Iranian Evolutionary Educational Psychology Journal
IEEPJ
University of Hormozgan
Online ISSN: 2588 - 4395
Homepage: https://ieepj.hormozgan.ac.ir

Incongruent Prosodic and Gestural Markers Enhance Iranian Children's Capability to Perceive Irony and Sarcasm

Raha Koochacki[1], Mohammad Hossein Sharafzadeh[2]✉, Ameneh Zare[3]

1. Ph.D., Candidate of General Linguistics, Department of Linguistics, Marv. D., Islamic Azad University, Marvdasht, Iran
2. Assistant Professor, Department of Linguistics, Marv. D., Islamic Azad University, Marvdasht, Iran, h.sharafzadeh@iau.ac.ir
3. Assistant Professor, Department of Linguistics, Marv. D., Islamic Azad University, Marvdasht, Iran

Article Info

Article type:
Research Article

Article history:

Received 28 Feb. 2025

Received in revised form 25 Apr. 2025

Accepted 22 May. 2025

Published online 01 Sep. 2025

Keywords:
Irony and sarcasm,
Prosodic-gestural cues,
Audiovisual task,
Mismatching condition,
Developmental pattern

ABSTRACT

Objective: This study examined whether incongruity between multimodal cues—specifically verbal content versus prosodic and gestural signals—enhances children's ability to detect ironic and sarcastic intent. A further aim was to determine the developmental onset of this ability.

Methods: Participants were 44 children aged 5, 8, and 11 years. They completed an audiovisual irony-detection task that included six congruent prompts and six incongruent prompts, designed to contrast negative contexts with positive responses. Incongruent prompts were delivered in three prosodic-gestural conditions: (1) matching, (2) weakly mismatching, and (3) strongly mismatching. Accuracy in identifying irony and sarcasm was recorded across age groups and conditions.

Results: Children across all ages demonstrated significantly higher accuracy in the strongly mismatching condition than in the weakly mismatching and matching conditions. Similarly, performance in the weakly mismatching condition was superior to the matching condition, indicating that greater incongruity between verbal and emotional cues facilitated irony and sarcasm detection. A significant main effect of age was also observed: 11-year-olds outperformed 8-year-olds, who in turn outperformed 5-year-olds, revealing a clear developmental progression. Additionally, correct response scores varied significantly across the three prosodic-gestural cue conditions, with performance highest under strong mismatches.

Conclusions: Findings suggest that children's ability to detect irony and sarcasm is strongly influenced by the degree of incongruity between verbal and emotional signals. This ability develops progressively with age, becoming more reliable by late childhood. The results underscore the importance of multimodal cues in pragmatic language development and the role of prosody and gesture in facilitating comprehension of nonliteral meaning.

Cite this article: Koochacki, R., Sharafzadeh, M. H. & Zare, A. (2025). Incongruent prosodic and gestural markers enhance iranian children's capability to perceive irony and sarcasm. *Iranian Evolutionary Educational Psychology Journal*, 7 (3), 1-25.
DOI: https//doi.org/ 10.22034/7.3.1

Publisher: University of Hormozgan.
DOI: https// doi.org/ 10.22034/7.3.1

Introduction

In daily social interactions, individuals must evaluate and synthesize various sources of information to effectively comprehend others. During the process of perception, listeners consider not only the verbal content of the message (i.e., what the speaker articulates) but also the prosodic and gestural components accompanying it (i.e., the manner in which the speaker conveys the message). These non-verbal multimodal cues assist listeners in deducing insights regarding the emotional and attitudinal states of their interlocutors, such as uncertainty, disbelief, anger, sadness and so on. This interpretative process is essential for effective communication. As noted by Van Lancker (2008: 206), "one of the most significant challenges in psycho- and neuro-linguistics research is to comprehend how these two components (i.e., verbal and emotional) interact within human communication."

Irony and sarcasm represent a type of non-literal language characterized by an incongruity between the literal meaning of an utterance and its intended meaning. While other forms of non-literal language, such as metaphors, exist, irony and sarcasm are arguably the most intricate, as comprehending it necessitates the listener to synthesize various types of information, including contextual markers (such as the specific circumstances and shared beliefs between the speaker and listener) and emotional-attitudinal markers, which are often expressed via prosody and gestures. Previous studies on the interpretation and perception of irony have demonstrated that detecting and comprehending ironic intent significantly depends on the perceptual contrast between the pragmatic context of a statement and its propositional content (Colston, 2002).

Literature review

Several experimental investigations have indicated that the extent of incongruity between the propositional content of an utterance and its situational context (for instance, the comment Well done! uttered in a negative context) correlates with the rates of irony detection (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003). Furthermore, recent research examining how contrasts among contextual, prosodic, and gestural signals influence the comprehension of irony and sarcasm has revealed that adults are more inclined to recognize them in a statement when they are presented with mismatching contextual cues (a) in conjunction with mismatching prosodic markers (Woodland & Voyer, 2011; Voyer et al., 2016) or (b) alongside mismatching prosodic and gestural markers (González-Fuente, Zabalbeascoa & Prieto, submitted).

Notably, the study by González-Fuente et al. (submitted) demonstrated that listeners place greater emphasis on prosodic and gestural cues compared to contextual ones, and they rely more on gestural information than on prosodic information for identifying irony. These findings imply that the detection of a speaker's ironic intent is significantly dependent on the ability to recognize discrepancies between the valence of the propositional content of the utterance and the valences of contextual, particularly prosodic and gestural, manifestations of that utterance.

The results align with relevance-theory accounts of irony, which assert that comprehending an ironic statement necessitates recognizing both the speaker's propositional and affective attitudes towards the statement (Yus, 2016). Regarding acquisition, research on the development of irony perception indicates that grasping the speaker's intention (determining whether the speaker aims to be agreeable or disagreeable) involves evaluating and synthesizing various cognitive and emotional signals, requiring a complex inference process that improves in accuracy as children grow older (Ackerman, 1983; de Groot et al., 1995; Creusere, 2000; Nakassis & Snedeker 2002; Harris & Pexman 2003; Filippova & Astington, 2008). Despite some variations among studies, evidence suggests that children start to recognize certain elements of ironic intent between the ages of 5 and 11 (e.g., Milosky & Ford, 2009), utilizing contextual and prosodic indicators. Nevertheless, to our knowledge, no prior research has specifically examined the impact of facial expressions in conjunction with prosodic markers on Iranian children's ability to detect irony and sarcasm.

Therefore, the primary objectives of this research are (a) to examine whether prosodic and gestural markers to emotion can aid young children in recognizing a speaker's ironic and sarcastic intent, and if so, (b) to identify the age at which this ability first emerges. Sarcasm, as a prevalent form of verbal irony, is typically characterized as a rhetorical device where the intended meaning is the opposite of the literal interpretation, often expressing a critical viewpoint regarding a specific event or individual (Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995; Cheang & Pell, 2008). Studies on the perception of verbal irony in both adults and children have predominantly concentrated on sarcastic expressions (Ackerman, 1983; Demorest, Mey, Phelps, Gardner & Winner, 1984; Capelli, Nakagawa & Madden, 1990; Nicholson, Whalen & Pexman, 2013).

To ensure that our findings align with the majority of existing literature, the present study will concentrate on this particular form of irony. Most developmental research concerning the acquisition of verbal irony concurs that both contextual markers (Ackerman, 1983; Capelli et al., 1990; Winner & Leekman, 1991) and prosodic markers (Ackerman, 1982, 1983; Capelli et al., 1990; Winner & Leekman, 1991; de Groot et al., 1995; Keenan & Quigley, 1999; Nakassis & Snedeker, 2002; Harris & Pexman, 2003; Climie & Pexman, 2008) serve as significant indicators for children in perceiving sarcastic comments. Nevertheless, there is a lack of consensus regarding the specific age at which children begin to utilize these markers effectively. In terms of contextual markers, whie some research indicates that they do not assist children in recognizing sarcastic remarks until the age of 11 (Capelli et al., 1990), other studies suggest that children as young as 6 can utilize them (Ackerman, 1983; Winner & Leekman, 1991). Likewise, although some studies have demonstrated that children can employ prosody as a marker for identifying sarcastic remarks by age 6 (Keenan & Quigley, 1999), others have found no evidence of this ability until children reach the age of 8 (Ackerman, 1983; Capelli et al., 1990) or even later (Winner, Windmueller, Rosenblatt, Bosco, Best & Gardner, 1987).

As noted by Nakassis and Snedeker (2002) and Laval and Bert-Erboul (2005), variations in experimental results may stem from differences in the materials and methodologies employed, particularly concerning the operationalization of the 'ironic tone of voice'. While the aforementioned studies generally differentiated between 'sincere' and 'ironic' tones, there was a lack of agreement on the precise definition of an ironic tone, with interpretations varying from a 'mocking intonation' (Capelli et al., 1990) to 'stressed intonation patterns' (Ackerman, 1983), or even a simple reference to an 'ironic tone of voice' (Nicholson et al., 2013). It is crucial to emphasize that there is no singular method for verbally conveying irony—there is no definitive 'ironic tone of voice' (Bryant, 2011; González-Fuente et al., 2015)—as the sentiments and emotions conveyed through an ironic statement can range widely from very positive to very negative (Laval & Bert-Erboul, 2005; Wilson, 2013; Yus, 2016). Notably, the only study that has thoroughly investigated it is Nakassis & Snedeker (2002), which examined the influence of intonational cues reflecting positive and negative emotions on the comprehending of irony by adults and 6-year-old children.

The researchers discovered that intonation served as a relational indicator, enhancing children's comprehending of ironic statements when the valence of the intonational markers matched the ironic meaning of the utterance. For instance, a negative intonation heightened the likelihood that a child would perceive the speaker's critical stance, thereby aiding their grasp of the irony being conveyed. Based on these findings, we proposed that the study question 'Does intonation influence irony and sarcasm comprehension?' should be rephrased to 'What types of intonations in which contextual relationships influence the comprehending of irony and sarcasm?'

Overall, we propose that one contributing factor to the varying findings regarding the role of prosody in irony and sarcasm detection across different studies may stem from insufficient control over the emotional valences expressed by prosody compared to those of the literal meaning of the statement. To address this limitation, our current study meticulously controlled for the emotional valence—spanning from positive to negative—expressed not only through prosody but also through facial gestures. It is widely recognized that facial gestures serve as a crucial cue for emotion recognition in children. For instance, Hübscher, Esteve-Gibert, Igualada, and Prieto (2016) conducted a task focused on uncertainty detection with children aged 4 to 6 years, utilizing a set of materials specifically designed to regulate the presence of lexical, intonational, and gestural indicators of uncertainty. Their findings indicated that children exhibited improved performance in recognizing uncertainty when gestural cues were included.

Furthermore, the researchers discovered that younger children exhibited a greater sensitivity to gestural and intonational cues indicating speaker uncertainty compared to lexical markers (such as adverbial expressions like 'perhaps'). This implies that the intonational and gestural aspects of communication may serve as foundational tools in the early stages of pragmatic development. These results align with those of Armstrong et al. (2014), which indicated that facial gestures similarly supported children's ability to recognize a different belief state, specifically incredulity.

While the evidence suggests that facial gestural cues aid in comprehending belief states during child development, to our knowledge, no studies have yet explored their influence on Iranian children's ability to detect irony and sarcasm. Notably, separate researches have demonstrated a significant correlation between children's perception of irony and their capacity to identify emotions in others, reflecting their empathy skills. Nicholson et al. (2013) conducted an experiment on irony perception and processing involving children aged 6 to 7 and 8 to 9 years.

The results revealed that the younger group struggled to recognize the speaker's ironic intent, achieving nearly zero accuracy for ironic statements. In contrast, the older group exhibited a strong correlation between their empathy skills, assessed via the Empathy Quotient for Children, and their success in identifying irony, with a correct response rate of 48% as determined through an object selection task.

As previously mentioned, the primary objective of this study is to examine whether the incongruity between multimodal cues—specifically, the verbal content contrasted with prosodic and gestural signals—can enhance children's ability to recognize a speaker's ironic and sarcastic intent. In accordance with the experimental framework established by Nicholson et al. (2013), we instructed children to participate in an audiovisual task designed to detect irony, which included six congruent prompts (comprising three brief narratives with positive conclusions followed by a video of a speaker displaying a positive response, and three brief narratives with negative conclusions followed by a negative response) as well as six incongruent prompts (six brief narratives with negative conclusions followed by a positive response).

The contrast between a negative context and a positive response in the latter instance was designed to evoke irony ad sarcasm. Importantly, these six 'ironic' remarks were delivered in three different conditions that altered the alignment between the literal meaning of the spoken words and the emotional tone conveyed through accompanying prosodic and gestural signals. In the matching condition, positive remarks were articulated with prosodic and gestural indicators that clearly expressed positive emotions. Conversely, in the strongly mismatching condition, these positive remarks were accompanied by prosodic and gestural signals that suggested a negative emotional tone. Additionally, a third condition introduced a weak mismatch by pairing ironic comments with prosodic and gestural cues where negative emotional content was minimized as much as possible. Our primary hypothesis suggested that a stronger mismatch between prosodic and gestural emotional signals and the literal meaning of verbal comments would lead to higher irony and sarcasm detection scores across all ages. We further anticipated that the enhancement provided by prosodic and gestural cues would be particularly evident in younger children. Essentially, while a negative event contrasted with a positive verbal message alone indicates irony, this contrast might become more pronounced and recognizable to children when accompanied by emotionally negative prosodic and gestural signals. Our secondary research question focused on identifying the

age at which children can detect irony, forecasting it to be earlier than findings from previous studies such as Nicholson et al. (2013). Ultimately, the study's main innovation was our method of controlling and manipulating prosodic and gestural information to examine whether these cues would assist children in recognizing irony in verbal messages.

Material and Methods

Preliminary Study: Discourse Completion Task

To collect the audiovisual materials required for the irony detection task involving children, we initially engaged 8 adult native Persian speakers (mean age: 26.3 years; standard deviation: 4.2) in a Discourse Completion Task (DCT). This method, based on prior research (Blum-Kulka, House & Kasper, 1989; Billmyer & Varghese, 2000; Félix-Brasdefer, 2010), is a semi-spontaneous elicitation task where participants are given a discourse scenario with a situational prompt, followed by a target sentence that they are instructed to produce aloud. The DCT was structured into two blocks. The first block aimed to elicit ironic and sarcastic utterances under three conditions: those accompanied by matching, weakly mismatching, or strongly mismatching prosodic-gestural cues. The second block focused on generating literal (non-ironic) utterances categorized into two conditions: positive (literal compliments) and negative (literal criticisms). These literal utterances were intended to serve as control stimuli for comparison.

a) **Ironic utterances**

Ironic utterances are central to this study, where researchers designed three discourse contexts to elicit sarcastic reactions for an irony detection task (refer to Table 1 for examples). Each context starts with a prompt depicting a negative situation, such as, "A friend of yours is riding a bike. Suddenly, the bicycle falls to the ground and is smashed." This prompt is then followed by a consistent positive comment like "Well done!" The goal is to examine how prosodic and gestural cues may aid in recognizing irony and sarcasm. Thus, speakers were asked to deliver the target sentences under three different conditions. These conditions were categorized based on the alignment between the sentence's inherent positivity ("Well done!") and the emotional tone conveyed, which could be positive, weakly negative, or strongly negative.

Consequently, our study involved three prosodic-gestural conditions: (1) matching, where participants expressed positive comments with an exaggerated, congratulatory tone using both

prosody and facial gestures; (2) weakly mismatching, where they minimized negative emotional elements in their prosody and facial expressions; and (3) strongly mismatching, where participants used prosodic and gestural cues to suggest a negative or critical perspective. Each speaker delivered nine utterances (3 discourse contexts multiplied by 3 prosodic-gestural conditions). For a detailed list of discourse prompts, refer to Appendix B in the author thesis as "Audio-Visual Representation of Linguistic Prosodic and Gestural Cues in Production and Perception processes of Sarcastic Irony and Sarcasm under the Experimental Conditions of Spontaneous and Acted Discourses: An Interpersonal Case Study". The conditions were randomly assigned to the speakers.

b) **Literal utterances**

To generate two sets of literal control stimuli, one representing positive feedback (literal compliments) and the other negative feedback (literal criticisms), participants were presented with the same three prompt scenarios utilized for eliciting ironic and sarcastic utterances. In this instance, each scenario had two possible outcomes: one positive and one negative, each followed by appropriate responses. For instance, the scenario 'A friend of yours is riding a bike. He/She makes the bike move in perfect balance, elicits the response 'Well done!', while the scenario 'A friend of yours is riding a bike'. Suddenly, the bike falls to the ground and is smashed' prompts the response 'What a terrible job!'. Consequently, each speaker generated six control utterances (3 discourse contexts × 2 conditions). The conditions were presented to the speakers in a random sequence.

c) Recording procedure

In relation to the DCT procedure, participants were instructed to read the situational prompt contexts and were subsequently recorded on video while providing the required follow-up comments. The recordings were conducted using a Nikon AF-P DX NIKKOR professional digital video camera in a soundproof room at the counseling department of Azad University Firuzabad Branch. Participants were directed to face the camera and were filmed against a white background, ensuring that their head and upper torso were captured within the video frame. The recordings were digitized at a rate of 25 frames per second, with a resolution of 720 × 576 pixels. The audio was sampled at 44,100 Hz with 16-bit quantization.

Table 1 illustrates an example of a discourse context utilized in the initial segment of the DCT to elicit ironic statements. The original Persian text is presented in italics, accompanied by its English translation beneath. The discourse context features a negative situational prompt in the left column, followed by a positive target comment, 'Well done!' (*آفرین!*) in the right column, categorized into three prosodic-gestural conditions in the middle column: matching (characterized by prosodic and gestural cues that express a positive emotion); weakly mismatching (cues that exhibit restrained emotion); and strongly mismatching (cues that convey a negative emotion).

Table 1. Example of a discourse context utilized in the initial segment of the DCT

situational prompt (a negative event)	Prosodic-gestural conditions	A positive target comment
"A friend of yours is riding a bike. Suddenly, the bicycle falls to the ground and is smashed."	Matching You exaggerate and tell your friend:	"Well done!"
	weakly mismatching You tell your friend with restrained emotions:	
	strongly mismatching you tell your friend in a critical tone:	

d) Analysis of the video recordings:

Analysis of the video recordings involved a comprehensive review of 72 ironic utterances (derived from 3 discourse contexts, 3 ironic conditions, and 8 participants) and 48 literal utterances (taken from 3 discourse contexts, 2 literal conditions, and 8 participants). These utterances were subjected to prosodic scrutiny using Praat (Boersma & Weenink, 2008) in line with the ToBI system (Prieto, 2014). Additionally, accompanying facial gestures were annotated using the ELAN system (Lausberg & Sloetjes, 2009), adhering to the MUMIN Multimodal Coding Scheme guidelines (Allwood et al., 2007: 278), with two extra gestures included: 'Wrinkled nose' and 'Averted gaze' (refer to González-Fuente et al., 2015).

Table 2 illustrates the distribution of intonational and gestural cues produced by participants in the DCT while delivering critical remarks across three specified conditions (the table only includes cues that were present in over 40% of instances). It is evident that in the matching condition, participants predominantly employed a L+H* L% intonational pattern (81%), accompanied by

notable or repeated head nods, raised eyebrows, and smiles (67%); the gestural elements in this condition comprised head nods, a stretched mouth, and raised eyebrows. In the weakly mismatching prosodic-gestural condition, the most frequently used prosody was the marked pattern L*!H% (76%), which is employed in Persian to convey skepticism or disagreement. The prevalent gestural cues in this scenario included head tilts, raised eyebrows, and averted gazes, occurring concurrently with the spoken comment, along with head tilts and a stretched mouth during the gestural coda. Lastly, in the prosodic-gestural strongly mismatching condition, participants exhibited a L* L% intonational pattern in 89% of instances. Disapproving gestures such as head shakes and tilts, furrowed eyebrows, nose wrinkles, and squinted eyes coincided with the speech. During the gestural coda, speakers displayed head shakes and furrowed eyebrows.

The prosodic and gestural markers employed by speakers during the production of the literal control utterances were also examined (refer to Table ۳). The findings aligned with expectations. Specifically, literal compliments were predominantly (89%) articulated using an emphatic L+H* L% intonational pattern, accompanied by gestures indicating approval, typically consisting of nods and raised eyebrows during the pronunciation of the sentence, followed by additional nods and smiles during the gestural conclusion. Conversely, literal criticisms were nearly always (93%) expressed with a L*L% intonational pattern, paired with disapproving gestures such as head shakes and tilts, furrowed brows, wrinkled noses, and squinted eyes, both at the time of utterance and during the gestural conclusion.

Table 2 presents the frequency of various intonational and gestural cues observed in the production of 72 ironic utterances within the DCT, categorized by condition. The rightmost column features video stills that depict the most representative facial gesture associated with each condition.

Table 2. Frequency of various intonational and gestural cues

Condition	Intonation	Gestures	Video still
Matching (positive verbal content matches exaggeratedly enthusiastic gestural prosodic markers) (N = 8)	L+H* L% (81%)	**During speech** Head nod (89%) Raised eyebrows (87%) Head tilt (54%) Smile (63%) **Codas** Head nod (75%) Raised eyebrows (46%) Stretched mouth (23%)	
Weakly mismatching (positive verbal content is accompanied by gestural prosodic markers with emotion restrained) (N = 8)	L*!H% (76%)	**During speech** Raised eyebrows (75%) Head tilt (46%) Averted gaze (29%) **Codas** Head tilt (42%) Stretched mouth (30%) Raised eyebrows (43%)	
Strongly mismatching (positive verbal content is contradicted by gestural prosodic markers signaling criticism) (N = 8)	L* L% (89%)	**During speech** Head shake (88%) Furrowed eyebrows (61%) Head tilt (43%), Wrinkled nose (22%), Squinted eyes (32%) **Codas** Head shake (57%) Furrowed eyebrows (23%)	

All in all, the findings from the DCT support earlier research on the audiovisual cues associated with verbal irony. Verbal irony is typically indicated by particular pitch patterns, such as distinct contrasting tonal-nuclear configurations (refer to González-Fuente et al., 2015), along with gestures that accompany speech both during and after the expression of irony and sarcasm, including facial expressions characterized by specific eye and eyebrow movements, as well as

laughter and smiles (see Attardo et al., 2003, 2011; Caucci & Kreuz, 2012; González-Fuente et al., 2015). The subsequent subsection will outline how the recorded materials were utilized as stimuli in an experiment aimed at assessing children's capacity to recognize irony and sarcasm.

Table 3 present a summary and distribution of prosodic and gestural cues detected in the 48 literal utterances gathered through the DCT under two literal control conditions. The rightmost column includes video stills that demonstrate the facial gestures most representative of each condition.

Table 3. A summary and distribution of prosodic and gestural cues detected in the 48 literal utterances

Literal control condition	Prosody	Gestures	Video still
literal compliment (N = 24)	L+H*L% (89%)	**During speech** Head nod (91%) Raised eyebrows (87%) **Gestural codas** Head nod (78%) Smile (64%)	
literal criticism (N = 24)	L* L% (93%)	**During speech** Head shake (78%) Furrowed eyebrows (77%) Squinted eyes (38%) Wrinkled nose (13%) Head tilt (29%) **Gestural codas** Furrowed eyebrows (31%) Head shake (63%)	

3.2. Experimental materials

The initial DCT produced video recordings, which provided the necessary material to develop stimuli for the irony detection experiment with children. This experiment, partially inspired by the design in Nicholson et al., 2013, involved presenting young children with 12 PowerPoint presentations. Each presentation consisted of a brief narrative followed by an embedded video

showing an adult's reaction to the narrative, with these videos originating from the DCT mentioned earlier.

In six of the trials, the adult's reaction aligned with the narrative's outcome, where a positive narrative led to a complimentary or congratulatory reaction, while a negative narrative resulted in a critical or hostile response. Conversely, in the other six trials, the adult's reaction was incongruent, as a negative outcome was met with a compliment, thereby creating an ironic effect. These six ironic responses were presented audio-visually across the three prosodic-gestural conditions previously described, with two responses in the matching condition (where the positive verbal content corresponded with enthusiastic prosodic-gestural cues), two responses in the weakly mismatching condition (where the prosodic-gestural cues displayed restrained emotion), and the final two responses in the strongly mismatching condition (where the positive verbal content contradicted the hostility indicated by the prosodic-gestural cues).

Each discourse context illustrated scenarios that were likely recognizable to children, featuring two distinct characters: a cartoon figure and a real human, which varied across different contexts. Each context was conveyed through a series of four slides. The initial slide introduced the characters, while the second and third slides provided a brief narrative that concluded with either a positive or negative result. The final slide included an embedded video of the human character responding to the event's outcome. In the following, Figure 1 depicted a sample slide from a PowerPoint presentation utilized in the irony detection task (see Appendix C in author thesis for the complete collection of 12 sequences).

Slide 1: introduction	**Slides 2 and 3: situational prompt**	**Slide 4: utterance**

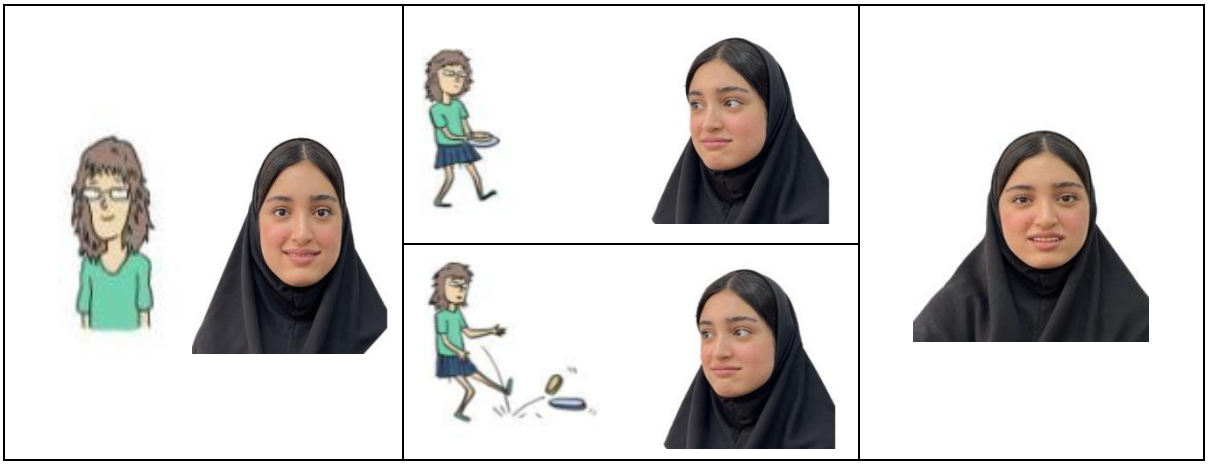

Figure1. Sample slides from a PowerPoint presentation utilized in the irony detection task

The first slide introduces the two characters, while slides two and three depict an event with a narrative conclusion. The fourth slide features an embedded video file displaying the human character's reaction to the event.

The researcher selected 12 embedded videos showing human reactions to the event, choosing from 120 available videos (48 literal and 72 ironic reactions) as part of the DCT task outlined in section 3.1. Particularly, the researcher selected 3 literal compliments, 3 literal criticisms, and 6 ironic comments, with 2 videos for each prosodic-gestural condition: matching, weakly mismatching, and strongly mismatching. The criteria for selection were based on the prosodic and gestural markers explained in section 3.1.d, ensuring that each video showcased cues most representative of its condition. Additionally, each selected video was performed by a different speaker.

Following the creation of the 10 stimulus presentations, a validation process was conducted utilizing the Digisurvey, that is an online survey platform. A group of 30 adults fluent in Persian participated by viewing the 12 presentations online. After each presentation, they were prompted to determine whether the human character's response was interpreted as literal or ironic. In cases of irony, they were further asked to assess the level of criticism perceived on a 7-point Likert scale. The findings revealed a remarkable consensus (98.9%) among the 30 participants in differentiating between the intended literal and ironic reactions.

A statistical analysis using a GLMM (Generalized Linear Mixed Model) was performed to determine significant differences in the degree of criticism ratings across three prosodic-gestural conditions. This was executed with SPSS Statistics 23 (IBM Corp., 2015). The dependent variable

was PERCEIVED CRITICISM, which was assessed on a 7-point scale ranging from '0' (not critical) to '7' (highly critical).

In this study, the constant variable was PROSODIC-GESTURAL CUES, categorized into three levels: matching, weakly mismatching, and strongly mismatching. SUBJECT and ITEM were treated as random factors. The findings indicated that participants were able to significantly differentiate among the three prosodic-gestural conditions ($F(2,346) = 24.732$, $p < .01$). The matching condition yielded a mean criticism rating of 2.7 (SD 0.5), the weakly mismatching condition resulted in a mean criticism rating of 5.3 (SD 1.2), and the strongly mismatching condition produced a mean criticism rating of 7.2 (SD 0.4).

3.3. Participants

A total of 44 children participated in the experiment, divided into three age groups: a group of 15 five-year-olds (mean age 5 years 6 months, standard deviation = 5.25), a group of 14 eight-year-olds (mean age 8 years 3 months, standard deviation = 3.66), and a group of 15 eleven-year-olds (mean age 11 years 5 months, standard deviation = 4.31). All participants came from middle-class families and were enrolled as preschoolers or students at three public schools in the Firouzabad region. To confirm that the children were primarily exposed to Persian rather than Turkish (in this region, some people are bilingual), a language exposure questionnaire (adapted from Bosch & Sebastián-Gallés, 2001) was given to the parents, revealing a mean exposure percentage of 87% for Persian (standard deviation = 12.2). Parents were also briefed on the purpose of the experiment and provided consent for their child's participation and for the experimental procedures to be recorded on video. Each child was tested individually at their respective schools and did not receive any compensation for their involvement.

3.4. Procedure

The experiment was conducted in a quiet room at each of the two participating schools. The child was positioned in front of a computer screen, accompanied by researcher who spoke Persian language and seated beside him/her. A second researcher (the author of this paper) was situated behind the child to document the child's actions manually, while a video camera was set up to capture the entire procedure from the child's perspective.

On the computer, four training PowerPoint presentations were displayed, followed by 12 stimulus presentations. The purpose of the training presentations was to instruct the child in evaluating

whether the human character's reaction to a particular story was "nice" or "mean". These presentations adhered to the same format as the stimulus videos as outlined in section 2.2, illustrating a narrative that concluded with either a positive or negative result, followed by a video showing an individual's reaction.

Among the four stories presented, two had positive outcomes while the other two ended negatively. Nevertheless, none of the responses during the training sessions exhibited irony. Following the observation of the individual's reaction to each narrative, the children were instructed to manually indicate their judgment of 'nice' or 'mean' by placing one of two toys into a designated 'answer bin' situated between them and the computer screen (refer to Figure 2 below). It was clarified that one toy represented the “nice rabbit” and the other the “mean octopus”. If a child perceived the human character's reaction as 'nice like the rabbit, they were to place the rabbit in the bin. Conversely, if they deemed the reaction as 'mean like the octopus, they were to place the octopus in the bin. After the completion of the four training presentations and the commencement of the main experiment, the child received no additional instructions regarding the placement of the toy. The positioning of the rabbit and octopus to the right or left of the child was systematically varied among participants.

Figure 2 illustrates the experimental setup for the irony detection task, featuring a plastic "answer bin" positioned between the child participant and the computer screen. In this instance, the “nice rabbit” is located to the child's right, while the “mean octopus” is to the left. Following the training sessions, the children engaged in a series of 12 experimental trials. These consisted of 6 stories with congruent, literal reactions—that is, 3 positive reactions paired with positive outcomes and three negative reactions paired with negative outcomes. The remaining 6 stories presented incongruent reactions: two featured aligned verbal and prosodic-gestural messages, two included mildly mismatched prosodic-gestural cues, and two showcased conflicting verbal and prosodic-gestural signals. The trials were presented to the children in varied sequences; 12 distinct PowerPoint presentations were generated to arrange the trials, and children were randomly assigned to one of these orders. In total, each participant underwent the complete procedure in roughly 15 minutes.

Figure 2. The experimental setup for the irony detection task

Results

Initially, 44 children were involved in the irony detection task; however, the results from 2 participants were excluded from the final analysis due to inappropriate responses during the training trials, indicating a lack of comprehending of the procedure. Consequently, data was collected from 42 children, each completing 12 experimental trials, resulting in a total of 504 responses (42 children × 12 responses). The data comprised the children's responses recorded by the second researcher (the author of this paper) during the experiment and verified through subsequent review of the video recordings. These responses were categorized as 'correct' or 'incorrect'. A 'correct' score was assigned in three specific conditions:

After a positive response to a positive event, the child selected the "nice rabbit"; after a negative reaction to a negative event, the child selected the "mean octopus"; or after a positive response to a negative event under any of the three prosodic-gestural conditions, the child selected the "mean octopus". Every other choice was deemed "incorrect".

The findings demonstrated that for both positive and negative literal control conditions, participants' response accuracy was at its maximum. The level of precision in the non-ironic positive condition (i.e. e. compliments after positive results) was 100%, and the level of precision in the negative literal condition (i.e. e. criticisms after negative results) was 94%. As a result, additional analyses did not include these two conditions. Using SPSS Statistics 23 software (IBM Corp, 2015), responses to the ironic reactions in the three prosodic-gestural conditions were

analyzed using a GLMM test. The dependent variable, RESPONSE, was quantified by calculating the mean proportion of correct to incorrect responses. The constant factors included PROSODIC-GESTURAL CUES (with three levels: matching, weakly mismatching, and strongly mismatching), AGE (with three levels: 5 years, 8 years, and 11 years), along with their interactions. SUBJECT and ITEM were designated as random factors.

The results from the Generalized Linear Mixed Model (GLMM) indicated a significant main effect of AGE ($F(2,1065) = 55.78$, $p < .001$), demonstrating that 11-year-olds outperformed 8-year-olds, who in turn outperformed 5-year-olds. This suggests a distinct developmental pattern in children's ability to detect irony as they age. Additionally, a significant main effect of PROSODIC-GESTURAL CUES ($F(2, 1065) = 501.86$, $p < .001$) was observed, with correct response scores differing significantly across all three levels of the PROSODIC-GESTURAL CUES condition. Consequently, the 'strongly mismatching' prosodic-gestural condition elicited significantly more correct responses than the 'weakly mismatching' condition, which also yielded significantly more correct responses than the 'matching' condition.

Importantly, the results from the Generalized Linear Mixed Model (GLMM) revealed a significant interaction effect between the variables AGE and PROSODIC-GESTURAL CUES ($F(4, 1065) = 61.38$, $p < .001$). This indicates that the influence of the prosodic-gestural condition on responses varied across different age groups. Figure 3 illustrates the average proportion of correct responses categorized by prosodic-gestural condition (i.e., matching, weakly mismatching, and strongly mismatching) and age (5-year-olds, 8-year-olds, and 11-year-olds). Subsequent pairwise comparisons indicated that the age groups exhibited significant differences in their responses based on the prosodic and gestural cues associated with an ironic reaction comment.

Individuals across all age groups demonstrated a significantly enhanced ability to detect irony and sarcasm in the strongly mismatching condition compared to the other two conditions, and also exhibited a notably better performance in the weakly mismatching condition relative to the matching condition. This suggests that children were considerably more inclined to detect irony and sarcasm when the emotional cues provided by prosody and gestures starkly contrasted with the literal meaning of the statement. For instance, when a negative outcome was met with the phrase 'Well done', yet the accompanying prosodic and facial expressions conveyed hostility or

criticism. Furthermore, for the 5-year-old participants, the strongly mismatching condition was the sole scenario among the three that was ever interpreted as ironic, occurring in 41% of instances. Generally speaking, these findings appear to support our hypotheses that prosodic and gestural cues can aid children in comprehending a speaker's ironic intent during the early stages of their development, particularly when these cues provide pragmatic information that significantly contradicts the semantic meaning of the statement.

Figure 3 illustrates the average percentage of correct responses in irony detection categorized by age and the prosodic-gestural condition. The blue columns represent instances where prosodic-gestural cues aligned with the verbal content, the pink columns indicate cases where these cues were only weakly mismatched with the verbal content, and the green columns depict scenarios where the prosodic-gestural cues were in strong contrast to the verbal content.

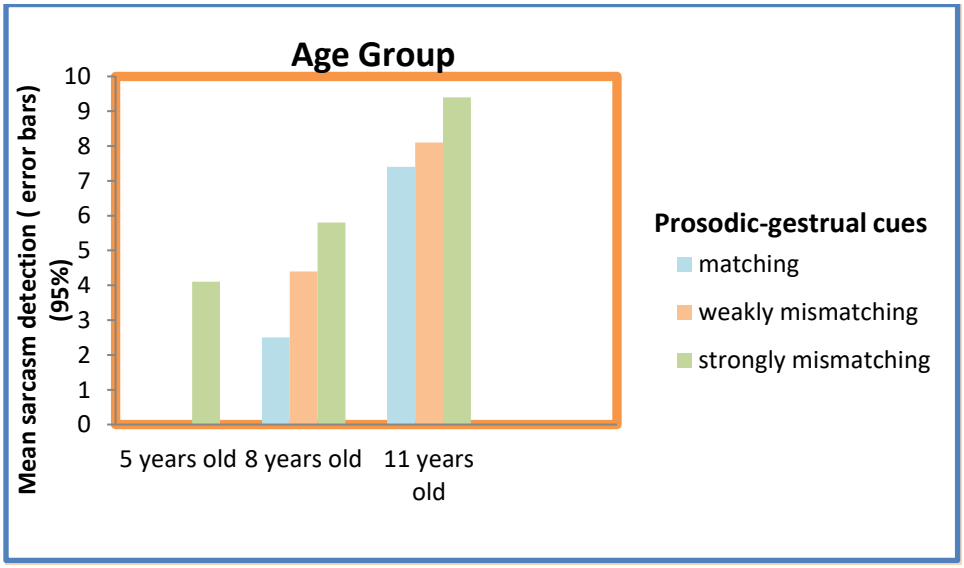

Figure 3. Average percentage of correct responses in irony detection categorized by age and the prosodic-gestural condition

Discussion

This paper investigated the function of prosodic and gestural markers of emotion in enhancing Iranian children's capability to detect a speaker's ironic and sarcastic intent. Through an audiovisual task designed for irony and sarcasm detection, we demonstrated that strongly mismatched prosodic and gestural signals resulted in higher rates of irony detection compared to

both weakly mismatched and matching cues, observed in children aged 5, 8, and 11. These findings aligns with previous studies indicating that prosodic signals enhanced the comprehending of sarcasm in 6-year-olds, particularly when the emotional tone of the prosodic cue conflicted with the literal meaning of the statement, such as when a negative intonation contrasted with a positive verbal message (Nakassis & Snedeker, 2002).

Building upon the research conducted by Nakassis and Snedeker (2002), our findings indicate that a greater degree of incongruity between the valences of prosodic-gestural cues and verbal comments correlates with increased irony detection scores among children. Specifically, strongly mismatched prosodic and gestural cues resulted in significantly higher rates of irony and sarcasm detection compared to weakly mismatched cues, which in turn elicited a greater number of correct responses than matching prosodic-gestural cues. These results further enhance the insights of prior studies that highlight the significance of contrast effects in the perception of sarcasm (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003; Woodland & Voyer, 2011; Voyer et al., 2016; González-Fuente et al., submitted).

Our findings align with previous studies indicating that the most effective combination of incongruent prosodic and gestural signals (such as employing an 'ironic' tone of voice alongside negative emotional facial expressions) in conjunction with conflicting discourse contexts (for instance, a negative scenario) resulted in the highest rates of irony detection among adults (Woodland & Voyer, 2011; Voyer et al. 2016, González-Fuente et al., submitted).

Conversely, our results reveal a distinct developmental pattern in children's ability to detect irony and sarcasm, as evidenced by the increase in their average irony perception scores with age across all prosodic-gestural contexts. In this context, our findings align with earlier research (e.g., de Groot et al., 1995; Creusere, 2000; Filippova & Astington, 2008; Harris & Pexman, 2003; Nakassis & Snedeker, 2002). A primary inquiry of this study was whether access to visual cues would enhance younger children's ability to recognize irony. As previously mentioned, our experimental framework was modeled after that of Nicholson et al. (2013), which uniquely incorporated visual elements into the experimental design, setting it apart from prior studies on children's irony perception.

In their research, the visual markers were represented by puppets. In the current investigation, we advanced the methodology by utilizing video recordings of actual individuals who could convey

verbal messages while exhibiting incongruent prosodic and facial gestures. Nicholson et al. (2013) examined children aged 6 to 7 and 8 to 9 years, discovering that the younger age group demonstrated almost no accuracy in detecting irony. Conversely, our findings for 5-year-old participants indicated that, although these children similarly struggled to detect irony when the prosodic and gestural signals accompanying a positively framed verbal message were either positive or mildly critical/negative, they successfully perceive irony 41% of the time when positive verbal content was paired with distinctly hostile prosodic and gestural signals.

The findings indicate that the ability to perceive irony may emerge at very early developmental stages, contingent upon children being exposed to significantly contrasting prosodic and gestural emotional cues. This occurs when these cues starkly differ from the literal meaning of a statement. Consequently, these results align with the research conducted by Armstrong et al. (2014) and Hübscher et al. (2016), which demonstrated that visual cues enhanced the ability of children aged 4 to 6 to recognize pragmatic meanings such as disbelief or uncertainty. All in all, our results support the increasing agreement that pragmatic gestures serve as essential tools in the development of language, particularly in the realm of pragmatics (e.g., McNeill, 1998; McNeill, Cassell & McCullough, 1994; Kelly, 2001; Butcher & Goldin-Meadow, 2000).

In conclusion, our findings indicate that children exhibit a heightened sensitivity to emotional expressions communicated through prosodic and gestural signals, enabling them to utilize these cues to assess a speaker's intentions. This observation is in line with prior research demonstrating a significant correlation between children's appreciation of irony and their empathetic abilities, particularly their capacity to recognize emotions in others (Nicholson et al., 2013). As previously mentioned, and as recent assertions from pragmatic cognitive theories like the Relevance Theory suggest, recognizing the speaker's 'affective attitude' is essential for comprehending an ironic statement (Wilson, 2013; Yus, 2016).

Considering the perspective of emotional valence, we assert that research focused on the development of irony detection must include a detailed examination of the emotionally charged visual and prosodic elements that accompany ironic statements. As noted by Bryant (2012), grasping verbal irony is a multimodal process where all factors involved in conveying an ironic statement influence its accurate interpretation. Therefore, the investigation of irony and sarcasm detection should not only explore the differences between the literal meanings of words and their

pragmatic contexts but also integrate the analysis of the interactions among contextual, propositional, prosodic, and gestural signals in the perception of irony and sarcasm.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committee of Islamic Azad University.

Author contributions

All authors contributed to the study conception and design, material preparation, data collection and analysis. All authors contributed to the article and approved the submitted version.

Funding

The authors did (not) receive support from any organization for the submitted work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Ackerman, B. (1982). Contextual integration and utterance interpretation: the ability of children and adults to interpret sarcastic utterances. *Child Development*, *53*, 1075-1083.

Ackerman, B. (1983). Form and Function in Children's Understanding of Ironic Utterances. *Journal of Experimental Child Psychology*, *35*, 487-508.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2007). The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management and Sequencing Phenomena. *Language Resources and Evaluation*, *41*(3/4), 273-287.

Blum-Kulka, S., House, J., & Kasper, G. (1989). Investigating cross-cultural pragmatics: an introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 1-34). Norwood, NJ: Ablex.

Boersma, P., & Weenink, D. (2008). PRAAT: doing phonetics by computer. Version 5.4.08. Computer program. Retrieved 2 April 2015 from http://www.praat.org/.

Bosch, L., & Sebastián-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29-49.

Bryant, G.A. (2011). Verbal irony in the wild. *Pragmatics and Cognition*, *19*(2), 291-309.

Bryant, G.A. (2012). Is verbal irony special? *Language and Linguistics Compass*, *6*(11), 673-685.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to twoword speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235-257). New York: Cambridge University Press.

Capelli, C.A., Nakagawa, N., & Madden, C.M. (1990). How children understand sarcasm: the role of context and intonation. *Child Development*, *61*, 1824-1841.

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*, 366-381.

Climie, E. A., & Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *Journal of Cognitive Development*, *9*, 257-285.

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics*, *34*, 111-142.

Colston, H. L., & O'Brien, J. (2000). Contrast of kind vs. contrast of magnitude: The pragmatic accomplishments of irony and hyperbole. *Discourse Processes*, *30*, 179-199.

Creusere, M. A. (2000). A developmental tests of theoretical perspectives on the understanding of verbal irony: children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbolic Activity*, *15*, 29-45.

de Groot, A., Kaplan, J., Rosenblatt, E., Dews, S., & Winner, E. (1995). Understanding versus discriminating nonliteral utterances: Evidence for a dissociation. *Metaphor and Symbolic Activity*, *10*, 255-273.

Demorest, A., Mey, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: understanding deliberately false remarks. *Child Development*, *55*, 1527-1534.

Filippova, E., & Astington, J.W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, *79*, 126-138.

Gerrig, R. J., & Goldvarg, Y. (2000). Additive effects in the perception of sarcasm: Situational disparity and echoic mention. *Metaphor and Symbol*, *15*, 197-208.

González-Fuente, S., Escandell-Vidal, V., & Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics*, *90*, 26-47.

González-Fuente, S., Prieto, P., & Noveck, I. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in 268 French. In *Proceedings of the Speech Prosody 2016* (pp. 902-906). Boston, MA (USA).

González-Fuente, S., Zabalbeascoa, P., & Prieto, P. (submitted). Communicating irony: when gesture cues are more powerful than prosodic and contextual cues. *Applied Psycholinguistics*.

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, *36*, 147-165.

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2016). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language, 37*(1), 24-41.

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Ivanko, S. L., & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Processes*, *35*, 241-279.

Keenan, T., & Quigley, K. (1999). Do young children use echoic information in their comprehension of sarcastic speech? A test of echoic mention theory. *British Journal of Developmental Psychology*, *17*, 83-96.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology*, *118*, 374-386.

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*, 3-21.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior research methods, Instruments & Computers, 41(3),* 841-849. Computer program. Retrieved 14 February 2015 from http://www.lat-mpi.eu/tools/elan/.

Laval, V., & Bert-Eboul, A. (2005). French-speaking children's understanding of sarcasm: The role of intonation and context. *Journal of Speech, Language, & Hearing Research*, *48*, 610-620.

McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction, 27*(3), 223-237.

Milosky, L., & Ford, J. (2009). *The role of prosody in children's inferences of ironic intent*. *Discourse Processes, 23*(1), 47-61.

Nakassis, C., & Snedeker, J. (2002). Beyond sarcasm: Intonation and Context as Relational Cues in Children's Recognition of Irony. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the Twenty-sixth Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Frontiers in Psychology*, *4*, 691.

Prieto, P. (2014). The Intonational Phonology of Catalan. In: Sun-Ah Jun (Ed.), *Prosodic Typology 2* (pp. 43-80). Oxford, UK: Oxford University Press.

Van Lancker, D. (2008). The Relation of Human Language to Human Emotion. In B. Stemmer & H.A. Whitaker (Eds.), *Handbook of the Neuroscience of Language* (pp. 199-208). New York: Academic Press.

Voyer, D., & Vu, J. P. (2016). Using sarcasm to compliment: Context, intonation, and the perception of statements with a negative literal meaning. *Journal of Psycholinguistic Research*, *45*, 615-634.

Voyer, D., Thibodeau, S.H., & Delong, B.J. (2016). Context, Contrast, and Tone of Voice in Auditory Sarcasm Perception. *Journal of Psycholinguistic Research, 45*, 29-53.

Wilson, D. (2013). Irony comprehension: A developmental perspective. *Journal of Pragmatics*, *59*, 40-56.

Winner, E., & Leekman, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, *9*, 257-270.

Woodland, J., & Voyer, D. (2011). Context and Intonation in the Perception of Sarcasm. *Metaphor and Symbol*, *26*, 227-239.

Yus, F. (2016). Propositional attitude, affective attitude and irony comprehension. *Pragmatics & Cognition*, *23*(1), 92-116.