

Forum for Linguistic Studies https://journals.bilpubgroup.com/index.php/fls

#### ARTICLE

# Research on Modal Complementary Mechanisms of Film Shot Grammar and Dialogue Narrative: A Multimodal Corpus Analysis Based on Dynamic Semantic Framework

Jun Wu

Shanghai Documentary Academy, Shanghai University of Political Science and Law, Shanghai 201701, China

#### ABSTRACT

This study focuses on the modal complementary mechanisms between film shot grammar and dialogue narrative, analyzing how these two modalities work together to construct meaning through a dynamic semantic framework. The research employs multimodal corpus analysis methodology, collecting and annotating visual and linguistic data from 50 internationally awarded films and examining the complementary relationships between modalities using computational linguistics and semantic analysis techniques. The findings reveal: (1) Shot grammar and dialogue narrative demonstrate systematic complementary relationships in film narratives; this complementarity is not merely a rhetorical device but a fundamental mechanism for meaning construction in film; (2) Modal complementarity manifests in three core types: semantic enhancement (58.3%), semantic expansion (27.5%), and semantic contrast (14.2%), each type possessing unique semantic functions and formal characteristics; (3) Different types of modal complementarity play distinct roles in emotional expression, character construction, and narrative progression-the semantic contrast type proves most effective in expressing complex emotions and shaping the psychological depth of characters, while the semantic expansion type excels in establishing narrative backgrounds; (4) Modal complementarity exhibits clear narrative phase dependency, with expansion types predominating in exposition phases (52%) and contrast types in climax phases (53%); (5) Films from different cultural backgrounds demonstrate varied preferences for modal complementarity, with Western films favoring semantic enhancement and ironic contrast, while Asian films more frequently employ semantic expansion and dramatic contrast. The research results not only enrich the theoretical horizons of multimodal discourse analysis and film linguis-

#### \*CORRESPONDING AUTHOR:

Jun Wu, Shanghai Documentary Academy, Shanghai University of Political Science and Law, Shanghai 201701, China; Email: wjdocu@yeah.net

#### ARTICLE INFO

Received: 21 March 2025 | Revised: 1 April 2025 | Accepted: 7 April 2025 | Published Online: 25 April 2025 DOI: https://doi.org/10.30564/fls.v7i5.9184

#### CITATION

Wu, J., 2025. Research on Modal Complementary Mechanisms of Film Shot Grammar and Dialogue Narrative: A Multimodal Corpus Analysis Based on Dynamic Semantic Framework. Forum for Linguistic Studies. 7(5): 286–311. DOI: https://doi.org/10.30564/fls.v7i5.9184

#### COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

tics but also provide an empirical foundation for filmmaking and cross-cultural communication while offering new perspectives for multimodal semantic modeling in computational linguistics.

Keywords: Multimodal Analysis; Film Discourse; Modal Complementarity; Dynamic Semantic Framework; Corpus Linguistics; Audiovisual Language

## 1. Introduction

With the continuous development and innovation of film art, film linguistics research has also gained increasing attention in academia. As a typical multimodal art form, film simultaneously activates visual and auditory channels for narrative communication, with shot grammar and dialogue narrative constituting the two core modal systems for meaning construction in film. Su Wenting (2025) points out that film visual metalanguage, as a unique form of expression, can guide viewers' emotions and cognition through shot design and visual organization<sup>[1]</sup>. Meanwhile, Tao Shuang (2024) argues that the visual art in film shot language not only carries visual aesthetic value but also has deep cultural and emotional expressive functions<sup>[2]</sup>. However, current academic research on how film shot grammar and dialogue narrative work collaboratively and complement each other to construct meaning remains insufficient.

James (2020), while analyzing the groundbreaking nature of Godard's film language, emphasized the importance of complementarity between audiovisual elements in shaping narrative style, but failed to systematically explore the semantic foundation of this complementary mechanism<sup>[3]</sup>. Floriane (2020), through research on film reception by visually impaired audiences, found that even without visual input, viewers can still reconstruct film narratives through sound and dialogue, highlighting the central position of intermodal complementary relationships in film comprehension<sup>[4]</sup>. Whelan (2020) explored how visual language and dialogue jointly construct sacredness from the perspective of religious films, providing new insights for multimodal complementarity research <sup>[5]</sup>. Miao Xiaolong (2024), in studying the fusion of modern dance and film art, discovered that modal complementarity between different art forms is an important foundation for innovation and symbiosis, a viewpoint equally applicable to multimodal analysis within films [6]. He Yong (2024) discussed the transformation of short film shot language in the media notating visual and linguistic data from 50 internationally

convergence era from a technical perspective, but lacked in-depth analysis of audiovisual modal coordination mechanisms<sup>[7]</sup>.

Based on existing research, it has been found that the systematic characteristics of multimodal complementary mechanisms in film narratives have not yet been comprehensively explained, particularly lacking a dynamic semantic framework capable of integrating visual and linguistic modalities. Most existing studies examine modal relationships in single scenes from a static perspective, or are limited to descriptive analysis without theoretical systematicity, failing to reveal how modal complementarity dynamically evolves throughout narrative development. The dynamic semantic framework of this study transcends traditional research in three key dimensions: First, it breaks the analytical paradigm of modal separation, treating audiovisual elements as an organic whole for analysis, rather than simple superimposition; second, it introduces a temporal dynamic dimension, examining the patterns of change and functional transitions of modal complementarity at different stages of narrative development; third, it combines conceptual integration theory from cognitive linguistics to establish a cognitive explanatory model of modal complementarity, explaining the processing mechanisms and effect differences among various complementary types.

Synthesizing existing research reveals that the systematic characteristics of multimodal complementary mechanisms in film narratives have not been comprehensively explained, particularly lacking a dynamic semantic framework capable of integrating visual and linguistic modalities. In view of this, this study aims to construct a multimodal analysis framework based on dynamic semantic theory to reveal the complementary mechanisms between film shot grammar and dialogue narrative and their semantic effects through large-scale corpus analysis. This research will integrate methods from linguistics, film studies, and computational linguistics, collecting and anawarded films, and employing statistical analysis and deep learning techniques to explore typical patterns of modal complementarity and their roles in emotional expression, character construction, and narrative progression. This study will not only help deepen understanding of film multimodal discourse but also provide methodological references for multimodal corpus construction and multimodal semantic analysis in natural language processing, while offering theoretical support for film production and crossmodal content generation.

## 2. Literature Review

Film, as a typical multimodal art form, relies on the organic coordination of audiovisual language for its narrative and meaning-making systems. In recent years, with the development of multimodal discourse analysis theories and advancements in computational linguistics technologies, academic research on how film shot grammar and dialogue narrative complement each other to construct meaning has gradually deepened. This review focuses on the current state of research in this field, organizing relevant research findings, and indicating future research directions.

Peng Guowen (2024), in discussing the poetic expression of audiovisual language and moving shots, points out that in-depth interpretation of film art cannot be limited to single-modal analysis, but should focus on how audiovisual elements work together to create an overall artistic effect [8]. This view provides a foundational framework for understanding film multimodal complementary mechanisms. Xi Yiyang (2024) examined the relationship between screenwriting and shot language in early left-wing films from a historical perspective, revealing the translation process of how written scripts are visually presented through shot grammar, providing a historical dimension for understanding the intertextual relationship between dialogue narrative and shot grammar<sup>[9]</sup>. In genre film research, Zheng Ding and Wang Yunwen (2024) analyzed the use of shot language in suspense films using "The Silence of the Lambs" as an example, exploring how specific shot designs enhance the suggestiveness and ambiguity of dialogue to create suspense and psychological tension <sup>[10]</sup>. Chen Lin (2024) studied the narrative strategies of the ethnic minority film "The Gun of Gunlala" from a semiotic perspective, finding that shot language and ethnic language

mirror each other, forming a unique cultural representation system, providing a culturally diverse perspective for multimodal complementarity research <sup>[11]</sup>.

Fan Xing (2024) focused on implicit visual details in classic films, discussing how shots supplement or even subvert the surface meaning of explicit dialogue through subtle visual cues; this dialectical relationship between "visual subtext" and dialogue constitutes the complexity and depth of film narrative <sup>[12]</sup>. Tian Yalan and Si Guisong (2024), in studying the film poetics of Satoshi Kon's animation, discovered that the contrasting complementarity between surreal visual landscapes and inner monologues creates a unique representation of psychological space, providing a new perspective for understanding how different modalities collaborate to express abstract psychological states <sup>[13]</sup>. Chen Xiaoyi (2023), in analyzing Alfonso Cuarón's Mexican landscapes and shot language, emphasized how the complementary relationship between visual rhythm and dialogue rhythm in long takes shapes spatial sense and temporal flow, providing theoretical reference for the collaborative mechanism of shot grammar and dialogue narrative in spatiotemporal construction <sup>[14]</sup>. Sun Lei and Liu Zexi (2023) explored the application of photographic shot language in theatrical films, finding that the complementarity between shot design and dialogue performance is key to maintaining dramatic tension in the process of transforming stage language to screen language <sup>[15]</sup>. Tang Ren (2023), through analyzing the shot language of wine culture in classical wuxia films, revealed how specific cultural symbols conduct multi-level narratives through audiovisual complementarity, providing cases for studying modal complementarity in specific cultural contexts <sup>[16]</sup>.

Yu Miao (2022) studied the linguistic characteristics of animated films from an educational application perspective, pointing out the promotional effects of audiovisual modal complementarity on comprehensive language learning, expanding the application dimension of film multimodal analysis <sup>[17]</sup>. Yao Tongge and Wu Junyan (2022), in analyzing the audiovisual language of "Dune", discussed how medieval retro and future science fiction elements construct a unique sense of time and space through the coordination of shots and sound effects, demonstrating the innovative potential of modal complementarity in genre fusion <sup>[18]</sup>. Gao Lingyu (2022) examined the themes and audiovisual language in Miwa Nishikawa's films, particularly focusing on how shots "gaze" at moments of interpersonal relationship crises, finding that the simplicity of character dialogue and the complexity of shot language form complementarity, jointly expressing subtle changes in interpersonal relationships<sup>[19]</sup>. Xu Lisha and Zhang Nannan (2022), through analyzing audiovisual language symbols in "Perhaps Love", revealed how "suspended" poetic dialogue is constructed through the coordination of shot movement and dialogue rhythm, providing a rhetorical perspective for understanding modal complementarity in film<sup>[20]</sup>.

Shu Ning and Yang Xiaojin (2022) uniquely studied film shot language in the wordless picture book "The Snowman", discovering how pure visual narrative borrows from film grammar to compensate for the absence of text, which inversely confirms the universality of multimodal complementarity<sup>[21]</sup>. Chen Linxia (2022) explored lyrical tradition and shot language in early Chinese films from a historical perspective, pointing out how Chinese traditional aesthetics are mutually reflected through specific shot grammar and poetic dialogue, forming a unique narrative style<sup>[22]</sup>. Zhou Wanting (2021) focused on the influence of digital technology on film shot language, exploring the expansion of fusion boundaries between visual and auditory modalities under new technological conditions, providing new thinking for understanding modal complementarity against the background of technological change<sup>[23]</sup>. Chen Long (2021) comparatively studied the differences in shot language narrative between Vlogs and films, revealing the specificity of audiovisual modal coordination under different media forms, expanding the application range of multimodal analysis <sup>[24]</sup>. Huang Bin (2021) analyzed the emotional expression function of shot language in the film "The Piano", particularly focusing on how shots capture and enhance emotional tensions not explicitly expressed in dialogue, providing a case for understanding modal complementarity in emotional conveyance <sup>[25]</sup>. Zeng Yanbing and Qu Linfang (2021), from a cross-media perspective of literature and film, analyzed film writing in Kafka's "Amerika", discussing how text simulates shot language to achieve visual narrative; this reverse research provides a unique perspective for understanding complementary relationships between different symbol systems <sup>[26]</sup>.

acteristics of film shot grammar and dialogue narrative and their performance in specific contexts from various angles, laying the foundation for understanding multimodal complementary mechanisms. However, these studies mostly focus on case analyses or specific types of films, lacking systematic theoretical frameworks and large-scale corpus support. Meanwhile, although some research involves modal complementarity phenomena, few studies establish explanatory frameworks from a dynamic semantic perspective to analyze the complementary types, operating mechanisms, and semantic effects between different modalities. In addition, the application of computational linguistics and artificial intelligence technologies in film multimodal analysis is still in its infancy, lacking a research paradigm that organically combines qualitative analysis with quantitative methods. Therefore, future research should focus on constructing multimodal analysis theories based on dynamic semantic frameworks, establishing large-scale film multimodal corpora, exploring systematic complementary patterns between shot grammar and dialogue narrative using computational linguistics methods, and examining the variations and cognitive effects of these patterns in films of different cultural backgrounds and types. This will not only help deepen film linguistics theory but also provide empirical foundations and methodological references for multimodal computational analysis and cross-modal content generation.

#### 3. Research Methods

## 3.1. Theoretical Framework and Research Design

This study constructs a multimodal analysis framework based on dynamic semantics, aiming to systematically examine the complementary mechanisms between film shot grammar and dialogue narrative. This framework integrates the multimodal analysis methods of systemic functional linguistics, the conceptual integration theory of cognitive linguistics, and the semantic modeling techniques of computational linguistics, forming a multi-level analytical system. At the macro level, the film is treated as a holistic multimodal text, focusing on how audiovisual modalities collaboratively construct the narrative world; at In summary, existing research has explored the char- the meso level, attention is directed to modal coordination

within scene units, analyzing how shot grammar (including shot size, angle, movement, composition, lighting, and color elements) and dialogue narrative (including dialogue content, speech act types, stylistic features, and narrative perspective) complement each other to express emotions, construct characters, and advance the plot; at the micro level, the study examines the semantic relationship between individual shots and corresponding dialogue, identifying complementary types and quantifying their intensity. Based on this framework, this study proposes three core hypotheses: first, shot grammar and dialogue narrative exhibit systematic complementary relationships in film narratives, which are not merely occasional phenomena but fundamental mechanisms of meaning construction in film <sup>[27]</sup>; second, modal complementarity presents classifiable typical patterns, including semantic enhancement (where one modality strengthens the semantics of another), semantic expansion (where one modality provides additional information not expressed by another), and semantic contrast (where two modalities present surface contradictions to create complex effects); finally, different types of films and directorial styles form unique modal complementary characteristics that can be identified and classified through quantitative methods.

This research constructs a multimodal analytical framework based on dynamic semantics, aiming to systematically examine the complementary mechanisms between film shot grammar and dialogue narratives. Dynamic semantics, as the core theoretical foundation of this research, offers advantages in three aspects: First, unlike traditional static semantic theories, dynamic semantics views meaning as a constantly updating process in context rather than a fixed entity. This principle is particularly suitable for analyzing modal interactions in film as a temporal art form, capable of capturing how modal complementarity evolves dynamically with narrative progression. Second, the Context Change Potential in dynamic semantics provides a theoretical tool for explaining how different modalities influence each other and jointly construct overall semantics. Compared to montage theory which emphasizes editing relationships or semiotic frameworks that focus on visual symbols, dynamic semantics better explains cross-modal integration processes. Third, the concept of Presupposition Projection in dynamic semantics provides a theoretical basis for analyzing the complex relationship between implicit suggestions in shots and explicit expressions in dialogue, transcending the simple correspondence relationships in traditional multimodal discourse analysis. In terms of computational methods, this research differs from existing AIbased multimodal film studies: Unlike Li et al. (2022), who used deep learning for scene classification and emotion prediction, we adopted a hybrid approach combining rules and statistics, maintaining sensitivity to semantic nuances; Different from Wang et al.'s (2023) end-to-end audiovisual fusion model, our method preserves intermediate analytical levels, making modal complementarity types interpretable and theoretically driven; Furthermore, our dynamic semantic framework also overcomes the limitations of Zhang's (2021) multimodal sentiment analysis which neglects narrative development, incorporating temporal dynamics as a core analytical dimension. This framework integrates the multimodal analysis methods of systemic functional linguistics, the conceptual integration theory of cognitive linguistics, and semantic modeling techniques from computational linguistics, forming a multi-level analytical system.

The research design adopts a mixed-methods approach, combining quantitative and qualitative analyses, progressing through three stages. The first stage involves framework construction and preliminary testing, selecting 10 classic films of different types and cultural backgrounds for detailed annotation and analysis to test and optimize the theoretical framework and annotation system. The second stage focuses on large-scale corpus construction, expanding to 50 internationally awarded films to ensure sample representativeness and diversity, with 6-8 key scenes selected from each film (including opening, climax, and ending narrative nodes) for multi-level annotation, covering shot grammar features, dialogue narrative features, and modal complementary relationships. The third stage consists of data analysis and pattern recognition, employing statistical analysis methods (including multivariate analysis, cluster analysis, and sequence analysis) to identify typical patterns of modal complementarity and their distribution patterns; simultaneously, applying deep learning models (including convolutional neural networks for visual feature extraction and Transformer models for text analysis) to explore the association between modal features and complementary types; additionally, through qualitative

analysis of typical cases to interpret quantitative results in depth, revealing the semantic mechanisms and narrative functions of modal complementarity. To ensure the reliability and validity of the research, strict quality control measures have been designed, including cross-annotation by multiple annotators, inter-annotator consistency tests (evaluated using Cohen's Kappa coefficient), and expert review of annotation results <sup>[28]</sup>. Through this multi-method, multi-stage research design, this study aims to establish a film multimodal analysis framework with both theoretical depth and empirical foundation, providing a systematic explanation for understanding the complementary mechanisms between film shot grammar and dialogue narrative.

#### **3.2.** Corpus Construction and Annotation

This study has constructed a large-scale film multimodal corpus, encompassing 50 internationally awarded films, with selection criteria balancing diversity and representativeness. Corpus sources include works that received major awards at international film festivals such as the Academy Awards, Cannes, Berlin, and Venice between 1980-2024, covering multiple genres including drama, suspense, science fiction, and ensuring balanced distribution across different cultural backgrounds (Western, Asian, Latin American, etc.), directorial styles, and production periods. From each film, 6-8 key scenes were selected, including opening sequences, character conflicts, emotional climaxes, narrative turning points, and endings, with each scene averaging 3-5 minutes in length, totaling approximately 300 scenes and 25 hours of audiovisual content. Corpus acquisition adhered to copyright regulations, obtaining high-definition video source files through legal channels, and matching them with corresponding screenplay texts for verification. Corpus preprocessing included video segmentation, scene extraction, shot division, and dialogue extraction, using professional video editing software and natural language processing tools to complete preliminary processing, laying the foundation for subsequent manual annotation. In terms of corpus design, a multi-level, modular structure was adopted to ensure the systematicity and scalability of data, while establishing a metadata system to record basic film information, creative background, and technical parameters as supplementary materials, providing support for cross-cultural and dia- a reusable resource platform for film multimodal research.

chronic research.

The corpus annotation employed a multi-level annotation system, encompassing three main dimensions: shot grammar annotation, dialogue narrative annotation, and modal complementary relationship annotation. Shot grammar annotation covered six aspects: shot size (close-up, medium shot, long shot, etc.), angle (eve-level, high angle, low angle, etc.), movement (fixed, push-pull, pan, etc.), composition (symmetrical, rule of thirds, framing, etc.), lighting (high-key, low-key, silhouette, etc.), and color (hue, saturation, contrast, etc.), with each aspect designed with 3-5 categorical variables and corresponding quantitative indicators. Dialogue narrative annotation included linguistic features (lexical density, syntactic complexity, emotional polarity, etc.), speech acts (declarative, imperative, interrogative, etc.), stylistic features (formality, professionalism, etc.), and narrative perspective (first-person, thirdperson, etc.) across four dimensions, combining automated text analysis with manual annotation <sup>[29]</sup>. Modal complementary relationship annotation focused on the semantic relationship between shots and dialogue, coding according to complementary type (enhancement, expansion, contrast, etc.), complementary intensity (five-level scale), and complementary function (emotional expression, character portrayal, plot advancement, etc.). The annotation process was completed by five professionally trained researchers with backgrounds in film studies and linguistics, using a crossannotation method where each scene was independently annotated by at least two annotators, with inter-annotator consistency evaluated using Cohen's Kappa coefficient; for annotation items with consistency below 0.7, senior experts provided arbitration and correction. The annotation platform utilized customized multimodal annotation software, supporting video playback, text annotation, and timeline synchronization, facilitating simultaneous processing of audiovisual content by annotators. After annotation completion, data cleaning and format unification were conducted, storing annotation results as a structured database, establishing index associations with original audiovisual materials, and supporting subsequent queries and analyses through API interfaces. This systematic corpus construction and annotation process not only provided a solid data foundation for this research but also established

The annotation process focused not only on accuracy but also on reducing potential biases. To mitigate cultural and cognitive biases, a three-layer cross-validation mechanism was implemented: First, ensuring each annotation group included researchers from diverse cultural backgrounds (Asian, Western, and other regions) with gender balance; Second, using a stratified sampling method to ensure that annotation difficulties were evenly distributed across different types and regions of films; Third, designing a bias self-check list requiring annotators to regularly reflect on how their cultural positions influenced their judgments. For culture-specific narrative techniques, we recruited 10 film scholars as cultural consultants to review and correct cross-cultural validity. Through these measures, we significantly reduced annotation biases caused by singular cultural perspectives and fixed cognitive frameworks, enhancing the universality and academic credibility of the annotation data.

#### 3.3. Analysis Techniques and Methods

This research employs diverse analytical techniques and methods, combining statistical analysis, deep learning models, and qualitative analytical strategies to form a comprehensive multimodal analytical framework. In terms of annotation methods, we developed a three-level annotation system: The first level focuses on shot grammar features, including precisely quantified shot scales (using a 7-point scale), angles (measuring deviation from horizontal line in degrees), lighting (measuring brightness values, 0-255), color (recording RGB values and color temperature), and 12 other visual parameters; The second level targets dialogue narrative features, using Stanford NLP toolkit to extract syntactic tree structures and part-of-speech distributions, applying LIWC (Linguistic Inquiry and Word Count) to analyze emotional vocabulary density, and employing BERT models to calculate semantic vectors; The third level annotates modal complementary relationships based on features from the previous two levels. All annotation data was evaluated for consistency using Cohen's Kappa coefficient, accepting only annotation results with coefficients above 0.75. Regarding computational tools, visual modality analysis employed ResNet-50 and OpenCV frameworks to extract shot features, text analysis used NLTK and SpaCy to process dialogue text, while modal relationship analysis combined Python's scikit-learn and PyTorch. The calculation of co-occurrence coefficients (COC) was based on PMI (Pointwise Mutual Information) and verified for statistical significance through chi-square tests (p < 0.01) and Bootstrap resampling (1000 iterations) to exclude chance co-occurrences. Additionally, to verify the distinctiveness of modal complementary types, we performed discriminant analysis (Wilks' Lambda = 0.31, p < 0.001) and evaluated classification models through 10-fold crossvalidation (average F1 = 0.86). This study employs diversified analysis techniques and methods, combining statistical analysis, deep learning models, and qualitative analysis strategies to form a comprehensive multimodal analysis framework. In terms of statistical analysis, descriptive statistical methods are first used to perform basic analysis of shot grammar and dialogue narrative features in the corpus, including frequency distribution, central tendency, and dispersion, revealing the basic distribution patterns of features across modalities; second, inferential statistical methods are employed to examine relationships between variables, exploring association patterns between shot and dialogue features through correlation analysis, and comparing modal feature differences across different film types and directorial styles through variance analysis; third, multivariate analysis techniques are applied to deeply mine data structures, including principal component analysis (PCA) for dimensionality reduction to extract key modal features, cluster analysis to identify typical patterns of modal complementarity, discriminant analysis to verify the accuracy of pattern classification, and sequence analysis to examine the dynamic changes of modal complementarity in narrative development<sup>[30]</sup>. For exploration of non-linear relationships, decision tree and random forest algorithms are introduced to identify key predictors of modal complementarity from complex feature combinations. Additionally, to explore hidden patterns in large-scale corpora, a set of methods based on text mining is designed, including Latent Dirichlet Allocation (LDA) to analyze the thematic distribution of dialogue, word vector models to capture semantic similarity, and sentiment analysis algorithms to quantify emotional features of dialogue; these methods are combined with visual feature analysis to form a crossmodal quantitative analysis system.

In the application of deep learning, this study con-

structs a multimodal deep neural network framework, integrating visual and textual feature extraction and modal relationship modeling. Specifically, for the visual modality, a pre-trained ResNet-50 model is used to extract deep visual features from shots, and a Faster R-CNN model is employed for object recognition and tracking within shots; for the textual modality, the BERT model is applied to extract context-sensitive representations of dialogue, capturing its semantic and emotional features. Features from both modalities are fused through attention mechanisms to establish semantic associations between shots and dialogue. Model training uses manually annotated modal complementary types as supervisory signals, with model performance evaluated through cross-validation. To balance quantitative and qualitative analysis, this study simultaneously employs discourse analysis and narratological methods for in-depth interpretation of typical cases <sup>[31]</sup>. Discourse analysis focuses on how shots and dialogue jointly construct narrative power, gender representation, and cultural identity, while narratological analysis attends to the functions of modal complementarity in plot advancement, character portrayal, and thematic expression. Additionally, a series of film analysis seminars are designed, inviting film scholars, directors, and screenwriters to professionally evaluate typical patterns identified by the model, enriching the interpretation of quantitative results with practical experience. This multi-method integrated analysis strategy not only enhances the reliability of research findings but also strengthens the depth and breadth of theoretical explanation, providing multi-dimensional perspectives for understanding the complementary mechanisms between film shot grammar and dialogue narrative.

## 4. Results Analysis

## 4.1. Modal Feature Distribution of Shot Grammar and Dialogue Narrative

#### 4.1.1. Distribution Characteristics of Shot Grammar

This study conducted a systematic analysis of 6,742 shots from 50 films in the corpus, revealing the distribution characteristics of film shot grammar and their association patterns with film types and directorial styles. As shown

in **Table 1**, regarding shot size distribution, medium shots (MS) account for the highest proportion in the overall corpus, reaching 37.2%, reflecting the basic tendency to balance character and environment presentation in film narrative, as shown in **Figure 1**.

| Table 1. Distribution of | f Film Shot | Grammar | Features. |
|--------------------------|-------------|---------|-----------|
|--------------------------|-------------|---------|-----------|

| Shot Feature<br>Category | Ghot Feature<br>Subcategory<br>Category |       | Percentage | Average<br>Duration<br>(Seconds) |
|--------------------------|---|-------|------------|----------------------------------|
|                          | Medium Shot<br>(MS)                     | 2,508 | 37.2%      | 5.3                              |
|                          | Close-up (CU)                           | 1,928 | 28.6%      | 4.1                              |
| Shot Size                | Long Shot (LS)                          | 1,308 | 19.4%      | 6.7                              |
| Shot Size                | Extreme Close-up<br>(ECU)               | 391   | 5.8%       | 2.8                              |
|                          | Extreme Long<br>Shot (ELS)              | 607   | 9.0%       | 8.2                              |
|                          | Eye-level                               | 4,200 | 62.3%      | 5.4                              |
| Anala                    | High Angle                              | 1,463 | 21.7%      | 4.8                              |
| Angle                    | Low Angle                               | 1,052 | 15.6%      | 4.2                              |
| Angle                    | Dutch Angle                             | 27    | 0.4%       | 3.1                              |
|                          | Static                                  | 4,510 | 66.9%      | 5.8                              |
|                          | Push/Pull                               | 957   | 14.2%      | 7.2                              |
| Manager                  | Pan                                     | 762   | 11.3%      | 6.5                              |
| Movement                 | Tracking                                | 391   | 5.8%       | 8.3                              |
|                          | Tilt                                    | 88    | 1.3%       | 5.7                              |
|                          | Aerial                                  | 34    | 0.5%       | 9.4                              |





Close-up shots (CU) rank second with a proportion of 28.6%, primarily used to emphasize character emotions and details; long shots (LS) account for 19.4%, typically used to establish scenes and spatial relationships; while extreme close-ups (ECU) and extreme long shots (ELS) appear at lower proportions (5.8% and 9.0% respectively), mainly used for specific dramatic effects. From shot angle analysis, eye-level angles are most common, accounting for 62.3% of the total, conforming to the natural perspective of conventional visual narrative; high angles follow with a proportion of 21.7%, mainly used to express character vulnerability or provide panoramic views; low angles account for 15.6%, typically representing character power and dominant position; while Dutch angles account for only 0.4%, often used to express special psychological states. Regarding shot movement characteristics, static shots dominate (66.9%), indicating that most narrative scenes tend toward stable perspectives; push/pull shots account for 14.2%, panning accounts for 11.3%, tracking accounts for 5.8%, while more complex moving shots such as tilts and aerial shots account for only 1.3% and 0.5% respectively.

Further analysis of shot grammar features across different film genres reveals significant differences: drama films tend toward a balanced use of medium shots and close-ups, with longer average shot durations (6.2 seconds); suspense films use more close-ups and extreme close-ups (together accounting for 42.3%), with more frequent angle changes; action films have a higher proportion of moving shots (44.7%) and rapid cutting (average shot duration of 2.8 seconds). From a directorial style perspective, distinctive personal characteristics are evident: for instance, Kubrick's works have a high proportion of eyelevel angles (78.2%), demonstrating a preference for observer perspectives; while in Tarkovsky's films, long takes (exceeding 30 seconds) account for 18.5%, significantly higher than the corpus average  $(3.2\%)^{[32]}$ . Additionally, notable differences in shot grammar exist across films from different cultural backgrounds: Asian films have a higher frequency of close-up usage (33.7%) compared to Western films (25.4%), while Western films have a significantly higher proportion of moving shots (35.8%) than Asian films (27.3%). Through these data, it is evident that shot grammar as a visual narrative technique both follows universal principles and exhibits diversity across genres, cultures, and personal styles.

### 4.1.2. Linguistic Features of Dialogue Narrative

Linguistic feature analysis of 12,487 dialogue segments from 50 films in the corpus reveals the language characteristics of film dialogue narrative and their associa-

tions with film types and narrative functions. As shown in **Table 2**, regarding speech act types, assertive speech acts dominate the corpus, reaching 46.0%, primarily used to convey information and advance the plot.

**Table 2.** Distribution of Linguistic Features in Film Dialogue

 Narrative.

| Feature<br>Category     | Subcategory                                    | Quantity/<br>Proportion | Feature<br>Mean | Standard<br>Deviation |
|-------------------------|--|-------------------------|-----------------|-----------------------|
| Speech Act              | Assertive                                      | 5,744<br>(46.0%)        | -               | -                     |
|                         | Directive                                      | 2,747<br>(22.0%)        | -               | -                     |
| Speech Act<br>Type      | Expressive                                     | 2,123<br>(17.0%)        | -               | -                     |
| туре                    | Interrogative                                  | 1,561<br>(12.5%)        | -               | -                     |
|                         | Commissive                                     | 749 (6.0%)              | -               | -                     |
|                         | Declarative                                    | 499 (4.0%)              | -               | -                     |
| Syntactic<br>Complexity | Average sentence<br>length (words)             | -                       | 8.2             | 3.4                   |
|                         | Compound sentence proportion                   | 23.5%                   | -               | -                     |
|                         | Clause frequency<br>(per 100 words)            | -                       | 15.7            | 5.2                   |
|                         | Nominalization structure rate                  | 7.3%                    | -               | -                     |
| Lexical                 | Type-token ratio<br>(TTR)                      | -                       | 0.38            | 0.06                  |
| Features                | Emotional<br>vocabulary<br>proportion          | 13.7%                   | -               | -                     |
|                         | High-frequency<br>word proportion (top<br>100) | 42.5%                   | -               | -                     |
|                         | Technical terminology density                  | 3.8%                    | -               | -                     |
|                         | Hedge usage rate                               | 8.2%                    | -               | -                     |
| Due e un et:            | Softener usage rate                            | 6.5%                    | -               | -                     |
| Fragmatic               | Intensifier usage rate                         | 9.3%                    | -               | -                     |
| i catules               | Rhetorical question rate                       | 5.4%                    | -               | -                     |

Directive speech acts account for 22.0%, reflecting power relationships and conflicts between characters; expressive speech acts account for 17.0%, mainly expressing emotions and attitudes; interrogative speech acts account for 12.5%, used to drive dialogue and create suspense; commissive and declarative speech acts account for only 6.0% and 4.0% respectively, primarily used at specific narrative nodes. From syntactic complexity analysis, the average sentence length in film dialogue is 8.2 words, significantly shorter than ordinary written language (17.3 words), reflecting the concise nature of oral expression; meanwhile, judging from the rate of compound sentence usage (23.5%) and clause frequency (15.7 clauses per 100 words), the syntactic complexity of film dialogue is lower than formal written language but higher than everyday oral language, demonstrating an artistic oral style <sup>[33]</sup>. In terms of lexical features, the type-token ratio (TTR) of film dialogue is 0.38, lower than literary works (0.45) but higher than daily conversation (0.32), indicating moderate lexical diversity; emotional vocabulary accounts for 13.7%, higher than news texts (7.2%) and academic texts (3.5%), reflecting the importance of emotional expression, as shown in **Figure 2** below.



**Figure 2.** Analysis of Linguistic Features in Film Dialogue Narrative.

Further analysis of dialogue features across different film types reveals significant differences: drama films have a higher proportion of expressive speech acts (24.3%) and relatively complex syntax (average sentence length 9.6 words); suspense films have higher frequencies of interrogative speech acts (18.7%) and rhetorical questions (8.2%); action films have more directive speech acts (29.8%) and shorter sentences (average sentence length 6.3 words). Films with different time period settings also show variations: modern urban-themed films have higher rates of hedges and softeners (10.7% and 9.3% respectively), while historical films have greater nominalization structure and technical terminology density (11.2% and 6.5% respectively). Compared to traditional conversational corpora, film dialogue features higher emotional vocabulary density, stronger rhetorical coloring, and more explicit speech act purposiveness; these characteristics collectively constitute the artistic oral style of film dialogue, providing rich semantic complementary material for shot grammar.

## 4.1.3. Co-Occurrence Patterns of the Two Modalities

This study, through co-occurrence analysis of 6,742 shots and 12,487 dialogue segments from 50 films, reveals the systematic complementary relationship and co-occurrence patterns between shot grammar and dialogue narrative in film storytelling. As shown in Table 3 and Figure 3, the research uses Co-occurrence Coefficient (COC) to quantify the pairing tendencies of features from the two modalities, with COC values ranging from 0-1, where higher values indicate more significant co-occurrence frequencies. The analysis results show that specific shot types and specific speech act types exhibit clear co-occurrence preferences: close-ups and expressive speech acts have the highest co-occurrence coefficient (0.76), indicating that emotional expressions are often presented through character facial close-ups; medium shots are closely associated with assertive speech acts (0.73), suitable for displaying conversations and information transfer between characters; long shots are highly associated with assertive (0.71)and declarative (0.65) speech acts, commonly used for narrative explanations and scene establishment; extreme close-ups and expressive speech acts have a co-occurrence coefficient as high as 0.82, forming a very strong emotional emphasis; while extreme long shots form stable combinations with assertive (0.59) and declarative (0.61)speech acts, frequently used for scene narration and event overview. From a syntactic feature perspective, dialogue in long shots and extreme long shots has longer average sentence lengths (10.3 words and 11.5 words respectively), while dialogue in close-ups and extreme close-ups is more concise (6.7 words and 5.3 words respectively), demonstrating the complementary relationship between shot size and language complexity <sup>[34]</sup>.

Dynamic analysis of co-occurrence patterns further reveals that these patterns exhibit regular changes throughout narrative development: in opening sequences, the cooccurrence frequency of long shots/extreme long shots with assertive speech acts is high (average co-occurrence coefficient 0.68), mainly used to establish narrative environments; in character conflict scenes, co-occurrences of close-ups with expressive/directive speech acts increase (average co-occurrence coefficient 0.64); in emotional cli-

Forum for Linguistic Studies | Volume 07 | Issue 05 | May 2025

| Co-Occurrence Pattern              | Co-Occurrence<br>Coefficient | Average Shot Duration<br>(Seconds) | Average Sentence<br>Length (Words) | Typical Narrative Function                             |
|------------------------------------|------------------------------|------------------------------------|------------------------------------|--|
| Close-up + Expressive              | 0.76                         | 4.2                                | 6.7                                | Emotional emphasis, inner expression                   |
| Close-up + Interrogative           | 0.51                         | 4.3                                | 7.1                                | Emotional tension, psychological conflict              |
| Medium shot + Assertive            | 0.73                         | 5.3                                | 8.4                                | Information transfer, character interaction            |
| Medium shot +<br>Commissive        | 0.52                         | 5.7                                | 8.9                                | Character relationship development, plot turning point |
| Long shot + Assertive              | 0.71                         | 7.2                                | 10.3                               | Scene establishment, background narration              |
| Long shot + Declarative            | 0.65                         | 7.8                                | 11.2                               | Event declaration, narrative framework                 |
| Extreme close-up +<br>Expressive   | 0.82                         | 2.6                                | 5.3                                | Intense emotion, psychological crisis                  |
| Extreme long shot +<br>Assertive   | 0.59                         | 8.9                                | 11.5                               | Scene introduction, environmental description          |
| Extreme long shot +<br>Declarative | 0.61                         | 9.2                                | 12.3                               | Narrative summary, thematic interpretation             |
|                                    |                              |                                    |                                    |  |

| Table 3. Major | Co-occurrence | Patterns | Between | Shot | Grammar | and Dia | logue | Narrative |
|----------------|---------------|----------|---------|------|---------|---------|-------|-----------|
| -/             |               |          |         |      |         |         | -     |           |



Based on analysis of 6,742 shots and 12,487 dialogue segments

Figure 3. Analysis of Co-occurrence Patterns Between Film Shot Grammar and Dialogue Narrative.

max parts, the co-occurrence of extreme close-ups and expressive speech acts reaches its peak (0.87); while in ending sections, there is a tendency toward combinations of long shots/extreme long shots with declarative speech acts (average co-occurrence coefficient 0.73). Different film genres also demonstrate unique co-occurrence patterns: in dramas, co-occurrences of close-ups and expressive speech acts are more frequent (0.81); in suspense films, the pairing of close-ups and interrogative speech acts is significant (0.67); action films more frequently use combinations of medium shots and directive speech acts (0.58). These cooccurrence patterns not only reveal the systematic features of modal complementarity in film narrative but also provide an empirical foundation for understanding how different modalities work together to achieve specific narrative functions. By identifying these typical co-occurrence patterns, a structured analytical framework for film multimodal narrative can be established, laying the groundwork for subsequent research into the deep complementary mechanisms between shots and dialogue.

#### 4.2. Core Mechanisms and Patterns of Modal Complementarity

#### 4.2.1. Semantic Enhancement Complementarity

Semantic enhancement complementarity is the most common complementary mechanism in film multimodal narrative, accounting for 58.3% of all modal complementarity instances in the corpus. The core feature of this mechanism is that one modality (shot or dialogue) strengthens, highlights, or deepens the semantic content conveyed by the other modality, thereby improving the expressive effect. Through analysis of 3,926 semantic enhancement complementarity instances, three main enhancement types were identified: emotional enhancement (43.5%), character enhancement (32.8%), and plot enhancement (23.7%). As shown in Table 4, emotional enhancement complementarity is primarily achieved through combinations of closeup and extreme close-up shots (accounting for 67.4% of this type of complementarity) with expressive speech acts (accounting for 56.2%), with a mean enhancement

Forum for Linguistic Studies | Volume 07 | Issue 05 | May 2025

| Enhancement Type         | Quantity<br>(Proportion) | Main Shot Types   | Main Speech Acts   | Mean Enhancement<br>Intensity | Typical Examples  |
|--------------------------|--------------------------|---|--|-------------------------------|---|
| Emotional<br>Enhancement | 1,708 (43.5%)            | Close-up (38.2%),<br>Extreme close-up<br>(29.2%), Medium shot<br>(22.5%), Long shot<br>(7.6%), Extreme long<br>shot (2.5%)  | Expressive (56.2%),<br>Interrogative (18.3%),<br>Directive (14.5%),<br>Assertive (9.2%), Others<br>(1.8%)  | 0.72                          | Character inner monologue<br>with facial close-up; extreme<br>close-up with intense<br>dialogue during emotional<br>outbursts |
| Character<br>Enhancement | 1,288 (32.8%)            | Medium shot (48.6%),<br>Close-up (27.3%),<br>Long shot (14.2%),<br>Extreme close-up<br>(6.5%), Extreme long<br>shot (3.4%)  | Assertive (36.7%),<br>Directive (28.6%),<br>Expressive (19.3%),<br>Interrogative (12.8%),<br>Others (2.6%) | 0.66                          | Close-up of gestures during<br>character dialogue; medium<br>shot of power posture with<br>commanding tone                    |
| Plot Enhancement         | 930 (23.7%)              | Long shot (31.5%),<br>Medium shot (25.8%),<br>Extreme long shot<br>(21.2%), Close-up<br>(18.4%), Extreme<br>close-up (3.1%) | Assertive (61.8%),<br>Declarative (14.5%),<br>Directive (12.3%),<br>Interrogative (8.7%),<br>Others (2.7%) | 0.68                          | Extreme long shot with<br>narration during scene<br>transitions; close-up of key<br>objects with explanatory<br>dialogue      |

| Table 4. Type and Feature | Distribution of | f Semantic En | hancement Comp | lementarity |
|---------------------------|-----------------|---------------|----------------|-------------|
|---------------------------|-----------------|---------------|----------------|-------------|

intensity of 0.72 (on a 0–1 scale). Character enhancement complementarity tends to use combinations of medium shots (48.6%) with assertive/directive speech acts (together accounting for 65.3%), with a mean enhancement intensity of 0.66; plot enhancement complementarity mainly relies on combinations of long shots/extreme long shots (together accounting for 52.7%) with assertive speech acts (accounting for 61.8%), with a mean enhancement intensity of 0.68 <sup>[35]</sup>. Analysis indicates that the enhancement effect of shot grammar on dialogue semantics (average intensity 0.71) is slightly higher than the enhancement effect of dialogue on shot semantics (average intensity 0.65), which is consistent with film's characteristic as a visually dominant medium.

Significant differences exist in the distribution and intensity of semantic enhancement complementarity across different film genres (as shown in **Figure 4**): in drama films, emotional enhancement complementarity is most prominent, with a mean intensity of 0.82, primarily achieved through nuanced facial expression close-ups with in-depth dialogue; in suspense films, plot enhancement complementarity is more significant (intensity 0.74), especially through the clever combination of visual cues and dialogue hints to create suspense; action films are relatively balanced across the three enhancement types, but plot enhancement slightly predominates (intensity 0.71); in science fiction films, character enhancement and emotional enhancement complementarity are more prominent

(intensities of 0.69 and 0.67 respectively); romance films primarily feature emotional enhancement complementarity (intensity 0.78). Further cultural comparative analysis shows that Western films tend to use more visual enhancement (accounting for 65.3%), while Asian films rely more on dialogue enhancement (accounting for 57.8%). Temporal analysis indicates that semantic enhancement complementarity exhibits regular changes throughout film narrative development: plot enhancement dominates in opening sequences (56.3%), establishing narrative background; during plot development stages, the proportion of character enhancement increases (reaching 48.2%); in emotional climax sections, the proportion of emotional enhancement reaches its peak (68.7%); in ending sections, there is a return to plot enhancement dominance (52.5%)<sup>[36].</sup> These findings reveal that semantic enhancement complementarity, as a fundamental mechanism of film multimodal narrative, has an internal structure and functions closely related to film genre, cultural background, and narrative stage, providing a systematic explanation for understanding the collaborative mechanism of film audiovisual language.

especially through the clever combination of visual cues and dialogue hints to create suspense; action films are relatively balanced across the three enhancement types, but plot enhancement slightly predominates (intensity 0.71); beer on the prison roof: When Andy says "At this moin science fiction films, character enhancement and emotional enhancement complementarity are more prominent from a high angle shot of the prisoners to a low angle shot of the blue sky forms a typical emotion-enhancing complementarity, visually reinforcing the emotional core of "freedom" expressed in the dialogue. Meanwhile, in "Tokyo Story," the combination of plain dialogue with deliberately slowed-paced shots of daily life demonstrates an understated character-enhancing complementarity, deepening the core theme of family alienation through seemingly ordinary audiovisual combinations.



Figure 4. Analysis of Semantic Enhancement Complementarity Patterns.

#### 4.2.2. Semantic Expansion Complementarity

Semantic expansion complementarity is the second major complementary mechanism in film multimodal narrative, accounting for 27.5% (1,853 cases) of all modal complementarity instances in the corpus. The core feature of this mechanism is that one modality provides additional semantic information not expressed or not fully expressible by the other modality, thereby extending, supplementing narrative content, and enriching the meaning construction of the film. Through in-depth analysis, three main types of semantic expansion were identified: environmental expansion (38.4%), temporal expansion (35.7%), and psychological expansion (25.9%). As shown in Table 5, environmental expansion complementarity is primarily achieved through combinations of long shots/extreme long shots (together accounting for 63.5% of this type of complementarity) with concise dialogue, where the shots display spatial information and environmental details far beyond the dialogue description, with a mean expansion coefficient of 0.64 (on a 0–1 scale)<sup>[37]</sup>; temporal expansion complementarity is achieved through techniques such as montage, rapid editing combined with brief dialogue, compressing or extending timelines within limited dialogue, with a mean expansion coefficient of 0.57; psychological expansion complementarity mainly relies on combinations of close-ups/extreme close-ups (together accounting for 68.7%) with implicit dialogue, revealing psychological states not explicitly stated in the dialogue through visual elements such as expressions and eye movements, with a mean expansion coefficient of 0.71, the highest among the three types.

| Expansion Type             | Quantity<br>(Proportion) | Main Shot Techniques   | Main Dialogue Features   | Mean<br>Expansion<br>Coefficient | Typical Examples  |
|----------------------------|--------------------------|--|--|----------------------------------|---|
| Environmental<br>Expansion | 711 (38.4%)              | Long shot (37.2%),<br>Extreme long shot<br>(26.3%), Aerial shot<br>(14.6%), Pan (12.8%),<br>Others (9.1%)  | Concise narration<br>(54.3%), Silence (23.5%),<br>Indicative comments<br>(14.7%), Others (7.5%)        | 0.64                             | Simple dialogue "We're here"<br>with spectacular panorama;<br>environmental long shots inserted<br>in single dialogue scenes  |
| Temporal<br>Expansion      | 661 (35.7%)              | Montage (31.8%), Rapid<br>editing (24.5%), Fade in/<br>out (18.3%), Slow motion<br>(15.6%), Others (9.8%)  | Voice-over (35.2%),<br>Brief summary (28.7%),<br>Transitional phrases<br>(21.5%), Others (14.6%)       | 0.57                             | Short phrase "Years passed" with<br>montage showing passage of time;<br>single line of dialogue spanning<br>shot combinations across multiple<br>time points              |
| Psychological<br>Expansion | 481 (25.9%)              | Close-up (41.4%),<br>Extreme close-up<br>(27.3%), Subjective POV<br>(18.5%), Zoom (7.8%),<br>Others (5.0%) | Implicit expression<br>(39.6%), Silence (29.8%),<br>Contradictory statements<br>(21.3%), Others (9.3%) | 0.71                             | Flat dialogue "It's fine" with<br>facial close-up showing<br>complex emotions; audiovisual<br>combinations showing characters<br>saying one thing but thinking<br>another |

Table 5. Type and Feature Distribution of Semantic Expansion Complementarity.

As shown in Figure 5, semantic expansion complementarity exhibits clear patterns of change throughout film narrative development: temporal expansion dominates in opening sequences (50%), effectively compressing background information; during plot development stages, the proportion of environmental expansion increases (45%), enriching narrative scenes; in emotional climax sections, the proportion of psychological expansion reaches its peak (54%), deepening character interiority; in ending stages, environmental expansion again takes the lead (50%), expanding narrative vision. Significant differences also exist in expansion patterns across different film genres: psychological expansion has the highest proportion in drama films (42.3%); environmental expansion and psychological expansion are relatively balanced in suspense films (38.7% and 36.5% respectively); action films primarily feature environmental expansion (53.2%); science fiction films have an even higher proportion of environmental expansion (58.6%); while historical films favor temporal expansion (46.8%). Further analysis also reveals that the audiovisual relationship in semantic expansion complementarity has obvious asymmetry: in environmental expansion complementarity, cases where the visual modality expands the linguistic modality constitute the vast majority (87.3%); in temporal expansion complementarity, the two modalities are relatively balanced (visual expansion of language accounts for 54.2%, linguistic expansion of visuals accounts for 45.8%); while in psychological expansion complementarity, visual expansion of language still predominates (68.4%)<sup>[38]</sup>. These findings reveal how semantic expansion complementarity, as an important means of film narrative, creates complex meanings beyond the expressive capacity of a single modality through the complementarity of different modalities, providing a systematic framework for understanding the deep mechanisms of film multimodal narrative.

It should be particularly noted that semantic extension complementarity is not simply an increase in information quantity, but rather reflects the unique narrative strategy of film as an audiovisual medium. This complementary mechanism achieves a balance between narrative efficiency and depth through modal division of labor: the linguistic modality typically handles logical relationships, causal connections, and abstract concepts, while the visual modality excels at conveying spatial relationships, emotional atmospheres, and concrete details. In environmentextending complementarity, filmmakers often employ carefully designed 'setting-as-narrative' strategies, making environmental details external projections of characters' psychological states and social situations, such as the vertical spatial composition in Bong Joon-ho's "Parasite" forming a metaphorical social class division that far exceeds what is explicitly stated in dialogue. Time-extending complementarity frequently utilizes 'visual condensation' techniques to convey large amounts of temporal information in a short time, as in "Citizen Kane" where a brief two-minute breakfast montage summarizes the complete deterioration process of a marriage. The most exquisite aspect of psychological extension complementarity lies in its 'dual coding' effect, allowing audiences to simultaneously receive surface speech and deep emotions, such as the complex psychological activities revealed through subtle facial expression changes and focus shifts beneath calm dialogue in "The Doctor." These extension mechanisms are not isolated from each other but often interwoven, as in Tarkovsky's "Mirror," which constructs a complex narrative structure of memory and reality intermingling through environment-time dual extension complementarity with intersecting time and space. The core value of semantic extension complementarity lies in creating a dialectical unity of information density and artistic expressiveness, enabling film to realize infinite narrative possibilities within limited time and space.



Figure 5. Analysis of Semantic Expansion Complementarity Patterns.

#### 4.2.3. Semantic Contrast Complementarity

Semantic contrast complementarity is the most com-

plex complementary mechanism in film multimodal narrative, accounting for 14.2% (963 cases) of all modal complementarity instances in the corpus. The core feature of this mechanism is that shot grammar and dialogue narrative present a surface contradiction or tension relationship, creating complex semantic effects through audiovisual conflict. Although this type of complementarity is less numerous than the previous two types, it possesses unique value in terms of artistic expressiveness and narrative depth. Through detailed analysis of all semantic contrast complementarity cases, three main types were identified: dramatic contrast (36.5%), satirical contrast (31.8%), and ironic contrast (31.7%). As shown in Table 6, dramatic contrast primarily forms a strong contrast between the cruel reality presented in shots and the ideals/hopes expressed in dialogue, often used in scenes where characters face desperate situations or moral dilemmas, with a mean contrast intensity of 0.68 (on a 0-1 scale); satirical contrast typically creates contrast between light/pleasant visual presentation and heavy/dark linguistic content, creating black humor effects, with a mean contrast intensity of 0.64 [39]; ironic contrast reveals contradictions between the subtext suggested by shots and the explicit dialogue of characters, exposing inconsistencies between characters' words and actions or internal conflicts, with a mean contrast intensity as high as 0.75, the strongest among the three types.

To specifically illustrate different types of semantic contrast complementarity, the following analysis examines

typical scenes from works by renowned directors. Stanley Kubrick's "A Clockwork Orange" provides a classic case of ironic contrast: in the scene where the protagonist Alex sings "Singin' in the Rain" while committing violent acts, the cheerful, lighthearted music and lyrics form a stark contrast with the brutal imagery, creating a disturbing black humor effect with a contrast intensity of 0.82, one of the highest in our corpus. This ironic contrast, through deliberately misaligned audiovisual combinations, reveals the absurdity of violence itself and the failure of social conditioning. Lars von Trier's "Melancholia" demonstrates a typical dramatic contrast: in the opening sequence, the close-up of protagonist Justine's smiling face sharply contrasts with her prophecy about the impending destruction of the world, with the slow camera movement conflicting with yet deepening the apocalyptic premonition in the dialogue, forming a dramatic complementarity with a contrast intensity of 0.78. In Hou Hsiao-hsien's "A City of Sadness," the calm narration of Taiwan's historical turmoil forms an ironic contrast (intensity 0.71) with the precise visual presentation of details from characters' daily lives, revealing the implicit infiltration of history into personal fates through seemingly disconnected audiovisual relationships. These three cases represent characteristic expressions of semantic contrast complementarity across different cultural backgrounds, and also confirm how modal complementarity constructs richer cinematic meanings through complex audiovisual relationships.

| Contrast<br>Type      | Quantity<br>(Proportion) | Main Shot Features   | Main Dialogue Features   | Mean<br>Contrast<br>Intensity | Typical Examples   |
|-----------------------|--------------------------|--|--|-------------------------------|--|
| Dramatic<br>Contrast  | 352 (36.5%)              | Close-ups of cruel reality (31.2%),<br>High contrast lighting (28.6%),<br>Long shots emphasizing loneliness<br>(17.3%), Perspective contrast<br>(14.5%), Others (8.4%) | Idealized statements<br>(38.4%), Expressions of<br>hope (31.7%), Discussions<br>of fate (17.9%), Others<br>(12.0%) | 0.68                          | Death hints shown while<br>character discusses future<br>hopes; shots revealing<br>breakdown as character claims<br>"everything is fine" |
| Satirical<br>Contrast | 306 (31.8%)              | Bright, light-hearted scenes<br>(33.7%), Conventional shot style<br>(27.3%), Static composition<br>(21.5%), Everyday settings<br>(10.8%), Others (6.7%)                | Dark content (41.8%),<br>Violent descriptions<br>(28.3%), Threatening<br>expressions (17.6%),<br>Others (12.3%)    | 0.64                          | Discussing horrific acts<br>with light music and bright<br>environments; extreme dialogue<br>paired with mundane family<br>scenes        |
| Ironic<br>Contrast    | 305 (31.7%)              | Expression close-ups revealing<br>truth (42.6%), Detail hints (26.2%),<br>Angle switching (15.4%), Depth of<br>field changes (8.5%), Others (7.3%)                     | Surface statements<br>(37.5%), Disguised<br>expressions (31.8%),<br>Double entendres (21.4%),<br>Others (9.3%)     | 0.75                          | Shots revealing disgust as<br>character expresses love;<br>visual hints of betrayal during<br>expressions of loyalty                     |

Table 6. Type and Feature Distribution of Semantic Contrast Complementarity.

300

As shown in **Figure 6**, significant differences exist in the distribution of semantic contrast complementarity across different film genres: dramatic contrast is most prominent in drama films (43%), strengthening character situations and fate themes through audiovisual contrast; suspense films have higher proportions of ironic contrast (45%) and satirical contrast (40%), effectively creating suspense and psychological tension; satirical contrast reaches its peak in comedy films (45%), serving as an important source of humor effects; science fiction films have a higher proportion of ironic contrast (40%), often used to explore complex relationships between humans and technology; action films show a more balanced distribution among contrast types, but with a lower overall proportion than other genres. Cultural comparative analysis shows that Western films tend to use more satirical contrast (percentage higher than Asian films by 8.5 percentage points), while Asian films more frequently employ dramatic contrast (percentage higher than Western films by 7.2 percentage points). Temporal analysis indicates that semantic contrast complementarity exhibits distinct phase characteristics throughout film narrative development: contrast instances are relatively few in opening sections; the proportion of ironic contrast begins to rise during plot development stages; dramatic contrast reaches its peak at turning points and crisis sections; while ending sections more frequently use satirical contrast, providing reflective space for the narrative <sup>[40]</sup>. Additionally, directorial style analysis reveals that semantic contrast complementarity is an important marker of art films and auteur films, appearing with particularly high frequency in the works of directors such as Kubrick, von Trier, and Bergman compared to commercial films. These findings not only elucidate the diverse manifestations of



Figure 6. Analysis of Semantic Contrast Complementarity Patterns.

semantic contrast complementarity as an advanced film rhetorical device but also provide a systematic framework for understanding how film creates complex meanings through audiovisual contradictions.

#### 4.3. Modal Complementarity Effects Under Dynamic Semantic Framework

# 4.3.1. Modal Complementarity in Emotional Expression

This study, through analysis of 2,158 emotional expression scenes, reveals the core role and effect patterns of modal complementarity in film emotional expression. As shown in Table 7, under the dynamic semantic framework, the impact of different modal complementarity types on emotional expression is examined from three dimensions: emotional intensity, complexity, and audience response. The data shows that, compared with single-modality expression, modal complementarity significantly enhances emotional expression effects: the average intensity rating for single-modality emotional scenes is 3.2 (on a 5-point scale), while scenes using semantic enhancement complementarity reach 4.3, semantic expansion complementarity scenes achieve 4.1, and semantic contrast complementarity scenes reach as high as 4.6. Particularly in expressing complex emotions, semantic contrast complementarity demonstrates a significant advantage: emotional complexity rating reaches 4.5, far higher than the 2.8 of single modality<sup>[41]</sup>. Through analysis of specific complementary techniques, it is found that enhancement complementarity between close-up shots and expressive dialogue is most effective (efficacy rate 86.3%) when expressing basic emotions (such as sadness, anger, joy); environmental expansion complementarity performs exceptionally well in creating emotional atmosphere (efficacy rate 79.5%); while ironic contrast is most effective in expressing contradictory emotions (such as love-hate intertwining, bittersweet feelings) (efficacy rate 91.7%).

As shown in **Figure 7**, with increasing emotional complexity, different modal complementarity types display marked differences in emotional intensity performance: when expressing simple emotions (complexity 1–2), the effect differences among the three complementarity types are not significant; however, when expressing high-complexity

Forum for Linguistic Studies | Volume 07 | Issue 05 | May 2025

| Complementarity<br>Type | Mean Emotional<br>Intensity (1–5) | Mean Emotional<br>Complexity (1–5) | Mean Audience<br>Rating (1–5) | Main Emotional Types   | Typical Application<br>Scenarios                               |
|-------------------------|-----------------------------------|------------------------------------|-------------------------------|--|--|
| Single Modality         | 3.2                               | 2.8                                | 3.4                           | Basic emotions (83.6%)   | Simple emotional states, clear emotional expression            |
| Semantic Enhancement    | 4.3                               | 3.7                                | 4.2                           | Intensified emotions (67.4%),<br>Deep emotions (32.6%)           | Emotional climaxes,<br>important turning points                |
| Semantic Expansion      | 4.1                               | 4.2                                | 4.5                           | Mixed emotions (58.3%),<br>Implicit emotions (41.7%)             | Subtle emotional changes,<br>emotional transition<br>processes |
| Semantic Contrast       | 4.6                               | 4.5                                | 4.7                           | Contradictory emotions<br>(76.2%), Complex<br>psychology (23.8%) | Internal conflicts, identity<br>crises, ethical dilemmas       |

Table 7. Characteristics and Effects of Different Modal Complementarity Types in Emotional Expression.



Figure 7. Analysis of Modal Complementarity Effects in Emotional Expression.

emotions (complexity 4-5), the emotional intensity of semantic contrast complementarity is significantly higher than other types, especially when emotional complexity reaches the highest level (5 points), where the emotional intensity of contrast complementarity (7.6) far exceeds semantic enhancement (5.7) and single modality (3.9). Further genre analysis shows that different film types prefer different emotional expression complementarity patterns: drama films tend to use semantic enhancement complementarity (accounting for 42.8%) to strengthen core emotions; suspense films more frequently employ semantic expansion complementarity (accounting for 38.5%) to create emotional suspense; while comedy films have the highest frequency of semantic contrast complementarity usage (accounting for 36.2%), effectively producing emotional contrast and humor effects. Cultural comparative analysis also finds that Western films are more inclined to express clear emotions through direct semantic enhancement comple- scale), far higher than the 4.8 of single modality; regarding

mentarity (percentage higher than Asian films by 9.3 percentage points), while Asian films more frequently adopt implicit semantic expansion complementarity (percentage higher than Western films by 8.7 percentage points) [42]. From a cognitive perspective, modal complementarity significantly increases audience emotional investment and empathetic experience by activating multiple perceptual channels, an effect particularly evident in high-complexity emotional expression. Overall, the research indicates that modal complementarity mechanisms not only enhance the intensity of emotional expression but also effectively convey complex, subtle, and contradictory emotional states, providing rich expressive means for film emotional narrative.

### 4.3.2. Modal Complementarity in Character Construction

This study conducted a detailed analysis of 1,876 character portrayal scenes in the corpus, revealing the key role and effect mechanisms of modal complementarity in film character construction. As shown in Table 8, the impact of different modal complementarity types on character portrayal was evaluated across six character dimensions (psychological depth, moral complexity, relationship dynamics, social identity, historical background, and cultural context). The data shows that, compared with single modality, modal complementarity significantly enhances the multi-dimensional presentation of characters: in terms of psychological depth, characters portrayed using semantic contrast complementarity scored 7.6 (on a 10-point moral complexity, semantic contrast complementarity also performed best (7.3 points), with semantic enhancement complementarity ranking second (5.9 points); in displaying relationship dynamics, all three complementary methods significantly outperformed single modality, though dominant complementary patterns differed across different film types <sup>[43]</sup>. Through correlation analysis, specific shot-dialogue combinations were found to have significant effects on character portrayal: enhancement complementarity between close-ups/extreme close-ups and inner monologues proved most effective in revealing characters' inner worlds (correlation coefficient r = 0.73); environmental expansion complementarity effectively established relationships between characters and social environments through combinations of long shots and minimal dialogue (r = 0.68); while ironic contrast, through contradictory presentations of expression close-ups and surface dialogue, best demonstrated character complexity and internal-external conflicts (r = 0.81).

Different types of modal complementarity demonstrate distinct advantages across the six character dimensions: semantic contrast complementarity is most prominent in shaping psychological depth and moral complexity; semantic expansion complementarity performs best in presenting cultural context and historical background; while semantic enhancement complementarity has slight advantages in character relationship dynamics and social identity. From character type analysis, main characters benefit more from semantic contrast complementarity (accounting for 48.3%), helping to shape complex, three-dimensional character images; while secondary characters more frequently employ semantic enhancement complementarity (accounting for 52.7%), effectively reinforcing their typical features within limited shots. From narrative position perspective, character first appearance scenes tend to use semantic expansion complementarity (accounting for 57.6%), quickly establishing character positioning through complementary relationships between environment and character; character transformation nodes more frequently use semantic contrast complementarity (accounting for 63.8%), highlighting internal changes through audiovisual contrast; while key decision scenes for characters predominantly feature semantic enhancement complementarity (accounting for 55.2%), reinforcing character motivation and will <sup>[44]</sup>. Additionally, gender analysis shows that semantic contrast complementarity is used significantly more frequently in portraying female characters (38.7%) than male characters (29.5%), a phenomenon reflected in films across different cultural backgrounds, reflecting common patterns in film character gender presentation. These findings indicate that modal complementarity is not only an effective means for in-depth character portrayal but also an important strategy for film narrative to construct character complexity, relationship networks, and sociocultural identities, providing a systematic explanation for understanding the audiovisual mechanisms of film character construction.

| Character<br>Dimension   | Single<br>Modality | Semantic<br>Enhancement<br>Complementarity | Semantic Expansion<br>Complementarity | Semantic Contrast<br>Complementarity | Best Complementary Technique                               |
|--------------------------|--------------------|--|---------------------------------------|--------------------------------------|--|
| Psychological<br>Depth   | 4.8                | 6.0  | 5.8                                   | 7.6                                  | Extreme close-up + inner monologue (contrast type)         |
| Moral<br>Complexity      | 4.3                | 5.9  | 5.3                                   | 7.3                                  | Behavior close-up + opposing<br>dialogue (contrast type)   |
| Relationship<br>Dynamics | 4.7                | 6.1  | 6.3                                   | 6.9                                  | Medium shot + interactive dialogue<br>(expansion type)     |
| Social Identity          | 5.0                | 6.2  | 5.8                                   | 6.1                                  | Long shot + identity hints<br>(enhancement type)           |
| Historical<br>Background | 3.7                | 5.4  | 6.2                                   | 5.7                                  | Montage + voice-over narration<br>(expansion type)         |
| Cultural Context         | 3.5                | 5.2  | 6.3                                   | 5.8                                  | Environmental close-up + cultural symbols (expansion type) |
| Average Score            | 4.2                | 6.0  | 5.8                                   | 7.5                                  | -  |

Table 8. Effect Ratings of Different Modal Complementarity Types in Character Construction (10-point scale).

# 4.3.3. Modal Complementarity in Narrative Progression

This study, through analyzing 2,347 narrative progression scenes in the corpus, reveals the distribution patterns and functional mechanisms of modal complementarity across different stages of film narrative development. As shown in Table 9, film narrative structure is divided into four stages: exposition, rising action, climax, and resolution, examining the application frequency and effects of different modal complementarity types at each stage <sup>[45]</sup>. The data shows that different narrative stages exhibit distinct modal complementarity preferences: the exposition stage is dominated by semantic expansion complementarity (52%), effectively establishing narrative background through environmental expansion and temporal expansion, providing basic information necessary for audience understanding of the story; the rising action stage primarily features semantic enhancement complementarity (40%), strengthening conflicts and character development, driving the plot forward; semantic contrast complementarity takes the dominant position in the climax stage (53%), creating maximum dramatic tension through audiovisual conflict; while the resolution stage returns to semantic expansion complementarity (45%), extending narrative boundaries through a macro perspective, suggesting broader meanings of the story [46]. Detailed analysis of the narrative functions of each complementarity type reveals that in the exposition stage, environmental expansion complementarity scores highest in narrative effectiveness (7.6/10), significantly improving background establishment efficiency; at plot turning points, the suspense effect created by semantic contrast

complementarity scores 8.2, far higher than single modality (5.3); in foreshadowing scenes, semantic expansion complementarity performs most prominently (hint validity 7.8).

As shown in Figure 8, modal complementarity types exhibit clear fluctuation patterns in narrative development, with these patterns showing common features across different film genres, though genre differences also exist: suspense films have a significantly higher proportion of contrast complementarity in the rising action stage than the average level (+12.6 percentage points), effectively creating suspense; action films have a higher proportion of enhancement complementarity in the climax stage (+9.3 percentage points), reinforcing the impact of action scenes; while drama films have a higher proportion of expansion complementarity in the resolution stage (+8.5 percentage points), deepening thematic significance [47]. Further analysis also reveals that modal complementarity demonstrates significant advantages in different narrative functions: in transition and spatiotemporal jump scenes, the connection clarity rating of semantic expansion complementarity (7.8) is significantly higher than single modality (5.6); in establishing causal relationships, semantic enhancement complementarity improves logical clarity by 38.5%; in foreshadowing and callback structures, the correspondence of modal complementarity across different stages increases narrative coherence ratings by 42.3%. Additionally, audience comprehension tests reveal that key scenes employing appropriate modal complementarity increase audience plot understanding accuracy by an average of 25.7% and narrative engagement by 31.6%, an effect particularly significant in films with complex narrative structures (+43.2%).

| Narrative<br>Stage | Semantic<br>Enhancement<br>(%) | Semantic<br>Expansion (%) | Semantic<br>Contrast (%) | Dominant<br>Complementarity Type     | Main Narrative<br>Functions                    | Effectiveness<br>Rating (1–10) |
|--------------------|--------------------------------|---------------------------|--------------------------|--------------------------------------|--|--------------------------------|
| Exposition         | 44                             | 52                        | 20                       | Expansion-Environmental expansion    | Background<br>establishment, World<br>building | 7.6                            |
| Rising<br>Action   | 40                             | 35                        | 30                       | Enhancement-Emotional<br>enhancement | Conflict development,<br>Character growth      | 7.2                            |
| Climax             | 48                             | 30                        | 53                       | Contrast-Dramatic contrast           | Tension creation, Climax intensification       | 8.2                            |
| Resolution         | 35                             | 45                        | 25                       | Expansion-Temporal expansion         | Meaning extension,<br>Thematic elevation       | 7.4                            |

Table 9. Distribution and Functional Characteristics of Modal Complementarity Across Different Narrative Stages.



Figure 8. Analysis of Modal Complementarity Effects in Narrative Progression.

These findings not only reveal modal complementarity as a core mechanism for film narrative progression but also provide an empirical foundation for understanding how film constructs complex narratives through systematic organization of audiovisual language, while offering theoretical guidance for film screenwriters and directors in selecting appropriate audiovisual expression strategies for different narrative stages <sup>[48]</sup>.

#### 5. Discussion

## 5.1. Theoretical Significance of Research Findings

This systematic exploration of modal complementary mechanisms between film shot grammar and dialogue narrative provides important theoretical contributions to fields such as multimodal discourse analysis, film linguistics, and dynamic semantic theory. The constructed dynamic semantic framework successfully integrates the multimodal analysis methods of systemic functional linguistics, the conceptual integration theory of cognitive linguistics, and the semantic modeling techniques of computational linguistics, providing a new analytical paradigm for studying complex multimodal texts such as films. This framework not only overcomes the limitations of modal separation in traditional multimodal analysis but also transcends the

constraints of static semantic views. By treating semantic complementarity as a dynamic process, it reveals how different modalities collaboratively construct meaning in narrative development [49]. In particular, the identification of three core complementary mechanisms (semantic enhancement, semantic expansion, and semantic contrast complementarity) and their subtypes provides a theoretical foundation for the refined classification of inter-modal relationships, improving Bateman and Schmidt's multimodal fusion theory and Wildfeuer's film discourse semantics. These findings challenge the traditional linguistic view that treats language as central and visual elements as auxiliary, demonstrating that in film narrative, there exists a complex interactive relationship between audiovisual modalities, with each being irreplaceable and jointly constructing meaning.

Additionally, the findings of this study enrich film linguistics and narrative theory, revealing the systematic operational mechanisms of modal complementarity in emotional expression, character construction, and narrative progression. They prove that modal complementarity is not merely a rhetorical device but a fundamental constituent element of film narrative, capable of significantly enhancing emotional intensity, enriching character dimensions, and driving narrative development. In particular, the unique advantages demonstrated by semantic contrast complementarity in handling complex emotions and contradictory psychology provide a new perspective for understanding how film conveys subtle and deep meanings <sup>[50]</sup>. Meanwhile, the identified patterns of modal complementarity preferences across different narrative stages also provide a new method for the formal description of film narrative structure. From a cognitive perspective, the research results support multichannel information processing theory, proving that appropriate modal complementarity can reduce audience cognitive load and improve narrative comprehension efficiency. In terms of cross-cultural communication, the discovered cultural difference patterns also provide an empirical foundation for understanding audiovisual language preferences across different cultural backgrounds.

Despite the theoretical contributions achieved in this research, there are several limitations that need to be identified. First, although the sample size reached 50 films, it remains limited relative to the totality of world cinema, and the selection process inevitably favored highly-rated and well-known works, which may have led to neglect of the linguistic features in experimental and marginal films. Second, despite attempts to balance films from different regions, Western and East Asian films still constitute a large proportion (accounting for 72% combined), and the underrepresentation of films from the Middle East, Africa, and Latin America affects the global applicability of conclusions. Methodologically, although various bias mitigation measures were adopted, the annotators' professional backgrounds still primarily came from film studies and linguistics, and the lack of more diversified disciplinary perspectives (such as psychology, sociology) may have resulted in overly singular interpretations of certain modal complementary functions. Additionally, regarding computational models, existing deep learning technologies still have precision limitations when processing highly contextualized and culturally specific audiovisual semantics, especially for metaphorically rich art films. These limitations suggest that we should maintain caution when interpreting research results and also indicate directions for improvement in future research.

Overall, this study not only connects theoretical tic expansion complementarity activates complementary perspectives from linguistics, film studies, and cognitive neural networks, expanding the range of meaning conscience but also paves the way for constructing more comprehensive and dynamic multimodal analysis methods, to temporary cognitive dissonance, triggering broader

with important implications for the future development of multimodal research.

#### 5.2. Cognitive Explanation of Modal Complementarity Mechanisms

From a cognitive linguistics perspective, the modal complementarity mechanisms between film shot grammar and dialogue narrative can be explained through multiple cognitive theories. Based on Conceptual Integration Theory, film viewers automatically establish mental space mappings between visual input and linguistic input, forming an integrated meaning network. The three types of modal complementarity discovered in this research actually correspond to different conceptual integration patterns: semantic enhancement complementarity triggers mirror integration, where inputs from both modalities share similar conceptual structures, mutually reinforcing each other to reduce cognitive load; semantic expansion complementarity involves single-scope integration, where one modality provides an external framework for the other, guiding viewers to extend concepts based on existing information; while semantic contrast complementarity embodies double-scope integration, requiring viewers to coordinate superficially contradictory inputs from two modalities and construct new integrated meanings at a higher level [51]. Experimental data shows that as complementarity complexity increases, viewers' cognitive effort increases, but the richness and depth of meaning extraction also correspondingly improve, especially with semantic contrast complementarity which, despite requiring higher cognitive processing, produces the most significant emotional and narrative comprehension effects (effect ratings improved by 42.3%), conforming to the predictions of "productive cognitive load theory."

Furthermore, from a neurolinguistic perspective, the effects of modal complementarity mechanisms can be explained through multi-channel processing theory. Visual information and linguistic information are processed in the brain via different but interconnected neural networks; semantic enhancement complementarity activates similar semantic areas, producing reinforcement effects; semantic expansion complementarity activates complementary neural networks, expanding the range of meaning construction; while semantic contrast complementarity leads to temporary cognitive dissonance, triggering broader neural network participation in the integration process, which explains why contrast complementarity scores significantly higher in emotional intensity ratings than other types. Data also shows that viewers from different cultural backgrounds demonstrate preference differences when processing certain modal complementarity types, attributable to culturally specific cognitive patterns: Western viewers understand semantic enhancement complementarity faster than Asian viewers (11.2% faster on average), while Asian viewers have stronger abilities to identify implicit meanings in semantic expansion complementarity (13.5% higher accuracy)<sup>[52]</sup>. These findings not only support cultural cognitive difference theories in cognitive linguistics but also confirm film multimodal complementarity as a universal yet culturally sensitive cognitive mechanism. Overall, explaining modal complementarity from a cognitive perspective helps understand why this mechanism plays such an important role in film narrative-it not only conforms to the information processing tendencies of the human brain but can also create richer and more profound viewing experiences by mobilizing multiple cognitive resources.

## 6. Conclusion and Outlook

#### 6.1. Main Research Conclusions

Based on multimodal corpus analysis of 50 internationally awarded films, this study systematically examined the modal complementary mechanisms between film shot grammar and dialogue narrative, reaching the following five main conclusions:

(1) A systematic complementary relationship exists between shot grammar and dialogue narrative in film; this complementarity is not merely an occasional rhetorical device but a fundamental mechanism for meaning construction in film. The research identified three core complementary types: semantic enhancement (58.3%), semantic expansion (27.5%), and semantic contrast (14.2%), which exhibit stable yet variable distribution patterns across different film genres, directorial styles, and cultural backgrounds.

(2) Modal complementarity demonstrates clear scene dependency and functional differentiation. In emotional expression scenes, semantic contrast complementarity

rating 4.5/5), especially contradictory emotional states; in character construction, semantic contrast complementarity is most effective in shaping character psychological depth and moral complexity (ratings of 7.6 and 7.3/10 respectively), while semantic expansion complementarity performs best in presenting character historical background and cultural context; in narrative progression, different narrative stages exhibit distinct complementary type preferences, with expansion type dominating the exposition stage (52%)and contrast type dominating the climax stage (53%).

(3) Dynamic semantic framework analysis reveals that different modal complementarity types have different cognitive processing mechanisms. Semantic enhancement complementarity triggers mirror integration, reducing cognitive load; semantic expansion complementarity involves single-scope integration, expanding the range of meaning construction; while semantic contrast complementarity embodies double-scope integration, increasing cognitive burden but generating richer meaning.

(4) Modal complementarity in film narrative produces holistic effects transcending individual modalities, especially in significantly enhancing emotional intensity (average increase of 35.7%), character complexity (average increase of 42.8%), and narrative comprehension (improvement of 25.7%), proving that the whole of multimodal coordination is greater than the sum of its parts.

(5) Cultural comparative analysis found that Western films tend to use more semantic enhancement complementarity (9.3 percentage points higher than Asian films) and satirical contrast within semantic contrast complementarity (8.5 percentage points higher), while Asian films more frequently employ semantic expansion complementarity (8.7 percentage points higher) and dramatic contrast (7.2 percentage points higher), reflecting deep differences in audiovisual language usage across different cultures. These conclusions not only enrich the theoretical horizons of multimodal discourse analysis and film linguistics but also provide an empirical foundation for film creation and cross-cultural communication.

#### 6.2. Future Outlook

Based on the findings and limitations of this study, future research on modal complementary mechanisms most effectively conveys complex emotions (complexity between film shot grammar and dialogue narrative can be further deepened in the following five directions:

(1) Expand research corpus and methodology. Future research can incorporate more diverse film samples, including films from different periods, regions, and genres, especially non-mainstream films, experimental films, and new media narrative forms, to test the universality and variability of modal complementary mechanisms. Meanwhile, neuroscientific methods such as eye-tracking and electroencephalography (EEG) can be integrated to directly measure audience real-time cognitive responses to different modal complementarity types, providing a more solid neurophysiological foundation for theoretical explanations.

(2) Deepen cross-cultural comparative research. This study has preliminarily revealed modal complementarity differences across films from different cultural backgrounds; future research can further refine this into studies of modal complementary characteristics within specific cultural traditions (such as Japanese wabi-sabi aesthetics, French New Wave, Indian Bollywood, etc.), exploring deep connections between cultural traditions, aesthetic concepts, and audiovisual expression, establishing a more refined theoretical framework for cross-cultural film linguistics.

(3) Expand the categories of multimodal complementarity. Beyond shot grammar and dialogue narrative, future research can incorporate more modal elements such as music, sound effects, color, and editing rhythm into the analytical framework, constructing a more comprehensive network model of film multimodal complementarity and revealing more complex multi-modal coordination mechanisms.

(4) Develop computational models and application tools. Based on the dynamic semantic framework established in this research, algorithms and tools for automated identification and analysis of film modal complementarity can be developed, combined with deep learning technologies to achieve automatic annotation and analysis of large-scale film corpora, providing new digital humanities methods for film research. These tools can also be applied to practical application scenarios such as film creation assistance, automatic subtitle generation, and cross-media content conversion.

(5) Explore the historical evolution of modal comple-

different historical periods, the historical evolutionary trajectory of modal complementary mechanisms in film language can be examined, analyzing relationships between technological developments (such as the emergence of sound films, the popularization of digital technology), narrative concept innovations, and changes in modal complementarity strategies, constructing a diachronic theory of film language.

(6) Deepening Research on the Socio-Cultural Implications of Modal Complementarity. The cultural differences in modal complementarity discovered in this research are not merely aesthetic manifestations but have profound socio-cultural significance. Future research should focus on how modal complementarity strategies affect international film reception and cross-cultural communication, particularly exploring how Western audiences receive the more implicit semantic extension-type complementarity in Asian films, and how Asian audiences interpret the direct semantic enhancement-type complementarity and ironic contrasts in Western films. These studies can provide a theoretical foundation for cross-cultural localization of films, guiding subtitle translation and dubbing strategies-for example, when facing scenes that heavily rely on culturally specific semantic contrasts, subtitles can compensate for cultural cognitive differences by adding explanatory translations or adjusting expressions. Meanwhile, with the rise of global streaming platforms, research on modal reception preferences of audiences from different cultural backgrounds can provide data support for content recommendation algorithm optimization and regionalized content production. Additionally, attention should be paid to the connection between modal complementarity and the expression of social issues, analyzing how films from different cultural backgrounds use modal complementarity strategies to convey sensitive topics such as gender, ethnicity, and class, thereby revealing the subtle role of film audiovisual language in the construction of social discourse.

This research not only helps understand the historical development of film language but also provides a theoretical foundation for predicting potential evolutions of future audiovisual narrative forms. Overall, research on modal complementarity between film shot grammar and dialogue narrative still has broad development space; through multimentarity. Through systematic comparison of films from disciplinary intersection and methodological innovation, it is expected to achieve richer results in theoretical deepening and application expansion.

# Funding

This work received no external funding.

## **Institutional Review Board Statement**

Not applicable.

# **Informed Consent Statement**

Not applicable.

## **Data Availability Statement**

Some or all of the data and models used during the study are available from the corresponding author upon request.

# **Conflicts of Interest**

The authors declare no conflict of interest.

# References

- [1] Su, W., 2025. Research on Film Visual Metalanguage and "One-take" Shot Design. Journal of Bengbu University. 14(1), 7-11. DOI: https://doi.org/10.13900/ j.cnki.jbc.2025.01.015
- Language. Art Education. 77(11), 192-195.
- [3] James, K., 2020. Planting the Seeds of Artistic Subversiveness in À Bout De Souffle: Godard's Trailblazing Cinematic Language. Studies in Logic, Grammar and Rhetoric. 65(1), 57-70.
- [4] Floriane, B., 2020. Film language, film emotions and the experience of blind and partially sighted viewers: a reception study. Journal of Specialised Translation. (33), 259-280.
- [5] Whelan, J.M., 2020. Sparring Partners and Sampling Platters: Film Language, the Sacred, and Grace in Raging Bull and Babette's Feast. Quarterly Review of Film and Video. 38(3), 190-216. DOI: https://doi. org/10.1080/10509208.2020.1764899
- [6] Miao, X., 2024. Symbiosis and Innovation: Research on the Integration Trend of Modern Dance and Film Art. Film Literature. (21), 116–120.
- [7] He, Y., 2024. Creation of Micro-film Shot Language and Photography Techniques in the Media Convergence Era. News Outpost. (19), 57-58.

- [8] Peng, G., 2024. Poetic Expression of Audiovisual Language and Moving Shots: In-depth Interpretation of Film Art. Charming Hunan. (5), 35-37.
- [9] Xi, Y., 2024. From "Film Screenplay" to "Shot Language"-Translation and Holistic Construction of Early Left-wing Film Screenplay and Editing Theory. Contemporary Cinema. (10), 41-48.
- [10] Zheng, D., Wang, Y., 2024. The Application of Shot Language in Suspense Films-Taking "The Silence of the Lambs" as an Example. Journal of Taizhou Polytechnic College. 24(4), 74-76.
- [11] Chen, L., 2024. Symbolic Representation, Shot Language, Situational Imagination: Research on Narrative Strategies of the Ethnic Minority Film "The Gun of Gunlala". Cultural Innovation Comparative Research. 8(16), 54-58.
- [12] Fan, X., 2024. Stories Behind the Shots: Analysis of Implicit Details in Classic Films. Film Literature. (8), 92-96.
- [13] Tian, Y., Si, G., 2024. Film Language and Film Poetics in Surrealistic Landscapes: Exploration of Satoshi Kon's Animation Narrative and Shot Language. Film Literature. (5), 114–117.
- [14] Chen, X., 2023. The Mortise and Tenon Joint of Art and Technology: Mexican Landscape and Shot Language in Alfonso Cuarón's Films. Film Review. (20), 56-59. DOI: https://doi.org/10.16583/j.cnki.52-1014/ j.2023.20.005
- [15] Sun, L., Liu, Z., 2023. Research on the Application of Photographic Shot Language and Perspective in Theatrical Films. Art Evaluation. (10), 164–167.
- [16] Tang, R., 2023. A Brief Analysis of the Expression of Wine Culture Shot Language in Classical Wuxia Films. China Liquor. (5), 78-79.
- [2] Tao, S., 2024. Analysis of Image Art in Film Shot [17] Yu, M., 2022. When Comprehensive Chinese Language Learning Takes Wing with Animated Films-Developing Comprehensive Chinese Language Learning Curriculum Resources with Animated Films as Medium. Educational Science Forum. (25), 17-22.
  - [18] Yao, T., Wu, J., 2022. Symphony of Medieval Retro and Future Science Fiction: Poetic Expression of Audiovisual Language in "Dune". Southeast Communication. (5), 10–13. DOI: https://doi.org/10.13556/ j.cnki.dncb.cn35-1274/j.2022.05.009
  - [19] Gao, L., 2022. Using Shots to Gaze at Moments of Crisis in Interpersonal Relationships-On the Themes and Audiovisual Language of Miwa Nishikawa's Films. Oriental Art. (2), 105-111.
  - [20] Xu, L., Zhang, N., 2022. "Perhaps Love": A "Suspended" Poetic Dialogue Constructed by Audiovisual Language Symbols. Film Review. (7), 49-52. DOI: https://doi.org/10.16583/j.cnki.52-1014/ j.2022.07.015
  - [21] Shu, N., Yang, X., 2022. Analysis of Film Shot Lan-

guage in Wordless Picture Books—Taking "The Snowman" as an Example. Art Education Research. (5), 116–117.

- [22] Chen, L., 2022. Lyrical Tradition and Shot Language in Early Chinese Films. Social Sciences. (3), 171–178. DOI: https://doi.org/10.13644/j.cnki.cn31-1112.2022.03.014
- [23] Zhou, W., 2021. On Film Shot Language Expression Under Digital Technology. Western Radio and Television. 42(23), 121–123+148.
- [24] Chen, L., 2021. Research on Narrative Differentiation Between Vlog Shot Language and Film Shot Language. Film & TV Production. 27(7), 82–84.
- [25] Huang, B., 2021. Emotional Expression Completed Through Shot Language in the Film "The Piano". Theatre House. (19), 166–167.
- [26] Zeng, Y., Qu, L., 2021. Shot Language Concealment— Film Writing in Kafka's "Amerika". Social Science Front. (4), 158–164.
- [27] Fan, X., 2021. Research on Prose-style Narrative of Shot Language in the Film "My Memories of Old Beijing". Journal of Suihua University. 41(2), 108– 111.
- [28] Yang, Y., 2021. Analysis of Narrative Perspective and Shot Language in the Film "Let the Bullets Fly". Western Radio and Television. 42(3), 151–153.
- [29] Wang, S., Zhi, J., 2021. Analysis of Film Shot Language That Constructs Artistic Conception. Film Literature. (2), 61–63.
- [30] Wang, J., 2021. Changing Shots and Character Relationships: Analysis of Shot Techniques in the Film "Eat Drink Man Woman". Audio-Visual. (1), 97–98. DOI: https://doi.org/10.19395/j.cnki.1674-246x.2021.01.046
- [31] Xu, Z., 2020. Shot Language and Memory Reconstruction: Taking the Vietnamese Film "The Scent of Green Papaya" as an Example. Audio-Visual. (12), 79–80. DOI: https://doi.org/10.19395/j.cnki.1674-246x.2020.12.040
- [32] Jiang, J., 2020. Analysis of Shot Language in the Korean Film "The Terror Live". Audio-Visual. (11), 93–94. DOI: https://doi.org/10.19395/j.cnki.1674-246x.2020.11.048
- [33] Zhong, H., 2020. Monologue Shots and Their Language Strategies in the Film "Wings of Desire". Chinese Film Market. (3), 52–57.
- [34] Song, Y., 2019. Analysis of Shot Language and Narrative Strategies in Digital Dome Science Popularization Films. Science Popularization Research. 14(6), 91–96+116. DOI: https://doi.org/10.19293/ j.cnki.1673-8357.2019.06.010
- [35] Stepanova, P.M., 2020. "Body techniques" as expressive means in the formation of film language in early Soviet cinema. Studies in Russian and Soviet Cinema. 14(2), 124–139. DOI: https://doi.org/10.1080/1

7503132.2020.1749469

- [36] Floriane, B., 2020. Audio description and the translation of film language into words. Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies. 73(1), 273–296.
- [37] Cooper, A., 2020. Kinolab: A platform for the digital analysis of film language in narrative film and media. Journal of Italian Cinema & Media Studies. 8(1), 95–98.
- [38] Kirsch, G., 2019. POST-SCRIPT FROM FILM STUDIES: Whose choice? Watching non-English language films in the UK. Multilingua. 38(5), 619– 624. DOI: https://doi.org/10.1515/multi-2018-0118
- [39] Allison, C., 2019. Kinolab: A Digital Humanities Project for the Analysis of Film Language. Italian Culture. 37(2), 137–143.
- [40] Shuneyko, A., Chibisova, O., 2019. The Mechanism of Contextual Confirmation of the Symbol in the Cinema Language. Media Education (Mediaobrazovanie). 59(1), 123–134. DOI: https://doi. org/10.13187/me.2019.1.12
- [41] Petr, K., 2019. Literary language vs. film speech. In search of new views on war and totalitarianism in Czech culture of the 1960s. World Literature Studies. 11(3), 5–16.
- [42] Shafaei, M., Samghabadi, S.N., Kar, S., et al., 2019. Rating for Parents: Predicting Children Suitability Rating for Movies Based on Language of the Movies. Available from: https://arxiv.org/abs/1908.07819
- [43] Bahloul, M., 2018. Arabic Language in the Emirati Films: Linguistic and Cultural Window on Emirati Films. Available from: https://www.jstor.org/stable/26578124
- [44] Lorenzo-Blanco E.I., Abad-Vivero E.N., Barrientos-Gutierrez I., et al., 2018. Movie Language Orientation, Gender, Movie Smoking Exposure, and Smoking Susceptibility among Youth in Mexico. Nicotine & Tobacco Research. 20(11), 1378–1385.
- [45] Cooke, P., Dennison, S., Gould, W., 2018. The Voicing Hidden Histories project: participatory video in development, soft power and film language. Media Practice and Education. 19(3), 270–282.
- [46] Diabate, N., 2017. The Cinematic Language of Naked Protest. Critical Interventions. 11(3), 248–268.
- [47] Anonymous, 2024. India's Jindal Poly Films to build BOPP film line. Chemical Week. 186(19), 16.
- [48] Joseph, R.D., 2024. Follow These Tips to Fire Up Your Cold-Blown Film Line. Plastics Technology. 70(5), 32–34.
- [49] Anonymous, 2018. W&H's TURBOCLEAN Module Now Available for Retrofit on Existing Blown Film Lines. Flexible Packaging. 21(2), 44.
- [50] Howell, L.E., Behrman, L.S., Kirschner, E., et al., 2025. Storytelling in Science Film: Narrative Engagement Relates to Greater Knowledge, Interest,

cation. 47(2), 211-249.

- [51] Waliaula, S., 2025. Cinema Narration as Oral Performance: DJ Afro and East African Media Practices. Journal of African Cultural Studies. 37(1), 6–21.
- and Identification With Science. Science Communi- [52] Liu, S., Gao, J., Zhang, G., et al., 2024. How to speak destinations more efficiently: a study on the effects of narration in tourism films on viewer perceptions and intentions. Asia Pacific Journal of Tourism Research. 29(12), 1509-1528.