a typical form of multimodal narrative, further demonstrates two values of multimodal analysis in theory and practice.

2. LITERATURE REVIEW

2.1 Research on Multimodal Discourse Analysis at Home and Abroad

Multimodal Discourse Analysis (MDA), as an interdisciplinary research approach, aims to break through the limitations of traditional discourse analysis that only focuses on language symbols and systematically explore how various symbolic resources such as language, images, gestures, sounds, and spatial layouts work together to construct meaning in social interactions. Since the rise of discourse analysis in the middle of the 20th century, scholars have gradually realized that the production and understanding of meaning are far from being accomplished by a single language modality, but are rooted in the complex interaction of multiple senses and multiple symbols.

The theoretical origin of multimodal discourse analysis dates back to the 1960s. Roland Barthes first studied the interaction between image and language in "Rhetoric of Images" [1], and put forward the concepts of "anchoring" and "relay". After that, M.A.K. Halliday's theory of Systemic Functional Linguistics made the core theoretical foundation of multi-model research. He said that there are three metafunctions: ideational metafunction, interpersonal metafunction and textual metafunction. Then he emphasized that language is the system of social symbols [2]. This idea throws the question to scholars: do other symbolic systems have similar metafunctions? Can you create a "grammar" to describe the meaning potential of non-linguistic modalities such as visual and auditory? Therefore, we think of the social semiotics path of multi-distributed analysis.

The 1990s is an important period for the formation of the theoretical framework of multimodal discourse analysis. In "The Language of Display Art" [3], O'Toole first transformed the idea of Systemic Functional Linguistics into visual arts such as painting, sculpture and architecture, and put forward" display art grammar "including representation, modality and compositional functions. Almost at the same time, Gunther Kress and Theo van Leeuwen systematically established the theory of "Visual Grammar" in the famous book "Reading Images: Grammar of Visual Design" [4]. They made it very clear. Images can also have three method functions corresponding to the three method functions of a language. It is representational function, interactive function and compositional function. In addition, the definition of analytical dimensions is introduced in detail, such as contact, social distance, modality, information value, saliency and framing. In this framework, the objects of discourse analysis have been greatly expanded. Visual texts (such as advertisements, news pictures and charts) can also be systematically deconstruct now. Later, in Multimodal discourse: Modalities and Media of Contemporary Communication [5], they further pointed out that multimodality is a meaning practice created by orderly combining multiple symbolic modalities in a specific social and cultural context. Now, we have established the mainstream paradigm of multimodal discourse analysis based on Systemic Functional Linguistics and social semiotics.

In the 21st century, this field shows the vigorous development of diverse theories, methodology innovation and cross-domain application. On the road of theory, in addition to the social semiotics in the center, other important branch have gradually grown up. Sigrid Norris made "multimodal interaction analysis [6]", and made language, gestures, postures and eye contact with face-to-face communication, which gave me a methodology framework to analyze the interaction events in detail. In addition, under the influence of Lakoff & Johnson's conceptual metaphor theory [7], the multimodal metaphor analysis appeared. This is how metaphor transcends language, how to make it with many models such as images, sounds and gestures, and how to express it with body. In addition, critical discourse analysis is added to create multimodal critical discourse analysis. This is to find out the power relations and ideologies hidden behind multiple texts. These roads are not loose, but help each other and deepen the understanding of multimodal meaning-making complexity together.

When it comes to research topics and methods, our views on research have expanded from static texts to moving texts with various senses. Early research focused on static images such as print advertisements and textbook illustrations. With the explosive growth of digital media, dynamic multimodal discourse, such as movies, TV shows, short videos, video games, web interfaces, etc., has become a research frontier. Scholars have begun to work on developing theoretical models and annotation tools for analyzing dynamic, temporal symbolic streams. For example, O'Halloran [8] tried to combine Systemic Functional Linguistics and dynamic visual analysis to study cinematic discourse. John Bateman [9] made multimodal corpora, made computer-aided analysis software, and used the quantitative and qualitative data of corpora and big data. Thanks to the progress of these technical tools, we can now make a more detailed systematic analysis of multimodal communication.

The introduction and study of multimodal discourse analysis by Chinese scholars began in the early 21st century and quickly integrated into the mainstream context of international academic dialogue. Li Zhanzi [10] was the first to introduce the Visual Grammar theory of Gunther Kress and Theo van Leeuwen into China, initiating multimodal research from the perspective of social semiotics in China. Hu Zhuanglin [11] made an important distinction between the concepts of "multimodal" and "multimedia", emphasizing the semiotic essence of multimodal research. Zhang Delu [12], on the basis of integrating Systemic Functional Linguistics, Visual Grammar and Lim's integrated model, proposed a "comprehensive theoretical framework for multimodal discourse analysis" that includes the cultural layer, context layer, content layer and expression layer, providing a well-structured and highly operational localized model for analyzing multimodal discourse. Since then, domestic research has advanced rapidly in both theoretical and applied directions. In terms of theoretical research, scholars have explored frontier cross-disciplinary fields such as multimodal metaphor, multimodal critical discourse analysis, and multimodal cognitive research [13]. In terms of applied research, it covered a wide range of genres, from static movie posters, news pictures, textbooks to dynamic classroom teaching videos, television commercials, national image promotional videos, short-video platform content [14]. These studies not only verified the applicability of

Western theories in the Chinese context, but also enriched the practical content of multimodal discourse analysis through case studies, and began to focus on the unique multimodal representations of Chinese culture.

2.2 Multimodal Film Discourse Analysis Studies at Home and Abroad

The western academic community started exploring this field earlier. In the 1960s, Christian Metz [15] was the first to introduce structural linguistics and semiotic theory into film studies, creating the important branch of film semiotics. He regarded film as a symbolic system with its own grammatical rules, laying the theoretical foundation for analyzing film from a linguistic perspective. Since then, the research perspective has expanded, from early semiotic analysis focusing on the structure of the film itself to the examination of the audience's receptive psychology and cultural context. The advent of the digital age has further driven the innovation of research methods, such as O'Halloran's attempt to integrate visual analysis with Systemic Functional Linguistics to interpret the meaning of dynamic images [16]. Recent studies, such as those conducted by Bateman and Schmidt [17], have placed greater emphasis on constructing a refined analytical framework centered on specific film clips, dedicated to empirically answering the core question of how films reliably convey meaning.

In China, this research started a little late, but it developed quickly and expanded in all directions. China scholars actively absorb and summarize the ideas of western theories, especially Systemic Functional Linguistics and social semiotics, which are also used in the analysis of film texts. Early studies, such as Li Miaoqing's exploration [18] of the film "Raise the Red Lantern", initially attempted to use multimodal discourse analysis theory to grasp the various symbolic resources in the film as a whole. Ji Jing [19] drew on Appraisal Theory to construct a multimodal interaction model to analyze the interactive meaning of images and language in films; Yuan Yanyan and Zhang Delu [20] verified the feasibility of Visual Grammar theory in the analysis of specific discourse through the analysis of movie posters. These studies together reveal that in cinematic discourse, there are often complementary, reinforcing, and even contradictory relationships among different modalities, and their synergy can produce meaning-making effects that cannot be achieved by a single modality.

2.3 Review of the Research

Looking at the existing research on multimodal film discourse, the academic community has accumulated rich achievements in theoretical construction and text analysis, but there is still room for further deepening on several specific issues. First, existing research has mostly focused on the narrative structure, emotional expression, and aesthetic style of films, with less emphasis on court discourse as an independent analytical category, especially lacking in-depth discussions on the multimodal construction and deconstruction mechanisms of the credibility of testimony. Second, core issues such as confrontational discourse strategies, courtroom performance, and power manipulation, which are characteristic of legal-themed films and TV works, have not received sufficient attention in the multimodal research perspective, and the depth and breadth of interdisciplinary integration need to be enhanced. Third, existing research usually focuses on a single modality or general narrative commentary when analyzing courtroom scenes, and fails to systematically integrate analytical tools such as the Appraisal System and Visual Grammar to provide a refined explanation of the accumulation and reversal of credibility. In view of this, this study organically integrates multimodal discourse analysis with the appraisal system and Visual Grammar theory, focuses on the classic courtroom suspense film "Witness for the Prosecution", and aims to analyze the synergy mechanism of language and visual modalities in credibility construction and reversal, attempting to make up for the deficiencies of adversarial discourse analysis and institutional critical perspective in multimodal studies of legal films and TV, to provide a new analytical approach for multimodal interpretation of the suspense genre.

3. METHODOLOGY

3.1 Research Objectives

Based on multimodal discourse analysis theory, this paper pays attention to the usage and function of multimodal resources in film "Witness for the Prosecution" and thinks about the following questions.

Firstly, it analyzes how film uses language and visual modalities to create and destroy the credibility of witness testimony.

Secondly, explore how prosecution and defense can run adversarial discourse strategies with multiple resources in the cross-examination of courtroom.

Third, look at how the multimodal design of the file represents the performative nature of courtroom discourse, and consider the cognitive biases of jury system.

3.2 Research Content

Witness for the Prosecution (1957) is a suspense TV series directed by Billy Wilder. Produced by Agatha Christie's short story and stage play of the same name. Then, he was nominated for six Academy Awards. In this film, the complicated murder case will unfold. The rich widow Emily French was killed, and the charming poor young man Leonard Vole became prime suspect. Despite his claim of innocence, all the circumstantial evidence is against him. Sir Wilfrid Robarts, a shrewd and sharp senior lawyer with heart disease, was determined to defend him. As the trial progressed, the situation took a sudden turn for the better as the defendant's calm wife, Christine Vole, appeared as a Witness for the Prosecution and delivered a devastating testimony. In the tight narrative, the film profoundly shows the complex interaction of language and power in the legal field through the successive court debates, evidence Revelations and character relationship reversals, and the actors' body movements, eye contact and spatial position changes complement or contrast with the dialogue content, together forming a system of non-verbal symbols. Set against the backdrop of the British adversarial criminal trial system, the film's core feature is that a jury composed of non-legal professionals independently determines the facts of the case. Under this system, the testimony of witnesses in court, cross-examination between the prosecution and the defense, and rhetorical strategies constitute the decisive factors in the formation of jury evidence, and subtle differences in language and visual symbols can influence the direction of the verdict. This makes the film naturally an ideal text for examining the multimodal construction of the operation of court discourse power and the credibility of testimony. This study selected the following representative clips from the film as samples for analysis (Table 1).

Table 1. The film clips selected.

Numble		Scheme 2
34min-38min50s	Initial Client Consultation	This is the first meeting between the two defense attorneys and Christine Vole. Christine Vole told the lawyer that her marriage with the defendants was actually invalid because her ex-husband was still alive. She appeared to lack genuine feelings for her current husband, but still promised to provide the defendant's alibi in court. Christine's unreliability made the defense of the case much more difficult, and Sir Wilfrid Robarts decided to take over the case.
55min30s-1h2min05s	Cross-examination of the police	In this clip, the prosecution and defense questioned the police officer as a witness. The police confirmed the time of death, the weapon, and the fingerprints of three parties at the scene. No items were lost. The debate focused on whether the marks left on the broken Windows were forged and whether the bloodstains on the defendant's coat were type O, but both the deceased and the defendant had type O blood.
1h5min36s-1h11min20s	Cross-examination of the housekeeper	In this clip, the prosecution and defense questioned the housekeeper. The housekeeper not only claimed that the modification of the will occurred a week before the incident, but also claimed that she heard the defendant talking to the deceased on the night of the incident. But the defense lawyer questioned the reliability of the evidence by revealing the housekeeper's hearing problem and also pointed to the possible bias that the housekeeper lost her inheritance interest due to the change of the will.
1h19min40s - 1h27min42s	Defense Closing Argument and Defendant Testimony	In this clip, Sir Wilfrid Robarts delivered a brilliant defense statement, summarizing the remarks of the police, the housekeeper, and Christine Vole, hinting at the witness's loopholes to the jury, and summoning the defendant himself as a witness to testify. However, during the questioning session, the prosecution presented eyewitness testimony that was seriously unfavorable to the defense. The defendant, Leonard Vole, broke down under the cross-examination of the prosecution's lawyers.
1 hour 39 minutes 24s-1 hour 43 minutes 54s	Prosecution Cross-examination	In this clip, The defense attorney cross-examined Christine Vole, presenting her forged handwritten letter, and Sir Wilfrid Robarts read parts of the letter showing that Christine Vole planned to charge her husband with murder by giving false testimony, completely refuting Christine Vole's previous testimony.

3.3 The Basis for Sample Selection

This study selected the above five clips from the film "Witness for the Prosecution" as the analysis sample, mainly based on the following considerations. First, the selection of clips followed the "relevance" and "typicality" principles to ensure that the selected content was highly consistent with the intended research objective. The above segments respectively cover core legal scenarios such as pretrial meetings, witness examination, defendant testimony, and key evidence reversal, concentrating on the verbal confrontation between the prosecution and the defense, the application of multimodal visual strategies, and the systematic construction and deconstruction of the credibility of witness testimony, providing an intensive and typical analytical field. Secondly, in order to achieve a deep focus of the analysis, this study deliberately excluded clips such as scene build-up and irrelevant transitions and casual chats during the plot development process. Such content either lacks the confrontational nature of legal discourse or has a low density of multimodal resources, which is not conducive to directly responding to the research question. By excluding the irrelevant content, it ensures that the analysis focus is always on the key scenarios that carry the core narrative function and rich expressive resources, allowing the research to more intensively reveal the deep workings of multimodal resources in the game of legal discourse.

3.4 Research Methods

Based on the common theoretical foundation of Systemic Functional Linguistics, this study integrates the interactive meanings of the attitude system and Visual Grammar in the Appraisal System to construct a theoretical framework applicable to the multimodal analysis of the film "Witness for the Prosecution".

3.4.1 Appraisal System Theory

The theory of the Appraisal System, proposed by Martin in the early 1990s, is an important extension of Systemic Functional Linguistics. The theory focuses on how language users express positions, attitudes, and emotions through lexical and grammatical resources. Although traditional systemic functional grammar has covered the three metafunctions of ideational metafunction, interpersonal metafunction and textual metafunction, it has not systematically dealt with the evaluative meaning in discourse. The proposal of the Appraisal System fills this theoretical gap. The system consists of three interrelated sub-systems: the Attitude system, the Engagement system, and the Graduation system. The Attitude system is further divided into affect, Judgement and Appreciation. Affect is the core of the Attitude system, belonging to the category of emotional responses. This is the speaker's direct emotional response and psychological tendency to the evaluated object, and it is the most basic and direct form of attitude expression. Judgement entered the ethics and morality area. This is the speaker's evaluation of human behavior according to social rules, ethical standards or rules, which is used to judge the legitimacy, standardization and social value of his behavior. On the other hand, endorsement enters the area of aesthetics and social value. It is the speaker's value judgment on the evaluated object, including aesthetic value, social significance and importance. In this article, affect resources shows various emotional reactions such as anger and sadness of the characters in the film. Judgement resources refers to the specific judgments made by the characters on the truth of the case, the contents of testimony, the actions of witnesses, etc. Appreciation resources is an abstract evaluation of some phenomena or people in terms of nature, value and significance.

3.4.2 Visual Grammar theory

Visual Grammar, developed by Gunther Kress and Theo van Leeuwen based on Systemic Functional Linguistics, extends the three metafunctions of language to the visual modality, forming a ternary framework of representational meaning, interactive meaning, and compositional meaning. Among them, the interactive meaning focuses on the construction of the relationship between the image and the viewer, mainly achieved through three core elements: contact, social distance, and attitude. Contact is divided into direct contact and indirect contact based on whether the image participants have life attributes, both of which have the dual functions of "Demand" and "Offer", breaking the either-or limitations in the original classification. Social distancing is indicated by the size of the frame of the shot. Close-ups or close-ups represent intimate relationships, while long shots correspond to distant social relationships. Attitude is determined by perspective: a horizontal Angle reflects the extent to which the viewer is involved in the world of the image, while a vertical Angle reflects the power relationship between the viewer and the participant. The theory provides a systematic and actionable approach to analyzing the construction of interpersonal meaning in images. In this paper, contact resources are used to analyze the impact of changes in the gaze of film characters on audience perception; Social distance resources are used to analyze strategies for the use of framing distances; Attitude resources are used to explore the different effects produced by different shooting angles.

3.4.3 Construction of the integrated framework

This study mainly referred to the multimodal discourse analysis synthesis theoretical framework proposed by Zhang Delu for the construction of the theoretical framework. Based on Systemic Functional Linguistics, this framework systematically examines the composition and operation mechanism of multimodal discourse from five levels: culture, context, meaning, form, and media, providing a clear theoretical path for multimodal analysis. Based on the common theoretical foundation of Systemic Functional Linguistics, this paper integrates the Attitude system in Martin's Appraisal System and the three core elements of interactive meaning in Gunther Kress and Theo van Leeuwen's Visual Grammar - contact, social distance, and attitude - to construct an analytical framework for multimodal analysis of the film "Witness for the Prosecution". To avoid repetition, this paper refers to the attitudinal element in the interactive meaning of the visual modality as the shooting Angle. Figure 1 presents the integrated analytical framework.

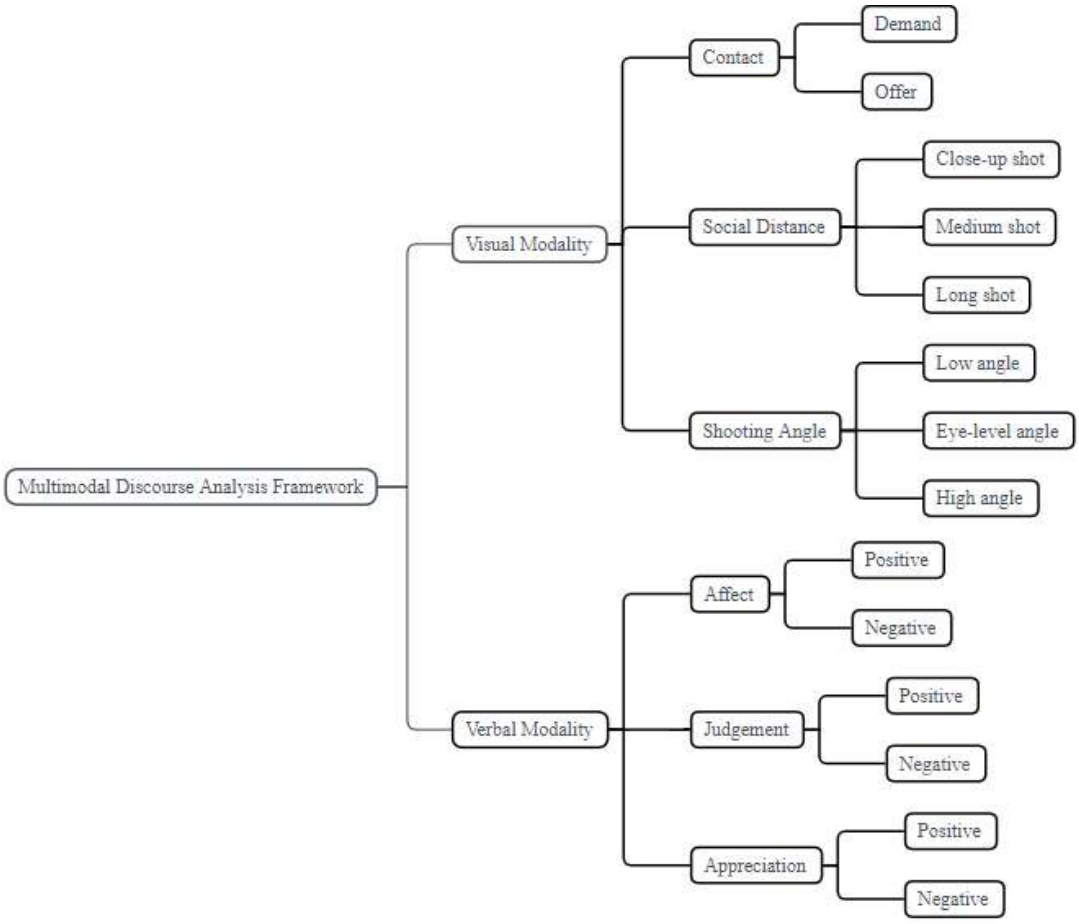


Figure 1. Screenshot of the integrated framework.

3.5 Research Tools

3.5.1 ELAN

ELAN is a multimedia annotation software developed by the Max Planck Institute for Psycholinguistics in the Netherlands and constitutes one of the core tools in Multimodal Discourse Analysis. The software enables timeline-based multi-tier annotation of video and audio files, allowing researchers to create hierarchical "tiers" for precise alignment and annotation of multimodal signals within the footage. In the present study, ELAN was employed to visualise the deployment of various elements in the visual modality, and to investigate the deployment strategies and combinatorial effects of Contact, Social Distance, Shooting Perspective resources (Figure 2).

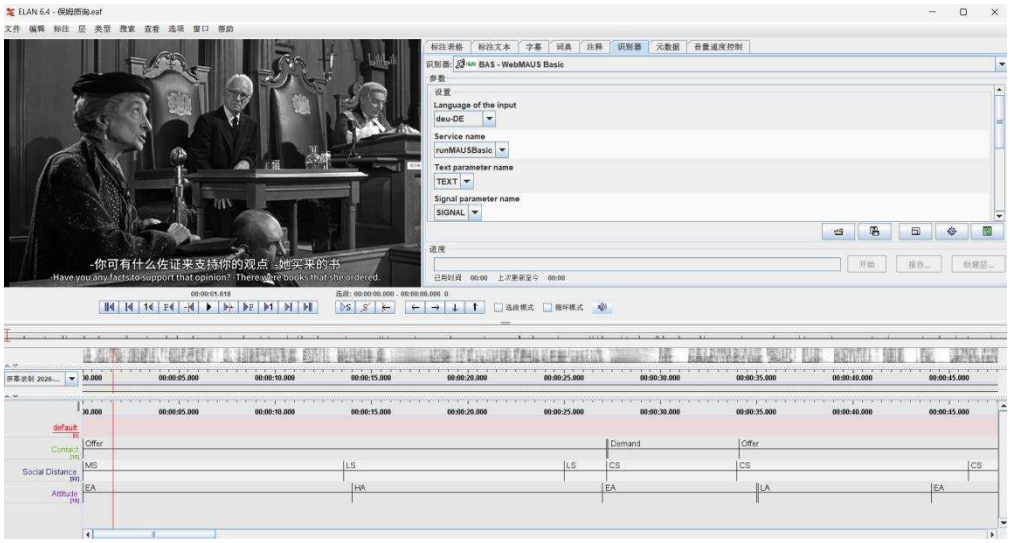


Figure 2. Screenshot of the ELAN software.

3.5.2 BFSU Qualitative Coder

BFSU Qualitative Coder is a qualitative data annotation tool developed by the research team at the National Research Centre for Foreign Language Education, Beijing Foreign Studies University. The tool is specialised in processing textual data and supports flexible coding and categorisation of transcribed film dialogue scripts. In this study, the linguistic modality texts were imported into the software for systematic annotation and quantitative analysis of the characters' lines, so as to explore the usage patterns of Affect, Judgement, and Appreciation resources (Figure 3).

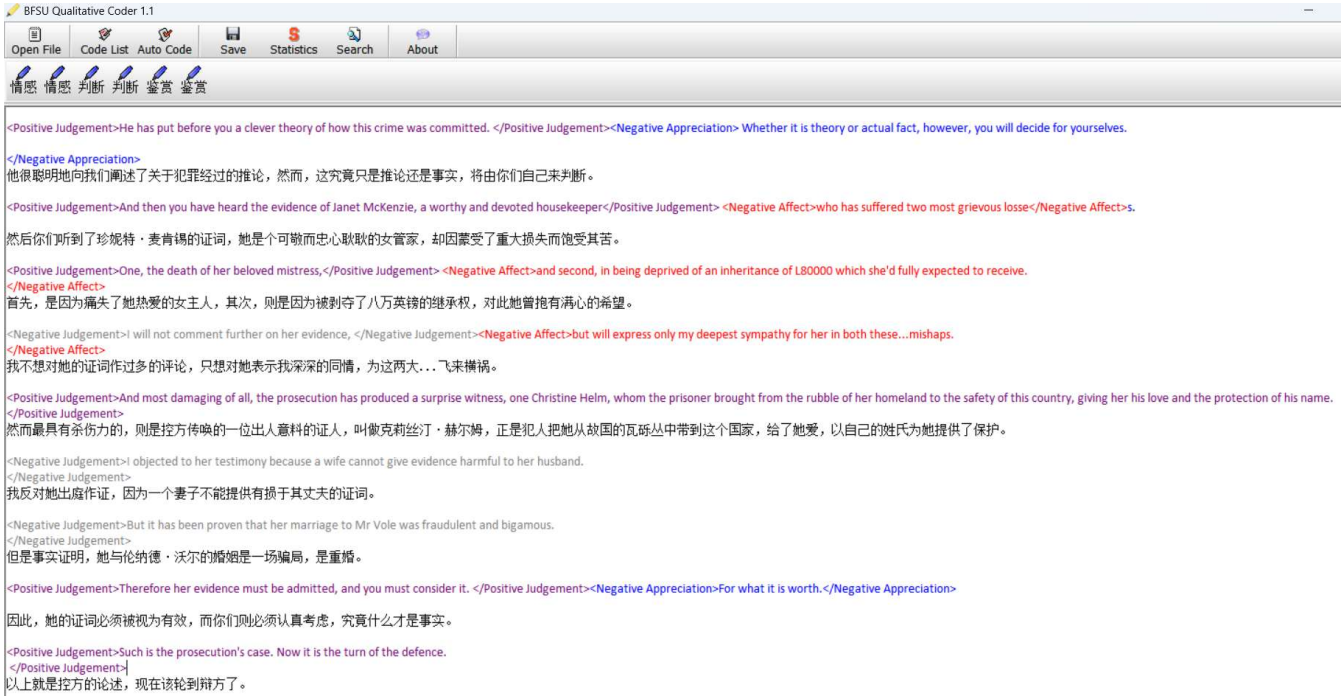


Figure 3. Screenshot of the BFSU Qualitative Coder software.

4. RESULTS AND DISCUSSION

4.1 Annotation Standards for Verbal Resources

4.1.1 Affect resource analysis

First, identify emotional expressions, label the various emotional response phrases of characters in the lines, and distinguish between positive and negative Affect resources, such as polite interactions, emotional outbursts, resistance and questioning, stress venting, and other emotionally inclined content; Secondly, explore how such emotional expressions interfere with the construction of the credibility of testimony and guide the emotional tendencies of the audience and the jury; Finally, analyze the strategic value of emotional discourse and explore how anomalous emotional expressions contribute to the overall construction of the film's suspense narrative.

4.1.2 Judgement resource analysis

Extract judgmental expressions, mark the critical remarks in the lines based on ethics, law and social norms regarding the authenticity of characters' behaviors, testimonies and their motives and characters, and distinguish between positive and negative Judgement resources. After that, analyze the adversarial logic of the judgment discourse, explore how the prosecution and defense use such expressions in cross-questioning to compete for the right to interpret the facts of the case, systematically construct the rationality of their own testimony and undermine the credibility of the other side's narrative; At the same time, explore how such judgment expressions accumulate performance tension and influence the logic of the jury's determination of the facts of the case.

4.1.3Appreciation resource analysis

Identify appreciative expressions, label the value evaluation discourse in the lines regarding the rationality of the relationship, the candor of the witness, the nature of the behavior and the social phenomenon, and distinguish between positive and negative Appreciation resources; Next, analyze the auxiliary function of appreciation discourse and explore how it works in conjunction with Judgement and Affect resources to influence the credibility of testimony. At the same time, it lays down detailed clues for the development of the film plot; Finally, explore how such expressions of appreciation, through value judgment, expose the performance nature of the legal process and complete a critical reflection on the legal system.

4.2 Data analysis

Table 2. The quantity and distribution of verbal modality resources.

Types	Positive/Polarity	Negative/Polarity	Total/Polarity
Affect	5 / 1.44%	40 / 11.53%	45 / 12.97%
Judgement	180 / 51.87%	99 / 28.53%	279 / 80.40%
Appreciation	6 / 1.73%	17 / 4.90%	23 / 6.63%
Total	191 / 55.04%	156 / 44.96%	347 / 100%

According to Table 2, among Affect resources, positive emotions accounted for 1.44%, negative emotions accounted for 11.53%, and the total accounted for 12.97%; Among the Judgement resources, positive judgments accounted for 51.87%, negative judgments accounted for 28.53%, and a total of 80.40%; In terms of Appreciation resources, positive appreciation accounted for 1.73%, negative appreciation accounted for 4.90%, and in total 6.63%. Overall, positive attitude resources accounted for 55.04%, negative attitude resources accounted for 44.96%, totaling 100%.

Among the three resources of Affect, Judgement, and Appreciation, Judgement resources are clearly dominant, mainly for the following reasons: First, the corpus selected in this study is mostly derived from cross-examination of witnesses by lawyers in court scenarios. Such questioning often uses leading questions in the form of declarative or multiple-choice questions, aiming to control the direction of the witness's response and force them to give a definite "yes" or "no" response, thereby locking in the facts and limiting their interpretive space. For example, "The blade is sharp, do you agree?" That is, by using closed-ended questions to make it difficult for witnesses to elaborate and thereby frequently activate Judgement resources.

Secondly, if one party's lawyer raises an inappropriate question, the other party's lawyer tends to use Judgement sentences to reveal the lack of rigor in the other party's argument when raising a challenge. For example: My knowledgeable friend refers to the attacker as the "murderer", but we are not sure if the murderer is a man or a woman. If it were a female murderer, that would be convincing."

Thirdly, during the defense statement stage, lawyers often extensively employ declarative sentences to create an objective and professional expression effect. Especially when dealing with the unfavorable evidence presented by the prosecution, lawyers often lead the jury to form a specific perception through formally neutral but actually implicitly inclined statements. For example: "It is up to you to judge whether this is an inference or a fact." Or "First, because of the loss of her beloved mistress, and second, because of the deprivation of her inheritance of £80,000, which she had hoped for." Such expressions, though not a direct conclusion, suggest a specific path of factual interpretation through the order of narration, emotional embedding and choice of detail. For juries not trained in systematic law, this kind of guidance can easily influence their independent judgment. In addition, when meeting with relatives of the parties involved or discussing the case with colleagues, lawyers also commonly use declarative sentences to describe the facts of the case, supplemented by critical discourse for analysis, thereby further expanding the scope of the use of Judgement resources.

The use of both Affect resources and Appreciation resources is relatively limited, with Affect resources outnumbered Appreciation resources. The vast majority of emotional expressions do not come from legal professionals, but from ordinary roles such as witnesses and audience members who have not received systematic legal training. For example, after learning that the defendant might inherit a huge sum of money, the housekeeper of the deceased grumbled indignantly: "If he could get a penny of that money, that would be a great injustice." Another example is when Christine Vole lost control of her emotions as her lawyer read out handwritten evidence against her, and shouted on the spot, "You go to hell!" Such expressions of emotion often undermine the credibility of the expresser as a witness because of their strong subjectivity and emotional nature. But at the same time they inject intense dramatic tension into the highly procedural trial process. For example, in the film, the performance in the court made by Christine Vole and defendant in order to overthrow the unfavorable situation makes jury an objective judgment of the audience through very emotional communication. This emotional performance not only breaks the fixed boundary between emotion and reason in court's story, but also strengthens the film's suspenseful narrative effect.

Although the number is small, the resource of Appreciation plays an important role in helping and consolidating the content of the story. First of all, the juxtaposition of appreciative language and Judge's resource will create internal tension that will reduce the speaker's credit. For example, the words of the deceased housekeeper: "I have no hostility towards him. He is a cunning, income, scheming rogue, but I have no hostility towards him. "Here are negative Appreciation ("Cunning and Incompetent", "Planning Rogue") and positive judgement no hostility. Secondly, the resource of Appreciation is also used to shape characters from the side. For example, nurse commented on lawyer's words: "This old fox, Wilfrid! This is the nickname we gave him, and he truly lives up to it." This connotative label not only portrays the character's professional style, but also reveals the performative quality that exists in the lawyer's courtroom strategy. In addition, Appreciation resources also have functional significance in regulating jury cognition. For example, after witness issued a subjective judgment, judge would immediately intervene: "Ladies and gentlemen of the jury, it is entirely possible that a woman read Disraeli without considering marrying a man younger than her. "This prevents the jury from forming a one-sided logic. However, this remedy is highly dependent on the judge's personal expertise and on-the-spot judgment, and its effectiveness is difficult to guarantee through institutionalization.

There are more Positive Judgement than negative Judgement. Judging from the overall distribution of Judgement resources, it is like this. On the one hand, lawyers from both the prosecution and the defense tend to use positive Judgement discourse in court statements to shape a professional, rigorous and reliable image, thereby enhancing the authority and persuasiveness of the language. This kind of expression often uses an objective sentence structure, downplays the subjective tone, makes the argument seem rational and legitimate, and is more likely to be accepted by the judge and the jury. On the other hand, lawyers generally use the "commitment-endorsements - cross-examination" strategy: that is, first guide witnesses to confirm their statements in court, then emphasize the reliability of their previous statements, and finally use that statement as evidence for face-to-face cross-examination. Under this pattern, witnesses answer "yes" more frequently than "no", which gives affirmative Judgement the dominant position in the trial interaction.

In contrast, the Affect resources are dominated by negative emotions. This distribution is closely related to British expression habits: British rhetoric prefers implicit irony and indirect subtlety, often conveys negative attitudes with polite and gentle irony, while maintaining social dignity and restrained etiquette with a clear stance. In the highly formal and conflict-intensive courtroom, this language strategy helps to reduce sharp oppositions, making the expression appear steady,

proper and more credible. In other words, the “amount” of negative emotions is large, but mostly presented in the form of implicit negative affect.

4.3 Annotation Standards for Contact Resources

Analyzing interactive resources first requires analyzing whether there is direct contact between the character and the audience.

In this study, living characters directly saw that the visual representation of camera was called direct contact, and the corresponding label was “Demand”. From the outside of the frame, it is indirect contact, labeled “Offer”. If it is inanimate object, use indirect contact and mark it as “Offer”.

The core function of “Offer” is to allow the viewer to view it from the observer’s point of view, thus causing rational thinking and judgment on the information in the picture. “Demand” provides audience with an immersive presence through the visual connection between the character and the viewer. Emotional tension and expressive tendencies of the character For example, when lawyer asks the witness, in the direct contact with the camera, audience can feel the oppression in words. Through witness to avoid the camera’s indirect contact, the viewer can also examine and judge testimony authenticity from subtle expressions or movements.

First of all, to analyze interactive resources, we must analyze whether there is a direct context between character and audience.

In this study, living characters classifies the visual representation of directly seeing the camera as direct contact, and takes the corresponding label as “Demand”. Looking outside the frame, it is indirect contact, and the label is “Offer”. If it is inanimate object, it is classified as indirect contact as “Offer”.

The core function of “Offer” is to allow the viewer to observe from the observer’s point of view, thus causing rational thinking and judgment on the information in the picture. “Demand” gives the audience a sense of immersive presence through the visual connection between the character and the viewer, allowing them to directly sense the emotional tension and expressive tendencies of the character. For example, when lawyer asks witness, directly seeing the direct contact of the camera can make audience feel the opposition in the language. Through witness to avoid the camera’s indirect contact, the viewer can also examine and judge the authenticity of their testimony from subtle expressions and movements.

4.4 Data Analysis

4.4.1 Demand image



Figure 4. Screenshot of the film “Witness for the Prosecution”.

Taking the first meeting in the film (Figure 4) as an example, the statistics show that demand images account for about 32.16%, which build a "call-response" two-way interaction through direct eye contact, achieving a strong emotional impact and narrative guidance. For instance, when Sir Wilfrid Robarts faces Christine Vole, the defendant's wife who appears to cooperate but actually has a sly tone in her words and deliberately conceals key information, he puts on his glasses and focuses his gaze to examine and interrogate her, using a direct look into the camera to directly convey the tension and oppression of the scene to the audience, completing the immediate construction of narrative tension.

4.4.2 Offer image



Figure 5. Screenshot of the film "Witness for the Prosecution".

67.84% of the offer images dominate the segment. These images, exemplified in Figure 5, in a one-way information transmission structure without eye contact, objectively present the background information of the case, the interaction between Christine Vole and the lawyers on both sides, and provide core information support for viewers to judge the credibility of the characters' testimony and deduce the truth of the case.

4.4.3 Overall Analysis

Table 3. The quantity and distribution of contact resources.

	Demand	Offer	Total
Number	31	47	78
Duration	1428.33s	2699.08s	4127.41s
Frequency	34.61%	65.39%	100.00%

According to the data in Table 3, the total time of contact resources is 4,127.41 seconds. Among them, demand images is 1,428.33 seconds, and offer images is 2,699.08 seconds longer. From the perspective of f frequency distribution, the demand image accounts for 34.61% of all content resources, and the offset image is more, accounting for 65.39%. Moreover, both of them cover all contact resources.

Due to the use of a large number of offer images, it can be seen that the film is centered on creating objective navigation. Let audience see the details of the story inside and outside the court, and put audience in the position of an objective viewer. Although the proportion of demand images is not large, it just enters the important place of story, moving the cognition and judgment of audience. By working together, it is helpful to make suspenseful narrative and lay the cognitive foundation for plot twists. Not only that, we also show how prosecution and defense use opposing and guiding legal discourse strategically through multimodal resources. This highlights the performance nature of court process and guides audience to think deeply about the narrative authenticity.

4.5 Annotation Standards for Social Distance Resources

When analyzing interactive resources, we must also analyze the social distance between children and audience, which was created by the framing of lens.

In this study, we think that the close-up above chest is close-up. “Medium distance” is above knees, and main figure is less than half of the frame. It is considered that the full sub-object and the surrounding environment are long distance when viewed completely or from a distance.

In Close-up shots, viewers can pay attention to the micro-expressions and body details of characters, and lead them to the emotional tension and verbal confrontation medium shots in courtroom to balance detail and scene, and viewers can feel the expression strategy of characters’ words and body movements. In the macro perspective of long shots, viewers can look at the spatial layout and power interaction of courtroom from bystander’s point of view and rationally understand the actions of legal process.

4.6 Data Analysis

4.6.1 Close-up shots



Figure 6. Screenshot of the film “Witness for the Prosecution”.

Take Christine Vole’s questioning section (Figure 6) as an example. According to statistics, close-up shots account for about 37.95%. Moreover, these shots are concentrated on the core dramatic conflict points of their paragraph. For example, when “prosecution lawyer” shows “key evidence” forcefully, camera will capture Christine Vole’s “emotion” change. She tidied up all the sophistry, through emotional outbursts, and finally confirmed it with tears. It also provides the immediate response of “defendant (Leonard Vole) to” trial process “. Close-ups firmly fixes the attributes of audience attention in the micro-expressions or body details of the character’s face by shrinking the frame space and embedding visual focus. In this way, the emotional tension of the characters in the courtroom will be as large as possible to create a narrative fit that matches plot pace very well.

4.6.2 Medium shot

Medium shots make up about 38.32%, roughly the same proportion as close-ups. This shot, illustrated in Figure 7, covers the core questioning process of lawyers questioning witnesses, presenting letter evidence, and reading letter texts in court. It can clearly show the details of lawyers’ body language and expressions, and also fully incorporate the expressions and immediate reactions of fellow lawyers and jury members, forming an effective visual coordination with the core figures. The camera language precisely and tenaciously portrays the masterful and shrewd character of Sir Wilfrid Robarts, presenting the narrative thread of how he uses fake letters to obtain key testimony from witnesses, turns the course of the trial, and ultimately takes the initiative in the trial. The medium shot, which combines the characteristics of the core

character and the narrative of the trial space, not only provides a suitable visual carrier for the plot development of this section, but also offers crucial support for the dramatic presentation of the trial discourse and body language.



Figure 7. Screenshot of the film “Witness for the Prosecution”.

4.6.3 Long shot

Long shots make up about 35.29% of the total, though a smaller proportion, they play an irreplaceable role in spatial narrative and perspective guidance. This shot size, illustrated in Figure 8, covers the entire courtroom from the public gallery, the witness stand, the prosecution and defense seats to the judge’s seat, including all the subjects involved in the trial, such as witnesses, lawyers, judges, juries and spectators, presenting the spatial structure and personnel layout of this special courtroom in an all-round way. The macroscopic panoramic view allows the viewer to stand in an objective spectator’s perspective and clearly examine the power hierarchy and interactions corresponding to different seats in the courtroom. This detach visual language downplays the dramatic emotional rendering of individual characters, guiding the audience to observe the progress of the trial process from a more rational and neutral perspective, and to clearly understand the logical norms of the legal process. It not only builds a complete spatial framework for the narrative of the trial segments, but also provides a visual support for the film to lead the audience to think deeply about the legal system, truth and justice.



Figure 8. Screenshot of the film “Witness for the Prosecution”.

4.6.4 Overall Analysis

Table 4. The quantity and distribution of social distance resources.

	Close-up shot	Long shot	Medium shot	Total
Number	55	43	45	143
Duration	675.07s	499.38s	556.36s	1730.81s
Frequency	39.00%	28.85%	32.14%	99.99%*

The statistics in Table 4 show that among all samples, the total cumulative duration of the three types of shots is 1,730.81 seconds. Among them, the cumulative duration of close-ups is the highest at 675.07 seconds, followed by medium shots (556.36 seconds), and the cumulative duration of long shots is the lowest at 499.38 seconds. In terms of frequency, close-ups accounted for 39.00%, followed by medium shots at 32.14% and long shots at 28.85%, covering 99.99% of the statistical sample in total.

The high-frequency close-up shots precisely manipulate the audience’s judgment of the credibility of the characters’ testimonies through close-range visual presentation, anchoring the audience’s emotional empathy and stance orientation; The complete presentation of the discourse confrontation and power game between the prosecution and the defense in the middle shot, complementing the close-up efficiently, clearly shows how lawyers strategically mobilize multimodal resources to transform them into core confrontational tools for seizing the initiative in discourse and reshaping factual narratives; The long shot builds up the sense of authority and ritual of the judicial process through panoramic presentation, while providing viewers with a space for reflection to detach from the narrative and examine the content. The three types of shots form a clear and complementary visual narrative synergy that works together to create a suspenseful atmosphere, reveal the specific operational logic of multimodal discourse confrontation in the courtroom, and build channels for the audience to question the judicial system itself.

4.7 Annotation Standard for Shooting Perspective Resources

Analysis of interactive resources requires further analysis of the construction of shooting angles. In this study, the visual representation of a camera position above the character’s eye level is classified as a high-angle shot, that of a camera

position level with the character's eye level is classified as an eye-level angle, and that of a camera position below the character's eye level is classified as a low-angle shot.

The "high-angle shot" is likely to create a scrutinizing perspective for the viewer, weaken the authority of the subject's discourse, and raise questions about the viewer's expression. The "eye-level angle" builds a neutral perspective, guiding the audience to objectively view the verbal expressions of both sides and the strategies of confrontation in court, and to make rational judgments. The "low-angle shot" creates an upward perception perspective for the audience, strengthening the deterrence of the discourse of roles such as judges and lawyers.

4.7.1 Eye-level angle



Figure 9. Screenshot of the film "Witness for the Prosecution".

Take the defense's statement and the defendant's testimony segments as examples (Figure 9). The proportion of eye-level shots is as high as 93%, constituting the absolutely dominant visual language. Eye-level angle plays an important narrative role in the defense statement of Sir Wilfrid Robarts. This scene is a very important point for defense to get jury's support. Lawyer should not only point out the loopholes in the prosecution's chain of evidence in an orderly way, but also have a psychological impact on jury by guiding words and emotional mobilization. By putting the camera on the eye level, audience can enter jury's position and intuitively feel the direction of defense strategy. As a result, there are strong presence and immersion.

4.7.2 Low-angle shot



Figure 10. Screenshot of the film “Witness for the Prosecution”.

“Low-angle shots” are usually used to show speakers who can be trusted more. Take Police witness in Figure 10 as an example. He has the meaning of objective and neutral symbolic in the organization. If you start shooting from Low angle, you can strengthen his authoritative image and make audience believe in his testimony. Similarly, authoritative people like judge often use this perspective. Further investigation shows that when making shooting angles, films often imitates the viewing position of character in the show. Especially imitating jury in courtroom and line of sight in spectator. For example, if the “witness stand” is physically designed higher, the camera will usually be placed lower. This perspective strategy enhances narrative immersion and involvement by imitating the experience seen by people present, and audience can intuitively participate in the trial process.

4.7.3 High-angle shot



Figure 11. Screenshot of the film “Witness for the Prosecution”.

It is even rarer to use high-angle shot in movies. Generally speaking, this is used to put forward a disadvantaged or disadvantaged person, or to show that this person's words are not credible. Take the scene in Figure 11 as an example. Prosecution just sent out the evidence to defendant, who is the weaker side of trial in that kind of pressure. At this point, jury will naturally be more cautious and skeptical about his testimony. High-angle shot further reduces that person's authority, and makes the audience criticize the character's position and the credibility of his speech inexplicably. This perspective approach not only conforms to the rhythm of narrative, but also conveys the position of the role in the power structure and the "offensive and defensive stances" of court in the game through vision.

4.7.4 Overall Analysis

Table 5. The quantity and distribution of shooting perspective resources.

	Eye-level angle	High angle	Low angle	Total
Number	54	12	29	95
Duration	1303.047s	187.18s	246.761s	1736.988s
Frequency	75.02%	10.78%	14.20%	100.00%

As shown in Table 5, of all the clips selected in this study, the total duration of the three shooting angles is 1,736.988 seconds. Among them, the eye-level angle was significantly longer, reaching 1,303.047 seconds; The duration of the high-angle shot was 187.18 seconds, and that of the low-angle shot was 246.761 seconds. In terms of frequency distribution, the eye-level angle accounts for the largest proportion, at 75.02%; High-angle shots accounted for 10.78%, and low-angle shots accounted for 14.20%. This indicates that in the sample, the eye-level angle is absolutely dominant in both duration and frequency dimensions, while the visual presentation of high and low angle shots is clearly secondary.

The three shooting perspectives serve the suspenseful narrative and the critical reflection of the legal process through a clearly coordinated strategy of primary and secondary. The absolutely dominant eye-level angle creates an objective neutral illusion of the courtroom scene by constructing an equal power relationship between the viewer and the character, placing the viewer in the main position of the juror and laying the visual foundation for the accumulation of credibility of testimony; Strategically distributed low-angle and high-angle shots intervene at key points. The former externalizes the power flow of discourse confrontation by giving the character visual authority, while the latter reveals the vulnerability and manipulability of the individual in the vast process with a bird's-eye view. This perspective configuration, with equal involvement as the main and power intervention as the secondary, reveals from the side the performance characteristics of fact-finding under the jury system.

5. CONCLUSION

Based on multimodal discourse analysis theory, this study systematically examined the visual and linguistic modal resources of the classic legal suspense film "Witness for the Prosecution" and fully responded to the pre-set research questions and core objectives.

The study found that the film achieved a systematic construction and subversive reversal of the credibility of testimony through the hierarchical synergy of visual and linguistic modalities. On the visual level, the dominant offer images, eye-level shots, and evenly distributed medium and long shots together form an objective and neutral narrative foundation; demand images, close-ups, and strategically distributed high-angle and low-angle shots precisely intervene at key plot points, manipulating the audience's emotional empathy and cognitive judgment, laying the groundwork for the final narrative twist. At the linguistic level, Judgement resources form the core of the trial discourse, constructing a rational narrative framework through closed-ended questions and guiding statements, supplemented by Affect and Appreciation resources to guide cognitive tendencies, and together complete the construction of the suspenseful narrative.

In the core courtroom scenario, both the prosecution and the defense transform multimodal resources into the core tools of adversarial legal discourse. On the visual level, the demand images by looking directly at the camera intensify the sense of oppression of questioning, close-ups capture micro-expressions to guide the judgment of the credibility of testimony, and high-angle and low-angle shots control the authority of character discourse; At the language level, the high-frequency Judgement resources lock the facts and limit the other party's interpretive space, the strategic emotional performance influences the jury's position, and the Appreciation resources deconstruct the other party's narrative rationality to achieve the struggle for the court's discourse power.

The multimodal design of the film fully reveals the inherent performative nature of the court process and the deep-seated flaws of the jury system. The eye-level angle places the audience in the main position of the jury, visually presenting the so-called objective fact-finding, which is essentially a cognitive manipulation performance by both the prosecution and the defense through multimodal symbols, exposing the core predicament of the jury system where non-professional judges are easily guided by discourse and fact-finding is easily influenced by subjective construction. This study provides a reusable analytical framework for interpreting legal suspense films as multimodal. At the same time, it is also a new reference to understand the role of symbols in the judicial scene and the power game.

5.1 Insufficient research

In this study, the selection method of sample has several limitations. During the research, only a few representative court core clips were selected from the files, and multimodal resource annotation and systematic analysis were carried out with the help of analysis software. The whole content of film is not included in the research scope. Because the non-core segments used for plot build-up, character profiling and background introduction in the file are not analyzed in detail, the multimodal resources in the file are not analyzed in detail. In the future research, by expanding the scope of sample analysis and including the whole content of film, we can draw a more comprehensive and persuasive research conclusion.

REFERENCES

[1] R. Barthes, Rhetoric of the Image. 1964.
[2] M. A. K. Halliday, Language as Social Semiotic: The Social Interpretation of Language and Meaning. London: Edward Arnold, 1978.
[3] M. O'Toole, The Language of Displayed Art. London: Leicester University Press, 1994.
[4] G. Kress and T. van Leeuwen, Reading Images: The Grammar of Visual Design. London: Routledge, 1996.
[5] G. Kress and T. van Leeuwen, Multimodal Discourse: The Modes and Media of Contemporary Communication. London: Arnold, 2001.
[6] S. Norris, Analyzing Multimodal Interaction: A Methodological Framework. New York: Routledge, 2004.
[7] Z. Zhong, S. Wen, and S. Chen, "Research trends in multimodal metaphor: A bibliometric analysis," Front. Psychol., vol. 14, p. 1144725, Apr. 2023, doi: 10.3389/fpsyg.2023.1144725.
[8] K. L. O'Halloran, Multimodal Discourse Analysis: Systemic-Functional Perspectives. London: Continuum, 2004.
[9] J. A. Bateman, J. Wildfeuer, and T. Hiippala, Multimodality: Foundations and Introduction. Berlin/Boston: De Gruyter Mouton, 2008.
[10] Z. Li, "Social semiotic analysis of multimodal discourse," Foreign Language Research, no. 5, 2003.
[11] Z. Hu, "Multimodalization in social semiotic research," Language Teaching and Research, no. 1, 2007.
[12] D. Zhang, "Exploration of a comprehensive theoretical framework for multimodal discourse analysis," China Foreign Languages, no. 1, 2009.
[13] D. Zhang and E. Guo, "The dual perspectives of multimodal discourse analysis: The connection and complementarity between the social symbol view and the conceptual metaphor view," Foreign Languages (Journal of Shanghai International Studies University), no. 3, pp. 20-28, 2013.
[14] D. Feng, "Implications of multimodal discourse analysis theory for film poster design," Film Literature, no. 16, pp. 157-158, 2012.
[15] C. Metz, "Le cinema: langue ou langage," Communications, no. 4, 1964.
[16] K. L. O'Halloran, Multimodal Discourse Analysis: Systemic-Functional Perspectives. London, U.K.: Continuum, 2004.
[17] J. Bateman and K. H. Schmidt, Multimodal Film Analysis: How Films Mean. New York, NY, USA: Taylor & Francis, 2013.
[18] M. Li, "Multimodal discourse analysis of adapted films: A case study of 'High Red Lantern'," Film Literature, no. 15, 2007.
[19] J. Ji, "Multimodal discourse analysis of the interactive significance of film discourse," Ph.D. dissertation, Shandong Univ., Jinan, China, 2010.
[20] Y. Yuan and D. Zhang, "Social semiotic analysis of multimodal movie poster discourse," Journal of Jining University, no. 2, 2012.