

Synergistic Construction of Expectational Stance and Multimodal Evaluative Meaning: A Linguistics Case Study of the Opus *Pon tu pensamiento en mí*

Bin Dang^{1,2*}

¹China Shaanxi Translators Association, Xi'an 710068, China

²Universidad Central "Marta Abreu" de Las Villas

*Corresponding author: Bin Dang, bd6298@outlook.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In both everyday talk and artistic storytelling, speakers rarely put their attitudes or evaluations into words in any direct way. Bringing together ideas from expectancy expectation-based sociopragmatics, and evaluative meaning research, the study builds a three-part model: Expectation Monitoring, then Cross-Modal Signal Coordination, and finally Stance Emergence. What emerges is that irony, a classic case of an expectational stance strategy, does subtle interpersonal work in the film. Rather than fraying social bonds, it tightens closeness, protects face, and stirs empathy. It acts as a bonding agent instead of a disruptor. The paper also draws attention to the multimodal evaluative meaning set out by Bin Dang^[1]. In that view, evaluative meaning is not a fixed property parked in one channel but forms piece by piece in time, through cross-channel evaluative tension and synergy. This insight turns out to be important for grasping how stance actually materializes in multimodal settings. Fresh experimental work backs up this cross-modal predictive framework: gestural discourse markers can raise discourse-level expectations even without words^[2], and iconic co-speech gestures are actively recruited to predict what is coming next, a pattern confirmed by both behavioral and electrophysiological measures^[3]. Meanwhile, computational studies show that expectancy violation itself, treated as a “surprise” signal, works as an efficient device for steering communication, matching the performance of more complex Theory of Mind approaches without demanding heavy mentalizing^[4].

Keywords: Expectational stance; Multimodality; Three-stage theoretical model; Evaluative meaning; Irony; Speech act; *Pon tu pensamiento en mí*; Emergent mechanism

Online publication: June 12, 2026

1. Introduction: How Does Stance "Emerge": An Expectation-Driven Perspective

In ordinary conversation, people do more than swap facts. They keep sending signals about how they view what is being said, how they see themselves relative to the listener, and how they want the back-and-forth to unfold. Older labels for this layer of meaning include “tone,” “modality,” “implicature,” or simply “what

is meant beyond what is said.” Whether someone remarks “What lovely weather” in the middle of a storm, or says “It’s so tidy in here” when stepping into chaos, a gap opens between the words on the surface and the speaker’s actual stance. That gap, more than explicit statements, creates the room for subtle human communication.

How do people bridge it? How does implicature come into being in the minds of those talking? For decades, linguists and philosophers leaned on Grice^[5] and the idea that implicature appears when conversation rules are deliberately flouted, forcing the hearer to draw an inference. This approach treats implicature as a retrievable object and concentrates on sorting it into types and describing the inferential steps. Cognitive science and interactional linguistics have revealed the shortcomings of this view. It pictures understanding as a fairly still, single-pass “decode-then-infer” routine. It struggles to capture how meaning is built in real time, by embodied minds, as talk unfolds. More importantly, it skips over a prior question: before a violation is picked up and inference starts, what conditions push implicature into the open in the first place?

Recent work points in a different direction. Understanding irony or indirect requests, it seems, draws on a cognitive capacity that is partly separable from understanding intonational contrasts or reasoning from world knowledge^[6]. This suggests that implicature is not one uniform thing, its comprehension calls on particular socio-cognitive underpinnings. The spotlight has begun to shift from the inferential act itself toward the cognitive soil that feeds it: socio-cognitive expectation. New experiments reinforce this reorientation. Ter Bekke *et al.*^[3] used Cloze tasks and EEG to show that comprehenders tap into the meaning carried by iconic co-speech gestures, gestures that tend to arrive before the related words, to predict upcoming speech. Those gestures cut alpha and beta power during anticipatory pauses and reduced the N400 amplitude, signs of active prediction, and eased semantic processing. Along similar lines, Scholman and Laparle^[2] found that gestural discourse markers can, on their own, call up discourse-level expectations in comprehenders, even with no lexical cue, though such gestural signals are weaker than words like “but” or “therefore.” Both studies underline that expectancy processing is deeply cross-modal.

Expectation works as a basic function of the human cognitive system across many levels of language. Listeners predict upcoming words, syntactic structures, referents guided by the implicit causality of verbs, and even logical relations between sentences cued by discourse connectives^[7]. The bigger question, then, is whether communicators also form and depend on expectations about how others will keep to social norms and take up particular attitudes. When a speaker’s multimodal output, for example, words, prosody, facial expression, gesture, diverges from such expectations, does implicature, or better, “expectational stance,” quietly rise from that rift? A computational angle on this query comes from Klomberg *et al.*^[4]. Their work indicates that a “Surprise” model, built on nothing more than expectancy violations, can compete with elaborate Theory of Mind strategies for guiding novel communication. It yields human-like message patterns while cutting the receiver’s cognitive load. Expectancy violation thus looks like a flexible, cognitively cheap signaling resource.

For these reasons, this paper advances a central thesis: implicature is not a pre-wrapped, after-the-fact “implied object.” It is an expectational stance, an attitudinal meaning that grows step by step out of a dynamic “expectation-violation-attunement” sequence, carried by the real-time interplay of multiple channels: language, prosody, facial expression, and gesture. The word “emerge” stresses that the whole is more than the sum of its parts and cannot be fully foreseen by looking at any single channel alone; it is born from cross-channel interaction and the temporal shape of the communicative event.

To lay out this perspective, the paper draws on the Cuban film *Pon tu pensamiento en mí* (Arturo Sotto, 1995) as cross-cultural multimodal material. The film shows seven scribes telling the story of the lead actor in their play, who portrays Christ. The actor performs tricks and fake miracles; the people take him for a new Messiah; he denies it, but the crowd needs to believe^[8]. The narrative frame itself thus constitutes a meta-commentary on expectation, that is, how the people's hope for a savior is fed by a performance, shaken by denial, and left wavering. The film's dense multimodal irony offers rich natural data for studying how expectational stance forms across channels. Crucially, Bin Dang's^[1] analysis of multimodal evaluative meaning in this film supplies a key tool. His expound shows how evaluative meaning is progressively built through the interaction of verbal, visual, and auditory channels, rather than being stored in any one of them beforehand.

The paper continues as follows. Section 2 traces the theoretical roots from expectancy linguistics to multimodal pragmatics and clarifies the conceptual origins of expectational stance. Section 3 constructs the three-stage model: Expectation Monitoring—Cross-Modal Signal Coordination—Stance Emergence. Section 4 takes representative scenes from the film as case studies and examines how expectational stance plays out in multimodal contexts, with particular attention to irony as a constructive interpersonal move. Section 5 proposes the Stance Attunement Hypothesis and discusses its implications. Section 6 wraps up and outlines future directions.

2. Theoretical Genealogy: From Expectation-based Linguistics to the Confluence of Multimodal Pragmatics

The idea of expectational stance does not come from nowhere. It grows out of several intersecting lines of research. Expectation-based Linguistics offers a cognitive and dynamic angle; multimodal sociopragmatics brings analytical tools for handling multiple channels; evaluative meaning theory supplies the semantic grounding for attitudinal meaning. This section reviews these three streams and shows how they merge in the notion of expectational stance.

2.1. Expectation-based Linguistics: Expectation as the Cognitive Engine of Language Processing

Over the last two decades or so, expectation has moved to the center of work in psycholinguistics and cognitive linguistics. Language users form expectations not only at the word or phrase level, which predicts a verb's arguments from its semantics, for instance, but also at the discourse level. Rohde and Kehler^[7], using eye-tracking, showed that when listeners hear "because" or a particular implicit-causality verb, they launch anticipatory eye movements toward a "cause" or "effect" referent before the sentence is finished. Expectations at the pragmatic level are thus built on the fly.

More recently, researchers have begun to uncover how multimodal cues feed linguistic expectation. Ter Bekke *et al.*^[3] found that listeners in natural conversation use the speaker's gestures to predict words ahead of time; gestures often arrive before the matching speech, giving an advance signal. EEG data showed that gestures not only nudge behavioral prediction but also modulate the N400, suggesting that they lighten the semantic load for what follows. This provides a strong demonstration that expectancy processing operates across channels. Extending this line to discourse, Scholman and Laparle^[2] ran a multimodal continuation study. Participants watched videos of a speaker telling stories containing contrast, list, or exception relations. Even when the audio cut off before the second relational element, gestural discourse markers

allowed comprehenders to figure out the intended relation. Gestures can independently stir discourse-level expectations.

Around the same time, Prieto *et al.* ^[9] put forward the notion of prosody as a grammatical component that spans both spoken and visual signals. Prosody, they argue, is realized not only through sound but also through visible signals like gestures and head movements, forming a “modality-neutral” prosodic system. This idea provides a helpful analytical frame for understanding how expectational stance can be jointly realized across channels. Zellers ^[10] further shows how conversational participants weave together prosodic and visual cues to anticipate what others will do, keeping interaction smooth.

Taken together, this work points to a conclusion: expectation is not some side effect of language processing but its cognitive engine. Yet studies to date have mostly stayed at lower levels, for example, lexical, syntactic, and have not devoted enough attention to higher-level socio-pragmatic expectations: the expectations communicators hold about each other’s attitudes, intentions, and interpersonal positioning. That gap is what motivates the introduction of expectational stance in this paper.

2.2. Multimodal pragmatics: The distribution of implicature across channels

For a long time, pragmatics kept its analytical focus almost entirely on the linguistic channel. More recently, Haugh *et al.* ^[11], in *The Cambridge Handbook of Sociopragmatics*, stated plainly that sociopragmatic meaning is carried multimodally, through prosody, gesture, and other non-verbal signals. This shift matters: implicature is no longer taken to be the exclusive property of the linguistic channel but is understood to be spread across several semiotic channels. Researchers need to adopt an integrated view.

When it comes to irony research, a multimodal perspective is especially important. Attardo *et al.* ^[12] did pioneering work on the multimodal markers of irony and sarcasm, mapping systematic features at the levels of facial expression, gesture, and prosody. Smith ^[13] examined ironic scenes from the TV series *Buffy the Vampire Slayer* through the lens of the “pragmeme of insolence,” paying close attention to how prosodic contours, gesture, and facial expression work cooperatively. Yana and Elena’s ^[14] study of gestures and prosodic structure in laboratory-elicited ironic speech further confirmed that paralinguistic features play a critical part in the perception of irony. These studies together show that irony is not simply a linguistic figure but a prototypical cross-channel speech act, its meaning depends on integrating information from several semiotic channels.

2.3. Evaluative meaning theory: Linguistic and visual realizations of attitudinal meaning

Evaluative meaning: the way speakers assign value judgments and emotional weight to what they say, which forms the semantic core of stance. Bin Dang’s ^[1] work on multimodal evaluative meaning in *Pon tu pensamiento en mí* pushes significantly beyond traditional appraisal theory. It proposes a viewpoint built around the notion that evaluative meaning is not a static property of any single channel but a dynamic process that emerges over time through “inter-channel evaluative tension” and “cross-modal evaluative synergy.”

In more detail, Dang’s viewpoint includes these key ideas: (1) **Evaluative Tension**: a mismatch in evaluative valence between channels (say, language vs. facial expression, or prosody vs. gesture) that acts as the cognitive trigger for generating complex evaluative meanings; (2) **Cross-Modal Evaluative Synergy**: the way signals from different channels unfold and modify each other along the timeline, eventually producing an overall evaluative meaning that goes beyond any single channel; (3) **Temporal Construction of Evaluative Meaning**: the fact that evaluative meaning is not delivered in an instant but becomes progressively clearer as

the receiver processes multimodal information in real time.

This model provides a direct theoretical footing for the present paper. If evaluative meaning emerges through cross-modal synergy, then the role of expectation as the cognitive frame that guides this synergy becomes especially salient. At this juncture, the paper weaves together the cognitive-dynamic perspective of expectation-based linguistics with the multimodal evaluative meaning model to develop the integrated concept of expectational stance.

2.4. New horizons in irony research: From “destructive” to “constructive”

Irony has often been treated as a negative use of language, implying a breach of the Cooperative Principle and typically read as sarcasm, ridicule, or criticism. In recent decades, however, research has increasingly highlighted another side: irony can serve as a positive social strategy.

Dews *et al.* ^[15], across a series of experiments, found that ironic criticism does less damage to the speaker-addressee relationship than literal criticism does; irony makes the speaker look less angry and more in control. Jorgensen ^[16] showed that sarcastic irony can have a face-saving function, making the speaker appear less rude and unfair, especially with mild criticism. More notably, some researchers ^[17] have argued from a Relevance Theory standpoint that irony works as a discourse strategy for boosting social intimacy. When “mild irony” is exchanged among close friends or colleagues, it can strengthen intimacy and solidarity rather than insult.

Nikabadze ^[18] looked at the socio-pragmatic functions of irony on social media and expanded the picture. Besides humor, social status enhancement, and emotion regulation, she identified an “empathic call for change” function, which ironic expression can draw shared attention to a desirable state of affairs. Mękarska and Witek ^[19] approached irony from a Speech Act Theory perspective and developed a formalized model, proposing that ironic utterances are social actions performed by invoking socially recognized procedures. This moves the discussion from irony as a mere rhetorical flourish to irony as a socially constitutive discursive practice.

What emerges from these studies is that irony is not a one-dimensional negative trope but a flexible sociopragmatic strategy. In close relationships and particular contexts, irony can soften criticism, protect face, attune emotions, and reinforce bonds. This body of work forms a crucial empirical base for the Stance Attunement Hypothesis presented later.

3. Theoretical construction: A three-stage emergence model of expectational stance

Drawing on the theoretical lineage traced above, this section proposes an integrative model meant to describe how expectational stance emerges in multimodal contexts. The model has three linked stages: Expectation Monitoring, Cross-Modal Signal Coordination, and Stance Emergence.

3.1. The hierarchical nature and social embeddedness of expectation

Expectation in language processing is layered rather than flat. At least four levels can be distinguished: (1) **Phonological-Lexical Expectation**: anticipating upcoming sounds or words from phonological sequences and co-occurrence probabilities; (2) **Syntactic-Semantic Expectation**: predicting syntactic category and semantic role from verb argument structure and semantic selection; (3) **Discourse-Logical Expectation**: predicting coherence relations between sentences from discourse connectives and rhetorical relations;

(4) **Pragmatic-Social Expectation:** predicting the speaker’s intentions, attitudes, and affective positioning from social norms, communicative conventions, and interpersonal relationships.

This paper puts the fourth level at the center: pragmatic-social expectation, the “expectation” within “expectational stance.” Expectations at this level are thoroughly social. They are rooted in a person’s internalized grasp of social norms, communicative conventions, situation-specific behavioral scripts, and the personal style of those they talk with. They include: **Situational Frame Expectations** (about the ground rules and scripts for the current type of interaction); **Relational and Identity Expectations** (about the right degree of politeness and directness given the relationship, power gap, and closeness); **Behavioral Sequence Expectations** (about what kind of response a particular turn should get next); and **Cross-Modal Coordination Expectations** (about what prosodic patterns, facial expressions, and gestures normally accompany certain verbal content).

These expectations work like a cognitive map. They guide where attention goes, help filter relevant channel signals, and supply an interpretive frame for combining those signals.

3.2. Expectational stance and “evaluative tension”

When incoming multimodal signals match the active pragmatic-social expectations, communication flows easily, and meaning comprehension demands little extra cognitive work. But when signals in one or more channels break from expectations, what Dang ^[1] calls “evaluative tension” sets in, a state of inconsistency or dissonance in evaluative valence across channels.

This evaluative tension acts as the cognitive trigger for expectational stance to emerge. For example, suppose the verbal channel delivers literal praise (“You did a great job”), while the prosodic channel sounds flat and uninterested and the visual channel offers a dismissive glance. Then a clear valence conflict emerges across the three channels. That conflict is not noise; it is a signal. It alerts the receiver that the surface linguistic meaning cannot be taken at face value and that a deeper interpretation is needed.

How strong the evaluative tension feels depends on how salient the cross-channel inconsistency is, how closely synchronized its presentation across channels is, and how firmly the underlying expectation is held. The more striking the mismatch, the tighter the temporal alignment, and the stronger the expectation, the greater the tension and the more pronounced the stance emergence process that follows.

3.3. The three-stage emergence model

Bringing the above points together, the model runs as follows:

Stage One: Expectation Monitoring. Before and during communication, the receiver keeps activating and updating a set of pragmatic-social expectations. These draw on contextual information, social knowledge, and familiarity with the speaker. They form a dynamic, layered framework that stretches from the situational frame down to cross-modal coordination.

Stage Two: Cross-Modal Signal Coordination and the Generation of Evaluative Tension. The receiver processes information in real time from multiple channels: language, prosody, facial expression, gesture, and holds it up against the current expectation framework. When signals align across channels and match expectations, comprehension goes ahead automatically. When cross-channel inconsistency appears what Dang ^[1] termed evaluative tension, the cognitive system is roused and shifts into a deeper processing mode.

Stage Three: Stance Emergence. Pushed by evaluative tension, the receiver launches a kind of “cognitive harmony restoration.” They seek an explanation that can make sense of the contradictory multimodal signals

and restore cognitive equilibrium. This explanation is not simply read off from any single channel. It emerges as the best hypothesis that fits the interplay and contradiction among the multimodal signals. That is the final form of expectational stance, an attitudinal meaning built through the integrative interpretation of multimodal signals, with the expectation framework as a guide.

The main contribution of this model is to recast the static notion of implicature from traditional pragmatics into a dynamic, temporal process of stance emergence, and to place the synergy and tension of multimodal information at the heart of that process. Implicature is no longer a fixed proposition hidden beneath the language; it is a meaning generated in real time through cross-channel interaction, steered by expectation.

4. Case analysis: Multimodal expectational stance in *Pon tu pensamiento en mí*

This section uses the Cuban film *Pon tu pensamiento en mí* as multimodal data, applying the model proposed above to analyze how expectational stance is realized across channels, with special attention to irony as a constructive interpersonal strategy. The film depicts seven scribes recounting the story of an actor playing a false Christ. The actor performs tricks and fake miracles, the populace takes him for the new Christ, he denies being Jesus, but the masses need to believe. The scribes manipulate the people's faith, spinning a tale inside an allegory of love, denial, and conspiracy^[8]. This narrative frame itself amounts to a meta-narrative about expectation: how the people's expectation of a Messiah is fed by the actor's performance, unsettled by his denial, and left vacillating between belief and doubt.

4.1. The meta-theme of “expectation” in the film’s narrative

The narrative design of *Pon tu pensamiento en mí* reads as a meta-commentary on expectation. The film begins with the scribes arguing about how to tell the story of Jesus; their talk is itself a meta-text about managing expectations. One scribe says: “Deberíamos conservar esa textura, una historia documental en blanco y negro.” (We should preserve that texture, a black-and-white documentary story.) Another pushes back: “Prefiero el documental, nos hace mucho más libres. Podemos jugar con el tiempo, con la imagen que se retuvo y con la que nosotros mismos nos inventamos.” (I prefer the documentary, it makes us much freer. We can play with time, with the retained image and with the one we ourselves invent.) This exchange reveals a central point: narrative itself is about managing what the receiver expects. By choosing how to tell the story, controlling when information is released, and orchestrating what is seen and heard, the narrator continually shapes and adjusts the audience's expectations.

Another snippet of dialogue among the scribes further illuminates the link between expectation and social power. One scribe comments: “Habrà que enriquecer la historia, adornarla un poco, sin que se vuelva oscura.” (We'll have to enrich the story, embellish it a little, without letting it become obscure.) Another puts it even more plainly: “Algo se sabe y lo que no, lo improvisamos o lo plagiamos.” (What we know, we use; what we don't, we improvise or plagiarize.) This awareness of narrative power, who sets expectations, who gets to break them, who controls their content, is exactly the deeper theme the film explores. In this sense, the whole film can be viewed as an allegory about expectation and its multimodal construction.

“Habrà que enriquecer la historia, adornarla un poco, sin que se vuelva oscura” (“We'll need to enrich the story, embellish it a little, but don't let it become obscure.”). The expectation here marks the speaker's anticipation of where the narrative might go wrong and a move to manage that. Its Chinese translation “但别

让它变得晦暗” (but don't let it become obscure), the word “别” (don't) is significant because it functions as a key expectation marker. After calling for enrichment and adornment, the speaker foresees that too much embellishment could cloud the story; “别” steps in early to cancel that negative potential. It works as a counter-expectation device, blocking the inference that decoration naturally slides into obscurity. The Spanish “sin que se vuelva oscura” (without it becoming dark) uses the subjunctive in a parallel way to ward off an anticipated adverse result. Thus, “别” encodes a deeper expectation: aesthetic elaboration must not sacrifice clarity, a sign of rhetorical self-awareness. The primary word of expectation is, without doubt, “别.”

4.2. Multimodal features of irony: The synergy of language, prosody, and visuality

What makes the film especially valuable for this study is its many ironic scenes, delivered in a strongly synergistic multimodal fashion. These scenes offer exemplary material for analyzing how expectational stance emerges.

Irony's multimodal signature can be described as a “misaligned synergy” along three dimensions. Linguistically, irony usually entails a split between what is said and what is meant. In the film, as the scribes talk about the Jesus story, one remarks: “Muy bien, nosotros somos los nuevos Escribas, comencemos nuestra santa versión de los hechos.” (Very well, we are the new Scribes, let us begin our holy version of the facts.) The word “holy” here carries a clear ironic weight; these scribes are not recording sacred history but making up a story.

But looking only at linguistic deviation falls short of capturing the full picture. Prosodically, irony often brings marked pitch contours, an unusually wide (exaggerated) or unusually narrow (flattened) pitch range, and characteristic patterns of contrastive stress. Visually, irony is frequently tagged by specific facial expressions, that is, brief lifting or lowering of the mouth corners, slight eyebrow raises, deliberate gaze aversion, and by gestures such as open palms, shrugs, or exaggerated pointing. Attardo *et al.* ^[12] systematically documented these multimodal markers, and Smith ^[13] confirmed that the interpretation of irony turns on the synergistic interplay of prosodic contours, gestures, and facial expression.

Bin Dang ^[1], in his study of the same film, observes that evaluative meaning is built up precisely through the tension and synergy of these cross-channel signals. The inconsistency among the linguistic, prosodic, and visual channels is not a breakdown of meaning but the very mechanism that produces it; this inconsistency prompts the audience to revise their expectations and lets the ironic stance surface.

4.3. The positive functions of irony: Social adhesive, not relationship disruptor

The habitual linking of irony to sarcasm and criticism overlooks how it works positively inside close relationships. The ironic expressions in *Pon tu pensamiento en mí* neatly illustrate irony's constructive role as a social bonding agent.

First, irony can protect face and soften criticism. Jorgensen ^[16] found that ironic criticism causes less relational damage than literal criticism because it blunts the edge. Dews *et al.* ^[15] confirmed that ironic criticism is less harmful than its literal counterpart and that it makes the speaker seem less angry and more self-possessed. Characters in the film use irony rather than direct criticism to achieve precisely this effect. Irony wraps criticism in humor, reducing face threat while still getting the critical point across.

Second, irony can boost intimacy and group identity. Researchers ^[17] have argued from a Relevance Theory perspective that mild irony, when shared among close friends, can increase intimacy and solidarity. Grasping irony requires shared background knowledge and common values between speaker and hearer;

that very sharedness marks off a cognitive community. Those who catch each other's irony become insiders; irony thus acts as an implicit marker distinguishing in-group from out-group. Nikabadze ^[18], studying irony on social media, likewise singled out functions like humor, status enhancement, emotion regulation, and the "empathic call for change." The ironic exchanges among the scribes in the film reinforce their internal solidarity as a narrative community.

Finally, irony can regulate emotional distance and keep relationships elastic. Mękarska and Witek ^[19] treat ironic utterances as social actions carried out by invoking socially recognized procedures, emphasizing irony as a socially constitutive discursive practice. The ambiguity and deniability built into irony give communicators a buffer zone for relational attunement. A speaker can put forward a certain attitude while keeping open the option of retreating behind "I didn't mean it that way." This elasticity makes irony a useful tool for navigating complicated social ties. Characters in the film deploy irony to transmit nuanced affective stances, putting this relational flexibility to work: they can maintain some distance while signaling closeness, and express criticism while preserving goodwill.

4.4. The cross-cultural dimension of expectational stance

As a Cuban film, *Pon tu pensamiento en mí* offers a special vantage point for looking at the cross-cultural reach of expectational stance. Expectation, especially at the socio-pragmatic level, is tightly woven into a culture's communicative conventions and social norms. Different cultures hold different default expectations about what counts as a normal tone, when irony fits, and how much is acceptable.

The ironic expressions in the film show both cross-cultural commonalities, which are the basic mechanisms of multimodal synergy that may rest on shared cognitive foundations, and cultural particularities. The distinctive Cuban humorous tradition, the familiar modes of political allegory, and the complex tension between the religious and the secular all give the film's irony a unique cultural imprint. This specificity highlights the need to analyze expectational stance within its sociocultural soil. Expectation itself is, after all, a product of sociocultural internalization.

The value of Bin Dang's ^[1] study of this film lies partly in showing that the cross-modal construction of evaluative meaning is not a universalist cognitive operation but a semiotic practice shaped by specific cultural contexts. The emergence of evaluative meaning is both cognitive and social, happening at the intersection of individual cognitive systems and sociocultural semiotic systems.

5. Theoretical advancement: The "Stance Attunement Hypothesis"

On the basis of the theoretical construction and case analysis above, this paper proposes the **Stance Attunement Hypothesis** as an original way of understanding the nature and function of expectational stance.

5.1. Statement of the hypothesis

The Stance Attunement Hypothesis holds that expectational stance is a finely tuned mechanism that evolved in human communication. Its main job is not to transmit extra propositional content but to manage and regulate, in a dynamic and fine-grained way, the psychological states, social relationships, and emotional distance between those talking. Expectation provides the baseline for attunement; expectation violation acts as the switch that turns attunement on; and the stance meaning that emerges through cross-modal synergy is the flexible medium through which attunement is carried out.

The hypothesis rests on four sub-propositions:

First, expectation is the baseline for attunement. Communicators continuously sustain expectations during interaction about how the other party should act and what attitudes they should express. These expectations make up a kind of default setting. When interaction runs according to expectation, things go smoothly with little extra cognitive effort. Expectation is not a bolt-on extra but the foundation.

Second, expectation violation is the activation signal for attunement. When incoming multimodal signals depart from expectations, whether through a deviation inside the linguistic channel or through cross-channel mismatch, the cognitive system flips from automatic processing to controlled processing and launches deeper meaning construction. Expectation violation is not a glitch but an attunement opportunity. Recent computational work backs this up. Klomberg *et al.* ^[4] showed that a Surprise model, resting solely on expectancy violations, can rival complex Theory of Mind strategies in guiding novel communication, producing human-like message patterns and faster receiver responses (indicating lower cognitive load). Expectancy violation thus seems to be a parsimonious and evolutionarily adaptive signaling mechanism in its own right.

Third, stance emergence is the process of attunement realization. Spurred by expectation violation, the receiver, through interpreting multimodal signals in an integrated way, constructs an attitudinal meaning that can restore cognitive equilibrium and expectational stance. This meaning is not stored in any single channel beforehand but arises progressively in time, driven by evaluative tension.

Fourth, the regulation of social relations is the ultimate purpose. The fundamental function of expectational stance is to regulate the social relationship, which is used to protect face, enhance intimacy, soften criticism, express empathy, and test relational boundaries. Implicature exists not because people cannot be direct but because the complexity of social relationships demands a more elastic, nuanced attunement tool.

5.2. Relationship to existing theories

The Stance Attunement Hypothesis does not reject existing theories but integrates and advances them. It operationalizes Grice's ^[5] "violation" as a violation of specific pragmatic-social expectations. It fleshes out the Relevance Theory process of "searching for optimal relevance" as dynamic integration and hypothesis testing within the multimodal information stream, guided by expectation. It extends the illocutionary force of Speech Act Theory to stance emergence in cross-channel contexts. At the same time, the hypothesis aligns closely with Bin Dang's ^[1] multimodal evaluative meaning model. If Dang's model describes "how" evaluative meaning emerges in multimodal interaction, the Stance Attunement Hypothesis answers "why" it needs to emerge that way: social relationships require this flexible, elastic, temporally unfolding mode of meaning realization.

Moreover, the Stance Attunement Hypothesis offers a unified framework for making sense of the recent positive turn in irony research ^[15-19]. Irony can serve multiple positive functions, such as face maintenance, intimacy building, and emotion regulation. This is precisely because it is a highly developed tool for expectational stance attunement. Through the tension between literal meaning and genuine attitude, it opens a flexible negotiation space between criticism and care, distance and closeness, seriousness and play.

5.3. Theoretical implications and future directions

Putting forward the Stance Attunement Hypothesis signals a shift in the study of implicature. It moves

from asking “What is the implicature of this utterance?” toward asking “In this interaction, what expectation was violated, and how? Which signals from different channels combined in what temporal rhythm to let this stance emerge, and how did the relational state between the interactants change?” This shift lifts the study of implicature from static taxonomy to dynamic process, broadens the lens from language alone to multimodal integration, and extends its reach from individual cognition to the terrain of social relationship attunement.

Future work can move in several directions. First, Cross-Cultural Validation: Computational Modeling Applications: Current large language models still show serious limits in understanding irony and implicature^[20]. The Stance Attunement Hypothesis and Bin Dang’s^[1] viewpoint of multimodal evaluative meaning could provide theoretical guidance for building dialogue systems that have something like social intelligence.

5.4. Affirmative Irony, Expectation Resonance, and the Positive Face of Social Bonding

This viewpoint (or model) of expectational stance finds convergent support in recent research on affirmative irony and multimodal expectation resonance. Dang proposes an Expectation Resonance framework in which affirmative irony operates through a cognitive functional mapping: $E_{set} = f(M_v, M_g)$, where M_v represents verbal input and M_g covers gestural and prosodic signals. Resonance emerges precisely when cross-modal discrepancy pushes the comprehender from literal reading toward ironic reinterpretation, a process driven by what Dang^[1] terms evaluative tension. The strength of resonance hinges on a perceptual threshold: below it, inconsistency signals speaker ineptitude; above it, the hearer recognizes intentional ironic play^[21]. This resonance mechanism aligns closely with the Stance Attunement Hypothesis articulated above. Crucially, affirmative irony in contexts such as Cuban cinema does not fray social bonds. Instead, as playful humour research on Cecilia demonstrates, joyful non-malicious wit serves to lower affective barriers, sharpen pragmatic awareness, and foster in-group solidarity^[22]. These findings corroborate the present paper’s claim that irony, when understood through the lens of expectational stance, functions not as a disruptive force but as a finely calibrated instrument for face protection, empathy cultivation, and the attunement of emotional distance.

6. Conclusion

Implicature is not the static, submerged underside of the linguistic iceberg. It is more like the bright spray thrown up as conversational currents run over the reefs of social cognition, the briefly visible pattern that appears when speech, tone, face, and hands move together along a multi-dimensional timeline. Through theoretical construction and case analysis, this paper has argued that implicature is, at bottom, an expectational stance. It is generated step by step through a three-stage process of expectation monitoring, cross-modal signal coordination, and stance emergence, and its core function is the regulation of social relationships.

The Stance Attunement Hypothesis pulls these ideas into a single framework: expectation is the baseline for attunement, expectation violation is the activation signal, stance emergence is the process, and the regulation of social relationships is the ultimate purpose. Converging evidence from multimodal discourse processing research and computational modeling gives solid support to this picture. Scholman and Laparle^[2] showed that gestural discourse markers can independently generate discourse-level expectations. Ter Bekke *et al.*^[3] demonstrated that iconic gestures facilitate predictive semantic processing at the neural level. Klomberg *et al.*^[4] provided computational evidence that expectancy violation

is a parsimonious but powerful signaling mechanism. All these findings converge on the same insight: expectation is not an epiphenomenon of communication but its foundational cognitive architecture. Within this architecture, irony stops looking like a negative linguistic deviation and starts looking like a positive, subtle social attunement strategy. It jump-starts cognitive processing through evaluative tension, achieves meaning emergence through cross-modal synergy, and ultimately plays an irreplaceable part in protecting face, enhancing intimacy, and regulating emotional distance.

Bin Dang's ^[1] research on multimodal evaluative meaning in *Pon tu pensamiento en mí* has provided direct theoretical support for this paper. His notion of cross-modal evaluative synergy reveals how evaluative meaning is progressively constructed as verbal, visual, and auditory channels interact, not as something that sits pre-formed in a single channel. Building on that foundation, this paper has woven together the cognitive-dynamic perspective of expectation-based linguistics with the theory of multimodal evaluative meaning, offering an integrated model of expectational stance emergence.

Continued exploration of expectational stance will not only deepen understanding of the nature of language. It will also throw light on how human beings, through semiotic means, live socially, how, beyond the words spoken, in the weaving and attuning of expectations, we build the real world of our mutual relations.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Dang B, 2026, Constructing Evaluative Meaning Across Modes: A Case Study of *Pon tu pensamiento en mí*. *Advances in Humanities Research*, 13(4): 45–63.
- [2] Scholman M, Laparle S, 2025, Can Gestures Speak Louder Than Words? The Effect of Gestural Discourse Markers on Discourse Expectations. *Discourse Processes*, 62(6–7): 457–478.
- [3] ter Bekke M, Drijvers L, Holler J, 2025, Co-Speech Hand Gestures Are Used to Predict Upcoming Meaning. *Psychological Science*, 36(4): 237–248.
- [4] Klomberg B, Schakel M, Spronck M, 2025, Expectation Violations as an Effective Alternative to Complex Mentalizing in Novel Communication. *iScience*, 28(7): 112795.
- [5] Grice HP, 1975, Logic and Conversation. In Cole P, Morgan JL, (Eds.), *Syntax and Semantics 3: Speech Acts*, Academic Press, New York, 41–58.
- [6] Floyd S, Jouravlev O, Poliak M, et al., 2025, Three Distinct Components of Pragmatic Language Use: Social Conventions, Intonation, and World Knowledge-Based Causal Reasoning. *Proceedings of the National Academy of Sciences*, 122(50): e2424400122.
- [7] Rohde H, Kehler A, 2014, Anticipatory Looks Reveal Expectations About Discourse Relations. *Cognition*, 133(3): 667–691.
- [8] Sotto A, (Director), 1995, *Pon tu pensamiento en mí* [Film], Instituto Cubano del Arte e Industria Cinematográficos (ICAIC) / Tabasco Films S.A.
- [9] Prieto P, Esteve-Gibert N, Shattuck-Hufnagel S, 2025, Towards a Novel Conceptualization of Prosody that Accounts for Spoken and Visual Signals. *Gesture*, 2025.
- [10] Zellers M, 2026, Multimodal Cues to Anticipation in Conversation, Research seminar, Stockholm University.

- [11] Haugh M, Kádár DZ, Terkourafi M, 2021, Gesture and Prosody in Multimodal Communication. In Haugh M, Kádár DZ, Terkourafi M, (Eds.), *The Cambridge Handbook of Sociopragmatics*, Cambridge University Press, Cambridge, 400–425.
- [12] Attardo S, Eisterhold J, Hay J, et al., 2003, Multimodal Markers of Irony and Sarcasm. *International Journal of Humor Research*, 16: 243–260.
- [13] Smith CA, 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 Jean Moulin.
- [14] Yana E, Elena K, 2023, Gestures vs. Prosodic Structure in Laboratory Ironic Speech. In *Speech and Computer*, Springer, 112–125.
- [15] Dews S, Kaplan J, Winner E, 1995, Why Not Say It Directly? The Social Functions of Irony. *Discourse Processes*, 19(3): 347–367.
- [16] Jorgensen J, 1996, The Functions of Sarcastic Irony in Speech. *Journal of Pragmatics*, 26(5): 613–634.
- [17] Lee J, 2020, Practising Irony as a Speaker’s Discourse Strategy for Increasing Social Intimacy, and Relevance. *Journal of English Language and Literature*, 66(4): 103–124.
- [18] Nikabadze M, 2024, The Socio-Pragmatic Function of Irony on Social Media: Using Examples from Johan Deckman’s Art Pictures on Instagram. *Proceedings of Tbilisi State University*, 9: 13–22.
- [19] Mękarska K, Witek M, 2022, Irony as a Speech Action. *Journal of Pragmatics*, 190: 76–90.
- [20] Yue S, Song S, Cheng X, et al., 2024, Do Large Language Models Understand Conversational Implicature? A Case Study with a Chinese Sitcom. *arXiv preprint*. arXiv:2404.19509.
- [21] Dang, B. (2026). Expectation resonance in affirmative irony: A multimodal analysis of verbal humor in Cecilia. *Scientific and Social Research*, 8(2), 45–62. <https://ojs.bbwpublisher.com/index.php/ssr/article/view/14918/13052>
- [22] Dang, B. (2026). Joyful wit on screen: Uncovering playful humour in Cecilia for Spanish heritage language learning. *Journal of Contemporary Educational Research*, 10(1), 22–38. <https://ojs.bbwpublisher.com/index.php/JCER/article/view/14755>

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.