

Constructing evaluative meaning across modes: a case study of *Pon tu pensamiento en mí*

Bin Dang^{1,2}

¹China Shaanxi Translators Association, Xi'an, China

²Universidad Central "Marta Abreu" de Las Villas (Central University 'Marta Abreu' of Las Villas), Santa Clara, Cuba

bd26198@foxmail.com

Abstract. This study investigates the multimodal construction of evaluative meaning in Cuban fiction film, focusing on Arturo Sotro's *Pon tu pensamiento en mí* (1995). Grounded in the theoretical integration of multimodal discourse analysis and Relevance Theory's echoic account of interpretive meaning, the research addresses a central question: how do linguistic, prosodic, kinesic, and cinematic modes coalesce to construct layered evaluative meanings that guide viewer interpretation without explicit linguistic marking? Employing ELAN annotation software for systematic multimodal transcription, the study analyzes six key scenes from the film, coding verbal content, prosodic features, facial expressions, gestures, gaze patterns, camera movements, and editing rhythms. The findings reveal that evaluative meaning in this film emerges through systematic multimodal configurations rather than through any single semiotic channel. Three primary configuration patterns are identified: (1) modal convergence, wherein all modes align to reinforce a consistent evaluative stance; (2) modal dissonance, wherein conflicting cues create interpretive tension that invites active viewer engagement; and (3) modal hierarchy, wherein certain modes serve as primary carriers of evaluative meaning while others function as reinforcing or modulating layers. The study further demonstrates that ELAN's layered annotation capabilities enable precise temporal mapping of modal co-occurrence, revealing the intricate choreography of meaning-making across semiotic resources. These findings contribute to the growing field of multimodal film analysis by offering both methodological innovations and theoretical insights into how cinematic texts construct complex evaluative meanings that transcend the limitations of verbal language alone. The study concludes by proposing the concept of "multimodal evaluative choreography" as a framework for understanding how filmic texts orchestrate semiotic resources to achieve particular interpretive effects.

Keywords: multimodal discourse analysis, evaluative meaning, Cuban cinema, ELAN annotation, relevance theory, film discourse

1. Introduction

The study of how meaning emerges from the interplay of multiple semiotic resources has emerged as one of the most dynamic frontiers in contemporary linguistics. As Kress and van Leeuwen [1] argue, meaning in contemporary communication is increasingly multimodal, drawing upon linguistic, visual, auditory, and

embodied resources that operate in concert rather than in isolation. This insight holds particular significance for the analysis of film discourse, where cinematic texts present an exceptionally rich site for investigating multimodal meaning-making. Films integrate spoken dialogue, prosodic contours, facial expressions, gestures, camera movements, lighting, sound design, and editing rhythms into a unified communicative experience that cannot be adequately understood through linguistic analysis alone. The complexity of film as a multimodal artifact demands analytical frameworks capable of capturing the simultaneous operation of multiple semiotic systems and their interaction over time.

Within this broader context, the present study focuses on a specific dimension of film discourse: the multimodal construction of evaluative meaning. Evaluative meaning, the expression of attitudes, stances, and subjective positions toward entities, states of affairs, or propositions, represents a fundamental dimension of human communication that has received extensive attention in linguistic pragmatics [2, 3]. In everyday interaction, speakers use language not merely to convey propositional content but to position themselves in relation to what they are saying, signaling approval, disapproval, doubt, admiration, or skepticism. In cinematic discourse, this evaluative dimension becomes particularly complex because it operates across multiple channels: a character's dialogue may express one evaluation while their facial expression communicates another, and the camera's framing may add yet another layer of meaning. How do filmmakers orchestrate verbal and non-verbal resources to communicate evaluative stances that are not explicitly stated in dialogue? How do viewers integrate cues from multiple channels to arrive at nuanced interpretations of characters' attitudes and the film's broader ideological positioning?

These questions are particularly salient for the analysis of Cuban cinema, a national cinema tradition renowned for its aesthetic sophistication and its capacity to engage with complex social realities through indirect expressive strategies. Following the Cuban Revolution of 1959, the establishment of the Cuban Institute of Cinematographic Art and Industry (ICAIC) fostered a vibrant film culture that combined revolutionary commitment with artistic experimentation [4]. Cuban filmmakers developed a distinctive cinematic language characterized by allegory, metaphor, and the strategic use of ambiguity to navigate political and social constraints. Arturo Sotelo's film *Pon tu pensamiento en mí* (1995), a film that reimagines the life of protagonist as a magician-actor navigating contemporary Cuban society, exemplifies this aesthetic tradition. The film's narrative structure incorporates multiple framing devices, including a chorus of scribes who debate how to construct the story, that foreground questions of representation, interpretation, and evaluative stance. Moreover, the film's dialogue is characterized by frequent use of indirect expression, ironic undertones, and layered meanings that depend crucially on non-verbal cues for their interpretation.

This study therefore addresses the following research questions:

(1) What multimodal resources contribute to the construction of evaluative meaning in *Pon tu pensamiento en mí*, and how do these resources interact across verbal, prosodic, kinesic, and cinematic modes?

(2) How can ELAN annotation software be systematically employed to capture and analyze the temporal coordination of multimodal resources in film discourse?

(3) What patterns of modal configuration characterize the film's strategies for conveying evaluative meaning, and what do these patterns reveal about the nature of multimodal meaning-making in cinema?

The significance of this research is both methodological and theoretical. Methodologically, the study demonstrates the application of ELAN as a powerful tool for multimodal transcription and analysis in film studies, offering a replicable framework for future research. The study provides a detailed annotation protocol that can be adapted by other researchers working with audiovisual data. Theoretically, the study contributes to the growing field of multimodal discourse analysis by developing the concept of "multimodal evaluative choreography", the orchestrated deployment of multiple semiotic resources to achieve particular evaluative

effects, as a framework for understanding how cinematic texts guide viewer interpretation. This concept advances beyond existing approaches by emphasizing the temporal and hierarchical dimensions of modal interaction.

The paper is structured as follows. Section 2 reviews the relevant literature on multimodal discourse analysis, evaluative meaning, and film discourse, situating the present study within current scholarly conversations. Section 3 presents the theoretical framework integrating multimodal discourse analysis with Relevance Theory. Section 4 details the methodological approach, including data selection, ELAN annotation procedures, and analytical protocols. Section 5 presents the analysis of six key scenes from *Pon tu pensamiento en mí*, tracing the multimodal configurations of evaluative meaning across each. Section 6 discusses the findings in relation to the research questions and proposes the concept of multimodal evaluative choreography. Section 7 concludes by summarizing the contributions of the study and suggesting directions for future research.

2. Literature review

2.1. Multimodal discourse analysis: foundations and developments

Multimodal Discourse Analysis (MDA) has emerged over the past two decades as a robust framework for understanding how meaning is constructed through the integration of multiple semiotic resources. The foundational work of Kress and van Leeuwen [1] extended Halliday's systemic functional linguistics to visual communication, proposing a grammar of visual design that identifies representational, interactive, and compositional meanings in images. This approach has since been extended to a wide range of semiotic modes, including typography, layout, gesture, and sound [5-7]. A key insight from this tradition is that different semiotic modes have different affordances, distinct potentials for meaning-making that derive from their material properties and social histories [8]. Language is temporally linear and relies on discrete symbolic units; gestures are embodied and indexical; images are spatially organized and capable of simultaneous presentation of multiple elements; sound is immersive and affective. When these modes are combined in multimodal texts, they do not simply duplicate meaning but rather contribute distinct dimensions that together create a whole greater than the sum of its parts.

Recent developments in multimodal research have focused increasingly on the temporal dynamics of multimodal interaction. Scholars have recognized that meaning in multimodal texts is not merely the product of simultaneous co-occurrence of different modes but emerges from their precise temporal coordination [9]. This temporal dimension is particularly crucial for film analysis, where the sequencing of shots, the timing of cuts, and the synchronization of sound with image all contribute to meaning. Advances in digital annotation tools, particularly ELAN [10], have made it possible to capture these temporal relationships with unprecedented precision, enabling researchers to analyze how different modes are coordinated across time and how this coordination shapes interpretation.

In the context of film, scholars have developed specialized approaches to multimodal analysis. Bateman and Schmidt [11] propose a "multimodal film analysis" framework that integrates linguistic analysis of dialogue with analysis of visual composition, editing patterns, and sound design. Their approach emphasizes the need to account for the hierarchical organization of filmic texts, where individual shots combine into sequences, and sequences combine into larger narrative units. Baldry and Thibault [12] offer a comprehensive methodology for multimodal transcription of audiovisual texts, identifying distinct resource systems including the visual frame, kinesic action, spoken language, and sound effects. These approaches share a commitment to systematic, empirically grounded analysis that respects the material specificity of each semiotic mode while

attending to their interrelations. What remains less developed, however, is a framework specifically focused on evaluative meaning, how films guide viewers to adopt particular attitudes toward characters, events, and ideas.

2.2. Evaluative meaning in language and multimodal contexts

The expression of evaluation, i.e., attitudes, stances, judgments, and feelings, has been extensively studied within linguistic pragmatics and discourse analysis. Martin and White's [2] Appraisal Theory provide a comprehensive framework for analyzing evaluative language, distinguishing between attitude (affect, judgment, appreciation), engagement (resources for negotiating dialogistic positioning), and graduation (scaling of intensity). Within this framework, evaluation is understood as a fundamentally intersubjective phenomenon, positioning speakers and hearers within communities of shared values. The theory has been applied to a wide range of discourse types, from news reporting to literary texts, demonstrating the pervasive role of evaluation in shaping meaning.

However, as recent research has demonstrated, evaluation in multimodal texts extend beyond verbal resources. Studies of multimodal stance-taking have shown how gesture, facial expression, and prosody contribute to the expression of evaluative positions [13, 14]. In film, the cinematic apparatus itself, i.e., camera angles, lighting, editing pace, can convey evaluative perspectives, aligning viewers with or against characters and shaping emotional responses [15]. A high angle shot may position a character as vulnerable or diminished; low-key lighting may create an atmosphere of menace; rapid editing may generate excitement or anxiety. These cinematic choices constitute a form of evaluation that operates alongside and interacts with verbal evaluation.

Of relevance to the present study is the phenomenon of indirect evaluative expression, including irony and other forms of non-literal meaning. Sperber and Wilson's [16, 17] Relevance Theory offers a cognitively oriented account of irony as echoic interpretation, a form of language use in which a speaker implicitly attributes a thought or utterance to someone else while dissociating themselves from it. This account has proven influential in multimodal analysis because it focuses on interpretive mechanisms that can be triggered by non-verbal as well as verbal cues [18, 19]. From a Relevance Theory perspective, irony does not require a special linguistic form but arises from the recognition that an utterance echoes another thought or utterance while communicating a dissociative attitude toward it. This framework can be extended to cinematic communication: a character's gesture, a camera angle, or a musical cue can function as an echoic device, signaling that the film is inviting viewers to adopt a particular evaluative stance.

Burczynska's [20] study of multimodal irony in film translation represents a significant advance in this area. Analyzing the Sherlock Holmes films, she demonstrates that multimodal irony is not carried by dialogue alone but emerges from the interplay of visual, kinesic, and acoustic modes. Her identification of "preservation" and "modification" strategies for irony translation reveals the complexity of multimodal meaning-making and the challenges of cross-cultural transfer. Smith's [21] analysis of irony in *Buffy the Vampire Slayer* further extends this line of inquiry, examining how prosodic contours, gestures, and facial expressions correlate with specific syntactic patterns to produce ironic effects. These studies collectively demonstrate the importance of attending to multiple semiotic channels in the analysis of evaluative meaning.

2.3. Cuban cinema and the aesthetics of indirection

Cuban cinema occupies a distinctive position within world cinema history, shaped by the Cuban Revolution's 1959 establishment of the Cuban Institute of Cinematographic Art and Industry (ICAIC). As Chanan 4 documents, ICAIC filmmakers developed a cinema characterized by both revolutionary commitment and aesthetic experimentation, drawing upon Soviet montage, Italian neorealism, and Latin American cultural

traditions. The early years of Cuban cinema produced landmark films such as *Memorias del subdesarrollo* directed by Tomás Gutiérrez Alea in 1968 and *Lucía* directed by Humberto Solás in 1968, which established a tradition of politically engaged yet formally innovative filmmaking.

A recurring feature of Cuban cinema, particularly known as the Special Period of the 1990s, is the use of indirect expression to navigate political and social constraints. Scholars have noted the prevalence of allegory, irony, and metaphorical discourse in Cuban films of this period [22, 23]. This aesthetic of indirection serves multiple functions: it permits engagement with sensitive topics while maintaining a degree of protective ambiguity; it rewards the cultural competence of local audiences who can recognize the referents of allegorical and ironic discourse; and it contributes to a distinctive cinematic language characterized by layered meanings and interpretive openness. For international audiences, these films often require additional interpretive work to access the full range of their meanings.

Arturo Sotelo's *Pon tu pensamiento en mí* exemplifies this aesthetic. The film's framing device, scribes debating how to construct the story of protagonist as a magician-actor, self-consciously raises questions about representation and interpretation. The casting of Fernando Hechavarría as Jesus and Luis Alberto García as his double further complicates questions of identity and authenticity. Throughout the film, characters speak in ways that invite multiple interpretations, and the cinematic techniques of camera movement, lighting, and editing contribute layers of meaning that supplement and sometimes contradict the dialogue. The film's title itself *Pon tu pensamiento en mí (Put Your Thought on Me)* invites viewers to engage in an act of imaginative projection, to construct meaning actively rather than passively receiving it.

Despite the film's significance within Cuban cinema, it has received limited attention in Anglophone film and linguistic studies. This study addresses this gap by bringing systematic multimodal analysis to bear on the film's strategies for constructing evaluative meaning.

3. Theoretical framework

The theoretical framework for this study integrates insights from multimodal discourse analysis, Relevance Theory, and film theory. This integration is motivated by the need for a framework that can account for both the semiotic specificity of different modes and the cognitive mechanisms through which viewers integrate multiple cues into coherent interpretations.

3.1. Systemic functional multimodal discourse analysis

The study adopts a systemic functional approach to multimodal analysis, building on the foundational work of Kress and van Leeuwen [1] and its extensions to film by Bateman and Schmidt [11]. From this perspective, meaning-making is understood as a social semiotic practice in which communicators select from available semiotic resources to achieve particular communicative goals. The focus on selection and combination, the "meaning potential" of resources, allows for systematic analysis of how specific choices contribute to overall meaning.

In the context of film, the systemic functional approach directs attention to the following resource systems:

- Linguistic system: The verbal dialogue, including lexical choices, syntactic structures, and rhetorical patterns.
- Prosodic system: Intonation, stress, rhythm, and voice quality.
- Kinesic system: Facial expression, gesture, posture, gaze, and head movement.
- Cinematic system: Camera movement, framing, angle, focus, lighting, color.
- Editing system: Shot transitions, pacing, rhythm.

- Sound system: Music, ambient sound, sound effects.

Each of these systems has its own semiotic affordances and constraints, and each contributes distinct dimensions to the construction of evaluative meaning. A systemic functional approach also emphasizes that these systems are not isolated but interact in systematic ways. For instance, the choice of a close-up shot (cinematic system) may be coordinated with a character's facial expression (kinesic system) to highlight an evaluative reaction, creating a multimodal configuration that carries meaning beyond what either system could convey alone.

3.2. Relevance theory and echoic interpretation

Relevance Theory [16, 17] provides a cognitive framework for understanding how viewers infer communicative intentions from multimodal input. Central to this framework is the principle of relevance: human cognition is geared toward maximizing the relevance of incoming stimuli, and communication operates on the assumption that speakers intend their utterances to be optimally relevant to hearers. Relevance is defined in terms of cognitive effects and processing effort: the more cognitive effects a stimulus yields, and the less effort required to process it, the more relevant it is.

For the analysis of indirect evaluative meaning, Relevance Theory's account of echoic interpretation is particularly valuable. Echoic utterances are those in which a speaker attributes a thought or utterance to someone else while implicitly communicating an attitude toward that attributed content. This account has been extended to multimodal contexts by Yus [18], who argues that non-verbal cues can function as "echoic devices" that trigger interpretive processes similar to those involved in verbal irony. In the context of film, echoic interpretation is activated not only by character dialogue but also by cinematic techniques that create interpretive distance. A camera angle that frames a character from a slightly distorted perspective, a musical cue that seems incongruous with the visual content, or a gesture that seems exaggerated, all can function as cues that the film is not simply presenting an event but rather inviting viewers to adopt a particular evaluative stance toward it.

Relevance Theory also provides a framework for understanding how viewers resolve ambiguity in multimodal texts. When multiple interpretive possibilities are available, viewers will select the interpretation that yields the greatest cognitive effects with the least processing effort, consistent with the principle of relevance. This cognitive perspective complements the semiotic approach of multimodal discourse analysis by providing an account of the interpretive processes that viewers engage in when processing multimodal texts.

3.3. Multimodal evaluative choreography: an integrative concept

Drawing on these theoretical resources, I propose the concept of "multimodal evaluative choreography" as an integrative framework for analyzing how films construct evaluative meaning. The term "choreography" is chosen deliberately to capture several key features of multimodal meaning-making in cinema:

First, temporal coordination: Like dance, filmic meaning emerges from the precise timing of multiple elements across time. A gesture that coincides with a particular word, a cut that aligns with a shift in gaze, a musical phrase that culminates with a character's utterance, these temporal alignments are not accidental but carefully orchestrated to produce specific effects. The choreographic metaphor emphasizes that meaning in film is not merely a matter of what modes are present but when and how they interact.

Second, hierarchical organization: In any multimodal configuration, certain modes may be foregrounded while others play supporting roles. The concept of choreography acknowledges that not all elements are equally significant at a given moment; rather, there is a dynamic hierarchy of salience that shifts across time.

This hierarchical organization guides viewers' attention, directing them to the most relevant cues for interpretation.

Third, interpretive guidance: Choreography is not merely a descriptive concept but a functional one: the orchestration of multimodal resources is oriented toward guiding viewers' interpretive processes. The film's choreography of modes directs attention, suggests relationships, and cues evaluative stances. From a Relevance Theory perspective, effective choreography reduces processing effort by making the most relevant cues salient at the appropriate moments.

Fourth, emergent meaning: Just as dance movements derive meaning from their combination into sequences rather than from isolated steps, multimodal evaluative meaning emerges from the integration of multiple resources rather than from any single channel. This emergent quality is what makes multimodal analysis essential, the meaning of a film cannot be adequately captured by analyzing its dialogue alone, nor by analyzing its visual composition alone, but only through the integrated analysis of how these resources work together.

This framework guides the analytical work presented in Section 5, enabling systematic examination of how verbal, prosodic, kinesic, and cinematic resources work together to construct evaluative meaning in *Pon tu pensamiento en mí*.

4. Methodology

4.1. Data selection

The primary data for this study consists of *Pon tu pensamiento en mí*, an 85-minute Cuban fiction film produced by ICAIC. The film was selected for several reasons. First, its narrative complexity and self-reflexive framing makes it a rich site for investigating multimodal evaluative meaning. Second, its dialogue exhibits frequent use of indirect expression that requires non-verbal cues for interpretation. Third, the film's status within Cuban cinema, a national tradition with distinctive aesthetic conventions, provides a specific cultural context for analysis. Fourth, the film has received limited scholarly attention, offering opportunities for original contribution.

From the full film, six key scenes were selected for detailed multimodal analysis. These scenes were chosen based on their centrality to the film's narrative development and their density of evaluative expression. The selected scenes are:

- (1) Scene 1: The scribes' opening debate about how to construct the story (approximately 4 minutes).
- (2) Scene 2: Tomás offering the protagonist in this movie Jesus a ride in the cart (approximately 2 minutes).
- (3) Scene 3: The scribes discussing the actor who will play the protagonist in this movie Jesus (approximately 3 minutes).
- (4) Scene 4: The protagonist in this movie Jesus's first encounter with the townspeople (approximately 5 minutes).
- (5) Scene 5: The scribes reacting to early footage (approximately 2 minutes).
- (6) Scene 6: The protagonist in this movie Jesus's final scene (approximately 4 minutes).

These scenes collectively represent approximately 20 minutes of screen time, providing a substantial corpus for detailed multimodal analysis while remaining manageable for manual annotation.

4.2. ELAN annotation protocol

EUDICO Linguistic Annotator (ELAN), developed by the Max Planck Institute for Psycholinguistics, was employed for multimodal annotation of the selected scenes. ELAN is a widely used tool for time-aligned

annotation of audio and video data, enabling researchers to create multiple tiers of annotation that can be synchronized with specific temporal segments of the media file. The software supports hierarchical tier structures, allowing for complex annotation schemes that capture relationships between different types of phenomena.

The annotation protocol for this study comprised the tier structure shown in Table 1.

Table 1. ELAN annotation tier structure

Tier Category	Specific Tiers	Description
Linguistic	Utterance	Transcription of dialogue
Linguistic	Lexical Evaluation	Words with evaluative meaning (judgment, affect)
Prosodic	Intonation	Pitch contour patterns
Prosodic	Stress	Emphatic stress patterns
Prosodic	Pause	Pauses and hesitations
Kinesic	Facial Expression	Emotional expressions, eyebrow movements, mouth movements
Kinesic	Gaze	Gaze direction, mutual gaze
Kinesic	Gesture	Hand gestures, head movements
Kinesic	Posture	Body orientation, posture shifts
Cinematic	Camera Movement	Pan, tilt, zoom, tracking
Cinematic	Shot Scale	Close-up, medium, long shot
Cinematic	Angle	High, low, eye-level
Cinematic	Focus	Depth of field, focus shifts
Editing	Cut Type	Match cut, jump cut, cross-cut
Editing	Shot Duration	Length of shots
Sound	Music	Musical score cues
Sound	Ambient	Environmental sounds

Each tier was annotated for the duration of each selected scene, with annotations time-aligned to the video. Annotations were reviewed and refined in multiple passes to ensure consistency. The use of ELAN enabled precise temporal analysis of modal co-occurrence, revealing how different modes align or diverge across time.

4.3. Analytical procedures

The analysis proceeded through several phases:

4.3.1. Phase 1: Descriptive annotation

The six scenes were annotated following the protocol above, creating a comprehensive record of multimodal resources in each scene. This phase involved repeated viewing of each scene to ensure accurate annotation of all relevant phenomena.

4.3.2. Phase 2: Modal configuration analysis

For each evaluative moment identified, I analyzed how different modes combined to construct evaluative meaning. This involved identifying patterns of modal convergence (where multiple modes communicate consistent evaluative stances), modal dissonance (where modes communicate conflicting cues), and modal hierarchy (where certain modes dominate). The ELAN interface was particularly valuable for this phase, as it allowed visualization of multiple tiers simultaneously, revealing temporal alignments and divergences.

4.3.3. Phase 3: Cross-scene pattern identification

Across the six scenes, I identified recurring patterns of multimodal configuration, grouping similar patterns and analyzing their functional differences. This phase involved comparing annotation files across scenes to identify both consistent patterns and variations.

4.3.4. Phase 4: Interpretive analysis

Drawing on Relevance Theory, I analyzed how the identified multimodal configurations might guide viewers' interpretive processes, examining the cognitive mechanisms through which viewers integrate multimodal cues into coherent interpretations. This phase considers how different configurations might affect processing effort and the range of possible interpretations.

4.4. Reliability and validity

To ensure analytical reliability, a second analyst independently annotated a subset of the data (approximately 20% of the selected scenes). Inter-annotator agreement was calculated using Cohen's kappa, with coefficients exceeding 0.85 for all annotation categories, indicating strong agreement. Discrepancies were resolved through discussion, with the annotation protocol refined to clarify ambiguous categories.

Validity is supported through multiple strategies: the use of established theoretical frameworks, systematic annotation protocols, triangulation across multiple scenes, and attention to disconfirming evidence. The analysis explicitly considered alternative interpretations of multimodal configurations and sought evidence that might challenge the proposed patterns.

5. Analysis

This section presents the multimodal analysis of the six selected scenes from the Cuban film *Pon tu pensamiento en mí*. For each scene, the configuration of linguistic was examined, prosodic, kinesic, and cinematic resources, tracing how evaluative meaning emerges through modal interaction.

5.1. Scene 1: the scribes' opening debate

The film opens with a chorus of scribes, i.e., seven individuals who appear to be constructing the story we are about to witness. This framing device immediately establishes a self-reflexive dimension, inviting viewers to consider the act of storytelling itself. The scribes debate questions of form, content, and interpretation, and their dialogue is rich with evaluative language that is significantly modulated by non-verbal cues.

5.1.1. Linguistic analysis

The scribes' dialogue employs explicit evaluative language that positions them in relation to the story they are constructing. Consider the following exchanges:

Escriba I: "Vida, Pasión y Muerte de un prestidigitador, la típica estructura cíclica. Es muy raro que fallen estas historias".

Life, Passion, and Death of a magician, the typical cyclical structure. It's very rare for these stories.

Escriba III: "Ni siquiera sabemos el posible argumento que tendrá esta historia. No tenemos una lista de sucesos y ya estamos hablando de la textura que tendrá la película".

We don't even know the possible plot this story will have. We don't have a list of events and we're already talking about what the texture of the film will be.

Escriba IV: "Detesto el blanco y negro y la libertad en el cine. Las leyes en este arte son como las mujeres, no se les puede abandonar por mucho tiempo".

I detest black and white and freedom in cinema. The laws in this art are like women; they cannot be left for long.

The evaluative language here ranges from judgment ("típica"/typical) to affect ("detesto"/I detest) to appreciation of artistic form ("la textura"/texture). However, the evaluative force of these statements is significantly shaped by their multimodal context. Escriba I's claim that such stories "very rarely fail" is presented with a certainty that the surrounding context suggests is ironic.

5.1.2. Prosodic and kinesic analysis

ELAN annotation reveals systematic patterns of prosodic and kinesic cues that modulate the scribes' evaluative stances. In the utterance of Escriba IV "Detesto el blanco y negro y la libertad en el cine", the prosodic contour shows a sharp rise and fall on "detesto", with extended duration on the stressed syllable. Simultaneously, the character's facial expression displays a slight smirk, with eyebrows raised and head tilted. This combination of emphatic prosody with a smirk that contradicts the apparent seriousness of the statement creates a complex evaluative signal. The smirk functions as what Relevance Theory would term an echoic cue, signaling that the speaker is not simply expressing personal distaste but rather performing a stance for effect.

The gaze patterns among the scribes are also significant. In the debate sequences, scribes frequently avert their gaze when making strong evaluative claims, only establishing mutual gaze when they appear to be seeking agreement. This pattern suggests that the evaluative statements are performative, part of a negotiation of group identity, rather than expressions of individual conviction. ELAN tracking of gaze direction across the scene shows that moments of mutual gaze consistently follow the most exaggerated evaluative statements, as if scribes are checking for others' reactions.

5.1.3. Cinematic analysis

The cinematic treatment of this scene reinforces the sense of evaluative negotiation. Camera placement shifts between medium shots that include multiple scribes and closer shots that isolate individuals. The editing rhythm is relatively slow, with shots averaging 8-10 seconds, allowing viewers time to process the interplay of verbal and non-verbal cues. This slow editing rhythm is itself an evaluative choice: it signals that the scene is not simply conveying information but inviting reflection.

Notably, when Escriba IV makes the comparison between cinematic laws and women, the camera holds on his face for an extra two seconds after he finishes speaking. This slight delay creates a moment for viewers to register the reaction, both of other characters and their own, to this provocative statement. The following shot shows other scribes exchanging glances, some with raised eyebrows, others suppressing smiles. This cinematic choice guides viewer interpretation by focusing attention on the reaction to the statement rather than the statement itself.

5.1.4. Multimodal evaluative configuration

In this scene, evaluative meaning is constructed through a configuration that might be termed modal dissonance with performative framing. The verbal content presents strong evaluative positions, but prosodic cues (emphatic delivery), kinesic cues (smirks, averted gazes), and cinematic cues (sustained focus on reactions) collectively suggest that these positions are being performed rather than earnestly held. The effect is to create interpretive distance; viewers are invited not to take the scribes' evaluative statements at face value but to consider them as part of a metadiscursive exploration of storytelling.

5.2. Scene 2: Tomás offers the protagonist in this movie Jesus a ride

This scene depicts Tomás encountering the protagonist in this movie Jesus walking with his bicycle in a rural setting. Tomás invites Jesus to join him in his cart, and their dialogue establishes a dynamic that will be central to the film.

5.2.1. Linguistic analysis

Tomás's invitation is framed in language that is simultaneously welcoming and evaluative:

Tomás: "Muy lejos ha de ser su destino y muy tarde su llegada, si persiste en esa vieja bicicleta. Los caminos de Dios son inescrutables. Suba, suba a esta carreta de divina o diabólica comedia, y que la suerte nos acompañe juntos".

Very far must be his destiny, and his late arrival, if he insists on that old bicycle. The ways of God are inscrutable. Climb, climb aboard this cart of divine or diabolical comedy, and may fortune accompany us together.

The language here is richly evaluative, blending religious reference ("los caminos de Dios"), theatrical metaphor ("divina o diabólica comedia"), and practical invitation. The qualification "divina o diabólica" is particularly significant, it presents an interpretive choice, inviting Jesus (and the viewer) to consider what kind of narrative they are entering. The repetition of "suba" creates a rhythmic insistence that reinforces the invitation's sincerity while the metaphorical framing of the cart as comedy suggests the scene's self-conscious theatricality.

5.2.2. Prosodic and kinesic analysis

ELAN annotation reveals the complexity of Tomás's embodied performance during this scene. As he delivers the line "suba, suba a esta carreta de divina o diabólica comedia", his gesture involves an open hand extended toward Jesus, palm up, with a slight shrugging movement of the shoulders. The repetition of "suba" is accompanied by a rhythmic nodding gesture that emphasizes the invitation's sincerity. The temporal alignment between the verbal repetition and the nodding gesture creates a multimodal rhythm that reinforces the invitation's urgency.

However, the facial expression during the phrase "divina o diabólica comedia" shows a momentary narrowing of the eyes and a slight pursing of the lips, micro-expressions that suggest a kind of amused skepticism. This combination of open, welcoming gestures with micro-expressions that signal interpretive complexity creates a layered evaluative stance. The micro-expressions function as echoic cues, suggesting that Tomás is aware of the multiple possible interpretations of his words and inviting Jesus to share in this awareness.

Jesus's responses are minimal and his kinesic profile is subdued. He maintains relatively neutral facial expression, with gaze directed at Tomás but with minimal eye movement. This contrast between Tomás's expansive expressiveness and Jesus's restraint itself communicate evaluative meaning, suggesting that Jesus is a receptive but enigmatic figure, one who observes more than he performs.

5.2.3. Cinematic analysis

The scene is shot largely in medium two-shots that frame both characters together, emphasizing their interaction. However, when Tomás delivers the phrase about "divine or diabolical comedy", the camera cuts to a slightly tighter shot of his face, allowing viewers to register the micro-expressions described above. This camera movement functions as an interpretive guide, directing attention to the subtle cues that carry evaluative complexity. The cut back to Jesus shows his reaction, a slight inclination of the head that could be interpreted as acknowledgment, agreement, or something more ambiguous.

The lighting in this scene is soft and naturalistic, with late afternoon sunlight creating long shadows. This aesthetic choice contributes to the scene's evaluative tone, it feels both mundane and slightly mythic, supporting the film's construction of Jesus as a figure who is simultaneously ordinary and extraordinary. The golden light evokes religious painting traditions while the setting (a rural road, a cart) grounds the scene in everyday Cuban reality.

5.2.4. *Multimodal evaluative configuration*

This scene exemplifies what I term modal convergence with differential weighting. All modes converge to construct a consistent evaluative stance: Tomás is welcoming, but his welcome comes with interpretive complexity; Jesus is receptive but enigmatic. However, different modes carry different weights in conveying this stance. Linguistic content carries explicit invitation and thematic framing; kinesic and prosodic cues carry the complexity and ambiguity; cinematic framing guides viewers' attention to the most significant cues. The differential weighting creates a hierarchy of salience that guides interpretation.

5.3. Scene 3: the scribes discuss the actor

This scene returns to the scribes, now discussing the actor who will play the protagonist in this movie Jesus in the story they are constructing. The scene is significant for its explicit metatheatrical dimension, it addresses directly the question of representation that pervades the film.

5.3.1. *Linguistic analysis*

The scribes' discussion focuses on the physical and personal qualities the actor should possess:

Escriba VI: "Debe llevar ternura en la mirada y tener el pelo un poco largo".

His gaze should convey tenderness, and his hair should be slightly long.

Escriba VII: "Estamos cazando estrellas o camioneros".

We are hunting stars or truck drivers.

Escriba VI's description is poetic and evaluative, focusing on the quality of the actor's gaze. Escriba VII's response is dismissive, contrasting the idealized "stars" with the mundane "truck drivers". This verbal exchange establishes a tension between idealization and realism that runs through the film. The question "stars or truck drivers" encapsulates a central tension in the film's representation of Jesus: is he to be presented as a transcendent figure or a grounded, ordinary one?

5.3.2. *Prosodic and kinesic analysis*

ELAN annotation of this scene reveals that the scribes' kinesic behavior is notably exaggerated compared to earlier scenes. When Escriba VI describes the actor's required qualities, his hand gesture traces a shape in the air, as if outlining an ideal form. His gaze is directed upward and slightly to the side, the classic posture of someone envisioning an ideal. The gesture and gaze together create a multimodal representation of idealization that reinforces the verbal content.

Escriba VII's response is delivered with a sharp, dismissive gesture, a quick downward movement of the hand, and a facial expression that combines raised eyebrows with a slight sneer. The prosodic contour of "estrellas o camioneros" shows a rising pitch on "estrellas" and a falling pitch on "camioneros", with a slight pause between the two nouns that emphasize their contrast. This prosodic pattern creates an evaluative contrast through sound alone: "estrellas" rises toward the ideal, "camioneros" falls toward the mundane.

The other scribes reacted to this exchange with laughter and exchanged glances. ELAN tracking of gaze direction shows that during Escriba VII's dismissive response, several scribes look toward each other rather than toward the speaker, suggesting that they are collectively registering the humor. This collective gaze pattern indicates that the evaluative stance is being negotiated among the group rather than simply expressed by an individual.

5.3.3. *Cinematic analysis*

The cinematography for this scene employs a more dynamic camera than in earlier scribe scenes. The camera moves slowly from one scribe to another, often lingering on their reactions. When Escriba VI makes the idealized description, the camera captures him in a medium close-up that emphasizes his expressive gestures.

The cut to Escriba VII for his dismissive response is slightly faster than previous cuts, creating a sense of verbal repartee. This editing pattern contributes to the scene's comedic rhythm, reinforcing the scribes' performances as theatrical rather than documentary.

5.3.4. *Multimodal evaluative configuration*

This scene demonstrates what I term modal reinforcement with intermodal complementarity. The verbal content establishes a contrast between idealized and realistic perspectives. Kinesic cues reinforce this contrast through gestures and facial expression. Cinematic choices, i.e., camera movement, cutting rhythm, shot selection, work to highlight the interactional dynamics among the scribes, emphasizing that their discussion is as much about group performance as about substantive content. The intermodal complementarity means that each mode contributes a distinct dimension to the overall evaluative meaning.

5.4. Scene 4: protagonist in this movie Jesus's first encounter with the townspeople

This scene depicts the protagonist in this movie Jesus's arrival in a town and his first interactions with its inhabitants. The scene is notable for its use of visual and auditory cues that construct an evaluative stance toward both Jesus and the townspeople.

5.4.1. *Linguistic analysis*

The dialogue in this scene includes several instances of indirect evaluative expression. When the protagonist in this movie Jesus performs a magic trick, a townsman responds:

Townsman: "Eso no es nada, eso lo hace cualquiera".

That's nothing, anyone can do that.

The utterance is superficially dismissive, but the prosodic and kinesic context suggests a more complex evaluative stance, admiration masked as dismissal. The use of "cualquiera" (anyone) is a common formula for expressing dismissiveness, but its pragmatic function in context is to express the opposite: the speaker is impressed but unwilling to admit it.

5.4.2. *Prosodic and kinesic analysis*

ELAN annotation of this scene reveals the crucial role of prosody in signaling the intended evaluative stance. The townsman's utterance is delivered with a rising-falling intonation on "nada" and "cualquiera" that is typically associated with teasing rather than genuine dismissal. His facial expression includes a smile that is not fully suppressed, and his gaze is directed at Jesus with an intensity that suggests fascination rather than contempt. The smile functions as a key echoic cue, signaling that the dismissal is not to be taken at face value.

Jesus's response is minimal, a slight smile and a nod, but his posture remains open and relaxed, suggesting that he interprets the utterance as the teasing it appears to be rather than genuine hostility. His embodied response acknowledges the teasing and accepts it, modeling for viewer show to interpret the exchange.

5.4.3. *Cinematic analysis*

The cinematic treatment of this scene is noteworthy for its use of shot-reverse-shot patterns that emphasize the interaction between Jesus and the townspeople. However, the editing departs from conventional patterns in two respects. First, the shots of Jesus are consistently longer than shots of the townspeople, suggesting his centrality to the scene's evaluative structure. Second, the shots of the townspeople often include other townspeople in the background, their reactions visible even when not the focus of the shot. This technique constructs a sense of collective evaluation; the townspeople are not just individuals responding to Jesus but a community negotiating its response.

5.4.4. *Multimodal evaluative configuration*

This scene illustrates modal convergence with interpretive layering. All modes converge to indicate that the townsman's apparent dismissal is actually admiration. However, the configuration also creates interpretive layering: the dismissal is still present as a surface meaning, with the admiration visible through prosodic, kinesic, and cinematic cues. Viewers must integrate these multiple layers to arrive at the intended interpretation. Layering invites active interpretive work rather than passive reception.

5.5. Scene 5: the scribes reacting to footage

This scene shows the scribes watching footage that has been shot, commenting on what they see. It functions as another metadiscursive moment, foregrounding questions of interpretation and evaluation.

5.5.1. *Linguistic analysis*

The scribes' comments include explicit evaluation of the footage:

Escriba VI: "Es horrible, sencillamente horrible, desde la primera escena, se sabe todo lo que va a pasar".

It's horrible, simply horrible, from the very first scene, you know everything that's going to happen.

Escriba IV: "A quién se le ocurre dejar una bicicleta tirada así en medio del desierto".

Who would leave a bicycle lying like that in the middle of the desert?

Escriba I: "Mejor volvamos a los orígenes, hagamos como en las Sagradas Escrituras".

Let us return to the origins; let us do as in the Holy Scriptures.

The evaluations are strongly negative, but the context suggests they may not be meant seriously. The repetition of "horrible" emphasizes the negativity, but the content of the criticism, that a bicycle in the desert is absurd, is itself absurd as a serious critique of a film about a magician.

5.5.2. *Prosodic and kinesic analysis*

The scribes' delivery in this scene is more emphatic than in earlier scenes. Escriba VI's "es horrible, sencillamente horrible" is delivered with exaggerated stress and a facial expression of mock horror, i.e., eyes wide, mouth open, hands raised. This exaggerated performance cues viewers not to take the evaluation at face value but to recognize it as part of the scribes' ongoing negotiation of their story. Exaggeration functions as an echoic cue, signaling that the evaluation is being performed for effect.

Escriba IV's comment about the bicycle is delivered with a gesture of exasperation, arms spread wide, palms up, that suggests the comment is as much about the absurdity of their own process as about the footage itself. The gesture invites viewers to laugh at the scribes' pretensions to critical authority.

5.5.3. *Cinematic analysis*

The camera in this scene frequently cuts to close-ups of individual scribes as they react, emphasizing the performative quality of their responses. However, the editing also includes reaction shots of other scribes, showing their responses to the evaluations being offered. This creates a recursive structure of evaluation, scribes evaluating the footage, and the film (and viewers) evaluating the scribes' evaluations. The recursive structure invites viewers to adopt a metaperspective on the act of evaluation itself.

5.5.4. *Multimodal evaluative configuration*

This scene exemplifies meta-evaluative choreography, a configuration in which the modes work together to communicate that the evaluation being performed is itself subject to evaluation. The exaggerated prosodic and kinesic cues signal that the scribes' negative evaluations are not final judgments but contributions to an ongoing process of interpretive negotiation. The cinematic structure reinforces this by showing the scribes' evaluations being received and reacted to by other scribes.

5.6. Scene 6: the protagonist in this movie Jesus's final scene

The final scene of the film shows Jesus's death and the community's response. The scene is both emotionally charged and interpretively complex, drawing together the thematic threads established throughout the film.

5.6.1. Linguistic analysis

The narrator's voice-over frames the scene with language that constructs evaluation through its religious resonance:

Narrador: "Descansa en paz Jesús, que tu alma penetre como ninguna en el reino de los cielos, pero que tu espíritu quede entre nosotros, para que nos ilumine y guíe".

Rest in peace, Jesús; may your soul enter the kingdom of heaven like no other, but may your spirit remain among us, to illuminate and guide us.

This language echoes Christian liturgical forms, but its context within a film about a magician-actor creates an evaluative stance that is simultaneously reverent and skeptical. The prayer-like structure invites a reverent response, but the framing of Jesus as a magician complicates this response. The concluding phrase "to illuminate and guide us" can be read as sincere or ironic, leaving interpretive openness.

5.6.2. Prosodic and kinesic analysis

The actors' performances in this scene are subdued, with restrained gestures and facial expression. The townspeople's reactions are shown through slow camera movements rather than close-ups, creating a sense of collectiveness rather than individual emotion. The restrained kinesic performances suggest a shift in register from the earlier scenes, the metadiscursive playfulness gives way to something approaching seriousness.

5.6.3. Cinematic analysis

The cinematography for this scene uses high-contrast lighting, with shadows emphasizing the drama of the moment. The camera movement is slow, with long takes that allow viewers to absorb visual composition. The final shot of the scene, and the film, holds on Jesus's face for an extended duration, allowing viewers to project their own interpretive closure onto the enigmatic expression. The extended duration creates space for reflection, inviting viewers to arrive at their own interpretation.

5.6.4. Multimodal evaluative configuration

This scene demonstrates modal convergence with hermeneutic openness. All modes converge to create a tone of emotional gravity, but they do not converge on a single evaluative interpretation. Rather, the configuration opens multiple interpretive possibilities, i.e., the protagonist in the movie can be understood as a tragic figure, a holy figure, a figure of resistance, or all these simultaneously. The hermeneutic openness is the evaluative stance: the film invites viewers to hold multiple interpretive possibilities in tension rather than resolving them into a single meaning.

6. Discussion

6.1. Summary of findings

The analysis presented in Section 5 yields several significant findings regarding the multimodal construction of evaluative meaning in *Pon tu pensamiento en mí*.

First, evaluative meaning in this film emerges not from any single semiotic channel but from the orchestrated interaction of multiple channels. Linguistic content provides explicit evaluative propositions, but these are consistently modulated by prosodic, kinesic, and cinematic cues. In many cases, the "true" evaluative stance, the one that the film guides viewers to adopt, is not fully captured by linguistic content alone but

requires integration of cues from multiple modes. This finding supports the central premise of multimodal discourse analysis: meaning in multimodal texts is distributed across semiotic resources rather than concentrated in any single mode.

Second, the analysis identified three primary patterns of modal configuration across scenes: (1) modal convergence, wherein all modes align to reinforce a consistent evaluative stance; (2) modal dissonance, wherein conflicting cues create interpretive tension that invites active viewer engagement; and (3) modal hierarchy, wherein certain modes serve as primary carriers of evaluative meaning while others function as reinforcing or modulating layers. These patterns are not mutually exclusive but often co-occur within single scenes, creating complex configurations. The identification of these patterns provides a vocabulary for describing multimodal evaluative strategies in film.

Third, the study demonstrates that ELAN's layered annotation capabilities enable precise temporal mapping of modal co-occurrence, revealing the intricate choreography of meaning-making across semiotic resources. The ability to align annotations across multiple tiers and to visualize their temporal relationships is essential for capturing the moment-by-moment coordination that characterizes multimodal communication. Without such tools, analyses risk oversimplifying the complexity of multimodal texts.

Fourth, the analysis reveals that the film's self-reflexive framing, the scribes who discuss how to construct the story, function as a meta-evaluative device. The scribes' performances model the act of evaluation for viewers, showing evaluation as a social, performative, and negotiated process. This meta-evaluative dimension contributes to the film's distinctive aesthetic of indirection, inviting viewers to adopt a reflective stance toward the act of interpretation itself.

6.2. Theoretical contributions: multimodal evaluative choreography

Building on these findings, I propose the concept of multimodal evaluative choreography as a framework for understanding how films construct evaluative meaning. This concept captures several key insights that emerged from the analysis:

6.2.1. *Temporal coordination as meaning-making*

The analysis revealed that the precise timing of modal elements is crucial to their interpretive effect. A gesture that coincides with a particular word, a cut that aligns with a shift in gaze, a musical phrase that culminates with a character's utterance, these temporal alignments are not incidental but central to the construction of evaluative meaning. The choreographic metaphor emphasizes that meaning emerges not just from what resources are present but from when and how they are coordinated.

6.2.2. *Hierarchical salience and interpretive guidance*

The analysis showed that not all modes carry equal weight at any given moment. In the scribes' opening debate, cinematic cues (shot duration, framing) served to guide attention to reactions, establishing a hierarchy of salience that shaped viewers' interpretive processes. This hierarchical organization is part of the film's choreography, directing viewers to the most relevant cues for interpretation.

6.2.3. *Emergent meaning through integration*

The concept of choreography emphasizes that multimodal evaluative meaning emerges from the integration of multiple resources rather than from any single channel. This emergent quality is particularly evident in scenes with modal dissonance, where conflicting cues do not cancel each other out but rather create interpretive tension that invites active viewer engagement. The integration of conflicting cues produces meanings that no single cue could produce alone.

6.2.4. *Interpretive openness as a communicative strategy*

The final scene's configuration of modal convergence with hermeneutic openness demonstrates that films can use multimodal orchestration not to narrow interpretation to a single meaning but to open multiple interpretive possibilities. This suggests that multimodal evaluative choreography encompasses not only strategies for directing interpretation but also strategies for strategically opening interpretation. The choreography can guide viewers toward ambiguity as well as toward clarity.

From a Relevance Theory perspective, multimodal evaluative choreography can be understood as a strategy for managing processing effort and cognitive effects. Modal convergence reduces processing effort by providing consistent cues across channels; modal dissonance increases processing effort but may yield greater cognitive effects by inviting active interpretive engagement; modal hierarchy guides effort toward the most relevant cues. The film's choreography thus reflects an implicit model of its viewers as active interpreters who can handle complexity.

6.3. Methodological contributions

The study also makes significant methodological contributions. The ELAN annotation protocol developed for this research provides a replicable framework for multimodal film analysis. The tier structure, annotation categories, and analytical procedures can be adapted for other films and other research questions. The protocol's emphasis on temporal precision and modal interaction addresses a gap in existing methodologies, which often treat different modes in isolation or fail to capture their temporal coordination.

Moreover, the study demonstrates that ELAN's capabilities extend beyond transcription to analytical insight, the process of annotating forces researchers to attend to temporal relationships and modal interactions that might otherwise be overlooked. The iterative process of annotation, review, and refinement leads to a deeper engagement with the data than might be achieved through less systematic approaches. The use of ELAN also enables collaboration and replication. The annotation files can be shared with other researchers, allowing for validation and extension of findings. This is particularly important for a field where qualitative analysis can sometimes seem opaque or idiosyncratic.

6.4. Limitations and future directions

Several limitations of this study should be acknowledged. First, the analysis focuses on a single film from a specific cultural context. While this allows for depth analysis, it limits generalizability. Future research should extend this analytical approach to other Cuban films and to films from other national traditions to identify whether the patterns observed here are specific to this film or reflect broader trends in Cuban cinema or in films more generally.

Second, the analysis did not include reception data. While the analysis makes claims about how films guide viewers' interpretive processes, these claims remain inferential. Future research could combine multimodal analysis with experimental methods (such as eye-tracking or post-viewing interviews) to investigate actual viewer responses. Such research could test whether the patterns identified here are indeed guide interpretation as proposed.

Third, the analysis was conducted by a single researcher with reliability checks by a second analyst. Future research could employ larger teams of analysts and more formal measures of inter-annotator reliability across all annotation categories. This would strengthen the reliability of findings and allow for more complex statistical analysis of modal co-occurrence patterns.

Fourth, the study focused on evaluative meaning broadly defined. Future research could examine specific subtypes of evaluative meaning, i.e., irony, sarcasm, parody, in greater detail, perhaps drawing on Relevance

Theory's echoic account as a more specific framework. Such research could investigate whether different subtypes of evaluative meaning are associated with distinct multimodal configurations.

7. Conclusion

This study has investigated the multimodal construction of evaluative meaning in Arturo Soto's Cuban film *Pon tu pensamiento en mí*, employing ELAN annotation software to conduct systematic multimodal analysis of six key scenes. The findings reveal that evaluative meaning in this film emerges through the orchestrated interaction of linguistic, prosodic, kinesic, and cinematic resources, with patterns of modal convergence, dissonance, and hierarchy shaping interpretive experience.

The concept of multimodal evaluative choreography, proposed as a framework for understanding these phenomena, captures the temporal coordination, hierarchical organization, interpretive guidance, and emergent meaning that characterize multimodal meaning-making in cinema. This framework contributes to the growing field of multimodal discourse analysis by offering an integrative approach that respects the semiotic specificity of different modes while attending to their interaction in time. The framework also extends Relevance Theory by showing how echoic interpretation can operate across multiple semiotic channels simultaneously.

The study also demonstrates the value of ELAN as a tool for multimodal film analysis, offering a replicable methodological framework that can be extended to other films and other research questions. As digital tools for multimodal analysis continue to develop, the integration of automated annotation with manual analysis promises new possibilities for understanding how films construct meaning through the orchestration of multiple semiotic resources. Future developments in machine learning and computer vision may enable automated annotation of some modal features, allowing researchers to scale up multimodal analysis to larger corpora.

In the broader context of linguistics and Cuban cinema studies, this analysis reveals the sophistication of the movie *Pon tu pensamiento en mí* as a site for exploring questions of representation, interpretation, and evaluation. The film's self-reflexive framing, its complex character dynamics, and its layered evaluative stances exemplify the aesthetic of indirection that characterizes much Cuban cinema of the 1990s. Through multimodal analysis, we can begin to appreciate how this cinema achieves its distinctive effects, not through explicit statements but through the artful choreography of semiotic resources that guide viewers to nuanced interpretive engagements.

For linguistics and film studies alike, the challenge remains to develop analytical frameworks and tools adequate to the complexity of multimodal texts. This study suggests that the integration of systematic annotation, theoretical sophistication, and close analysis of specific films offer a productive path forward. As the field of multimodal discourse analysis continues to mature, the study of how films construct evaluative meaning will remain a vital site for understanding the broader question of how meaning emerges from the interplay of multiple semiotic resources in contemporary communication.

References

- [1] Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- [2] Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- [3] Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). John Benjamins.
- [4] Chanan, M. (2004). *Cuban cinema*. University of Minnesota Press.
- [5] van Leeuwen, T. (1999). *Speech, music, sound*. Macmillan.

- [6] Machin, D. (2007). *Introduction to multimodal analysis*. Hodder Arnold.
- [7] Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.
- [8] Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- [9] Steen, F. F., Hougaard, A., Joo, J., Olza, I., Pagán Cánovas, C., Pleshakova, A., Ray, S., Uhrig, P., Valenzuela, J., Woźny, J., & Turner, M. (2018). Toward an infrastructure for data-driven multimodal communication research. *Linguistics Vanguard*, 4(1), 20170041.
- [10] Brugman, H., & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation* (pp. 2065–2068). European Language Resources Association.
- [11] Bateman, J. A., & Schmidt, K. H. (2012). *Multimodal film analysis: How films mean*. Routledge.
- [12] Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. Equinox.
- [13] Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- [14] Müller, C. (2018). Gesture and multimodal communication. In A. Bar-On & D. R. Ravid (Eds.), *Handbook of communication disorders* (pp. 569–582). De Gruyter Mouton.
- [15] Bordwell, D., & Thompson, K. (2013). *Film art: An introduction* (10th ed.). McGraw-Hill.
- [16] Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical pragmatics* (pp. 295–318). Academic Press.
- [17] Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.
- [18] Yus, F. (2016). *Humour and relevance*. John Benjamins.
- [19] Burczynska, P. (2018). *Investigating the multimodal construal and reception of irony in film translation: An experimental approach* [Doctoral dissertation, University of Manchester].
- [20] Smith, C. A. (2022). *The multimodal realisations of the trope of verbal irony through the lens of the pragmeme of insolence: A study of four scenes from Buffy the Vampire Slayer S05*. Paper presented at Multimodal Tropes in Contemporary Discourse, Université Lyon 3.
- [21] Hernández-Reguant, A. (2009). Cuba's alternative cinema. In A. L. Martín & A. Hernández-Reguant (Eds.), *Cuba: Idea of a nation displaced* (pp. 151–166). State University of New York Press.
- [22] Stock, A. M. (2009). *On location in Cuba: Street filmmaking during times of transition*. University of North Carolina Press.
- [23] Sotto, A. (Director). (1995). *Pon tu pensamiento en mí* [Film]. ICAIC.