Research in Generative Grammar
RGG 2025.16
Rivista di Grammatica Generativa

Every smile you fake, I'll be watching you. Visual cues, affective stance, and irony

Beatrice Giustolisi & Francesca Panzeri
University of Milano-Bicocca

Abstract. Irony mutes the aggressiveness of criticisms, and tinges compliments with unfriendliness. In this work we first aimed to investigate whether interlocutors are sensitive to the ironists' affective stance conveyed through visual cues. We asked 103 raters to evaluate the friendliness of a series of sincere/ironic praises/blames presented as videos from which audio has been removed. We then addressed the question of whether interlocutors rely on the perceived 'jocularity' of speakers to attribute them an ironic intent. We re-analyzed previously published data from a study on the recognition of irony based on visual cues, verifying whether the newly collected friendliness ratings might predict accuracy. Results confirmed the hypothesis that when presented with pairs of sincere/ironic videos, participants tend to attribute irony to the friendlier video of the pair.

Keywords: Irony recognition, Irony markers, Visual cues, Friendliness ratings

Cite as: Giustolisi, Beatrice & Francesca Panzeri. 2025. *Every smile you fake, I'll be watching you.* Visual cues, affective stance, and irony. *RGG*, 47(16). pp. 1–13. Lingbuzz Press.

1 Introduction

To avoid misunderstandings and communicative failures, ironic speakers may exploit metacommunicative cues to alert interlocutors that their remark should not be taken at face value but need to be reinterpreted (typically, as conveying the opposite meaning). The most studied irony marker is the so-called ironic tone of voice (Bryant & Fox Tree 2002: see); whereas scarce attention has been paid to the ironic face (Burgers & Van Mulken 2017: for a summary see). Ironic statements have been described as accompanied by a series of visual cues, neither obligatory nor specific, and often opposites in their direction, for example: raised or lowered eyebrows, wide-open eyes or squinting (Attardo et al. 2003), winking (Muecke 1978) or blank face (Attardo et al. 2003). Given the importance of multimodality for linguistic communication (e.g. Holler & Levinson 2019), the goal of our study is to focus on the contribution of visual cues for the recognition of irony.

Previous studies have shown that visual cues alone permit to distinguish ironic from sincere comments (Aguert 2022, Li et al. 2022); however, they are less reliable than contextual cues (Deliens et al. 2018) and their role seems to be indirect (Giustolisi & Panzeri 2021). Specifically, in Giustolisi & Panzeri (2021) we investigated the relative weight of visual and acoustic cues for the correct detection of irony. We used a Discourse Completion Task (Félix-Brasdefer 2010) to elicit a semi-spontaneous production of sincere and ironic positive (such as “It was a beautiful hike!”, N=5) and negative (such as “You are surely bad at drawing!”, N = 5) remarks by four different speakers, who were video recorded while they were uttering them. We thus obtained a total of eighty remarks: 20 sincere compliments, 20 ironic criticisms, 20 sincere criticisms and 20 ironic compliments (see Table 1). The eighty videos were then edited to obtain a muted version (the audio was removed) and a version with only the audio. In Study 1 and 3, we presented pairs of the same remark pronounced ironically and sincerely, asking participants to detect the ironic one. We found that visual cues were even more reliable than acoustic ones, since participants had a very good accuracy when only facial expressions were available. Nevertheless, it is important to note that in Study 1 and 3 the remark was written on the screen and so participants knew what was being said. However, when participants were not informed of the content of the sentence that was being pronounced (Study 4), the rate of accuracy dropped, for ironic compliments (from 89% to 79%), and even more for ironic criticisms (from

84% to 64%, with the majority of participants responding at chance level). These data cast doubt on the idea that there exist visual cues that specifically convey the speaker's ironic intent and suggest that in the first three studies participants might have recognized irony only indirectly, by comparing the polarity of the remark with the polarity of the speaker's attitude (see also Mantovan, Giustolisi & Panzeri 2019).

To account for the fact that in Study 4 participants wrongly attributed an ironic intent to speakers who were making a sincere compliment, we hypothesized that participants could mistake the speaker's positive stance in (sincerely) praising someone with the jocularity typically associated with irony. According to the most credited accounts of irony, ironic speakers intend to manifest their negative, contemptuous attitude towards a thought that is being echoed (Wilson & Sperber 2012), or the person who would be so foolish to entertain that thought (Clark & Gerrig 1984). In other words, irony is typically associated with a critical stance. At the same time, though, ironic speakers are considered to be funny (Dews, Kaplan & Winner 1995), and many instances of real life ironic remarks constitute what Kotthoff (2003) labels as "friendly-playful irony": describing a nearly unfurnished room as "orderly" does have a negative connotation ("too little furniture"), but it is perceived as teasing and as a good-humoured remark. We might then conjecture that participants associate irony with positive attitudes. This hypothesis is in line with the results of Deliens et al. (2018)'s first experiment, in which participants rated sincere positive statements as more ironic than their sincere negative versions, suggesting that the positive attitude conveyed by praises might be confused with irony. This line of reasoning could also explain why participants detected ironic compliments with higher accuracy compared to ironic criticisms: In fact, if irony is identified with the expression of a positive stance, it would be easier to recognize an ironic compliment paired with its sincere counterpart (a criticism) compared to an ironic criticism paired with its sincere counterpart (a compliment).

The main goal of the present work is to analyse the visual cues of irony with an investigation that builds on previous work on the acoustic cues of irony. Having noted that there do not seem to be clear acoustic correlates that univocally characterize ironic criticisms and ironic compliments (Bryant & Fox Tree 2005) Mauchand, Vergis & Pell (2020) investigated whether interlocutors would consider the 'friendliness' of speakers as a proxy for the correct detection of their communicative intent. They found general support for the Tinge Hypothesis (Dews, Kaplan & Winner 1995, Pexman & Olineck 2002) and for the Asymmetry of Affect Hypothesis (Matthews, Hancock & Dunham 2006): Irony mutes the aggressiveness of criticisms, but also attenuates the friendliness of compliments. Moreover, interlocutors seem to take into account the af-

fective contribution of prosody spontaneously in the case of ironic criticisms (perceived as less friendly than their sincere counterparts, i.e., sincere compliments), whereas ironic compliments were rated as unfriendly as their sincere counterparts (i.e., sincere criticisms) unless participants were explicitly asked to focus on the way they are pronounced. Interestingly, McKinnon & Prieto (2014) asked participants to rate the impoliteness of speakers pronouncing insulting comments sincerely and with a mocking attitude and found that participants rated mocking speakers as less impolite (and thus friendlier) when they based their judgments on visual cues compared to acoustical cues alone.

Inspired by these studies, our goal is to explore further the role of visual cues in transmitting the friendliness of sincere and ironic speakers, and the relationship between the perceived positive affective stance and the ironic communicative intent. To this aim, we presented participants with videos of speakers pronouncing literally positive and literally negative remarks sincerely (sincere compliments and sincere criticisms) and ironically (ironic criticisms and ironic compliments), and we asked them to rate their friendliness on a 5-point Likert scale. Crucially, the audio track was removed from the video, therefore responses were based on the observation of the speakers face and not on the speakers voice. We were expecting to detect a modulation of friendliness ratings across content (positive/negative) and attitude (positive/negative). Moreover, it was our aim to use these friendliness ratings to reanalyse Giustolisi & Panzeri (2021) Study 4 results and directly test their hypothesis.

2 Friendliness ratings

2.1 Methods

2.1.1 Participants

A total of 103 raters were recruited through flyers on social media and personal contacts. One rater did not complete the study, and it was removed from the final database of 102 raters (76 females, 26 males, mean age = 41 yrs, SD = 16). All raters had Italian as their first language.

2.1.2 Materials

We used the same video materials as Giustolisi & Panzeri (2021) in the "without audio" modality, i.e., we used 80 muted videos of a person uttering a comment that could be either sincere or ironic. Crucially, the 80 experimental items were composed of 10 pairs of comments, in which

Content	Attitude	Type of comment	N
Positive	Positive	Sincere compliment	20
Es. "It was a beautiful hike!"	Negative	Ironic criticism	20
Negative	Positive	Ironic compliment	20
Es. "You surely are bad at drawing!"	Negative	Sincere criticism	20

Table 1: Materials summary and examples of sentences with positive/negative content.

the same remark was produced either sincerely or ironically by four different Italian speakers (10 pairs * 4 speakers = 40 pairs of comments = 80 items). Fifty percent of the remarks had positive content, 50% negative content, therefore 25% of them were sincere compliments (remark with positive content, uttered sincerely), 25% sincere criticisms (remark with negative content, uttered sincerely), 25% ironic criticisms (remark with positive content, uttered ironically) and 25% ironic compliments (remark with negative content, uttered ironically) (see Table 1).

2.1.3 Procedure

The raters completed the study via Qualtrics. Experimental items were presented randomly, and for each muted video participants were asked to rate the friendliness of the speaker on a 5-point Likert scale. Participants were informed neither about the polarity of the comments, nor about the attitude of the speaker, and they did not know that the material was composed of ironic and sincere remarks.

2.1.4 Hypothesis

In the absence of linguistic cues, we hypothesized that raters base their friendliness ratings on the actors' facial expressions, which should be influenced both by content and attitude. If this is the case, we expect friendliness ratings to be higher for comments with positive content than for content with negative content, and to be higher for comments with a positive attitude compared with comments with a negative attitude.

2.2 Results and analysis

Friendliness ratings across types of comments are summarized in Table 2. Since participants might vary in the interpretation of the 1 to 5 scale, before proceeding with statistical analysis the ratings were transformed into z-scores based on each participant's mean rating and standard deviation (Figure 1).

Z-scores were analysed with linear mixed models with the lme4 package (Bates et al. 2015) in R, version 4.1.2 (R Core Team 2021). Fixed

Content	Attitude	Type of comment	Median	Mean	SD
Positive	Positive	Sincere compliment	3	3.12	1.07
	Negative	Ironic criticism	3	2.79	1.08
Negative	Positive	Ironic compliment	3	2.93	1.12
	Negative	Sincere criticism	2	2.19	0.91

Table 2: General median, mean, and SDs for the ratings in ironic compliments, ironic criticisms, literal compliments, and literal criticisms.

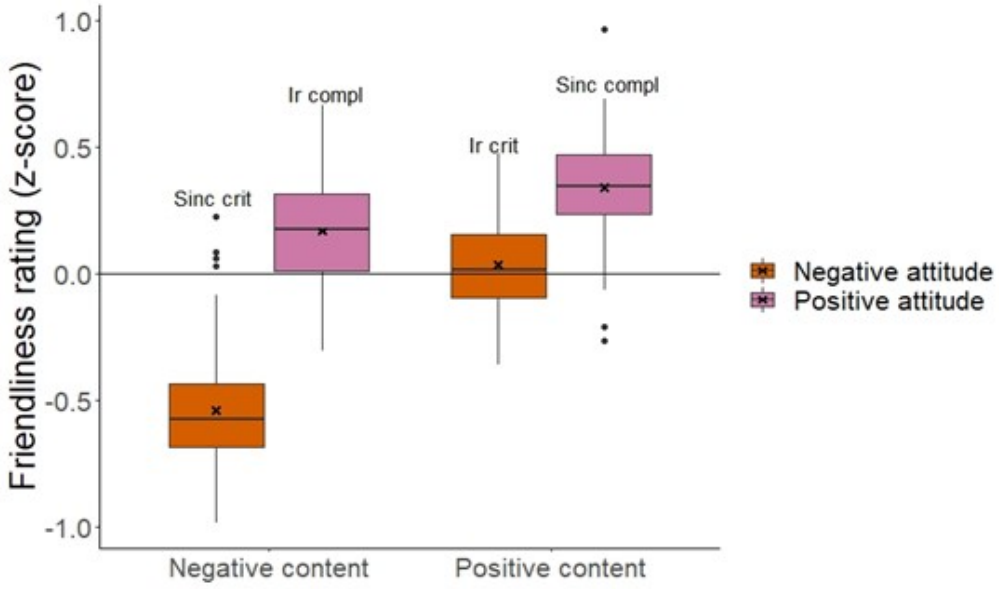

Figure 1: Distribution of participants' mean z-scores across content (x-axis) and attitude conditions (colors). The straight line indicates the median, whereas the cross the mean.

factors were content (negative coded as 0 and positive as 1), attitude (negative coded as 0 and positive coded as 1), and their interaction, whereas random factors were by pair and by actor random intercepts and by pair and by actor random slopes for the effect of attitude. Random intercepts for participants were not included to avoid singular fit. A likelihood ratio test comparing the model with the interaction and a model without the interaction revealed that the interaction was not significant ($\chi^2(1) = 2.92$, $p = .087$) and it was therefore removed from the model.

The analysis of z-scores showed a significant effect of content ($\beta = 0.31$, $SE = 0.12$, $t = 2.71$, $p = .027$), with higher ratings for comments with positive content compared to comments with negative content, and a significant effect of attitude ($\beta = 0.51$, $SE = 0.18$, $t = 2.76$, $p = .022$), with higher ratings for comments with positive attitude compared to comments with negative attitude.

2.3 Interim discussion

In a study in which only visual cues were available, we found that participants provided higher friendliness ratings for compliments than for criticisms, and for comments with a positive content than for comments with a negative content. Even if a direct comparison is not strictly possible, because in our stimuli the linguistic content of the remark was not accessible, this result might be considered the visual counterpart of the results of Mauchand, Vergis & Pell (2020), who found that when presented with auditorily comments with the task of focusing on the way they are pronounced (ignoring their linguistic content), participants are influenced by prosody (higher ratings for positive prosody than for negative prosody) and by content (higher ratings for positive content than for negative content).

2.4 Reanalysis of Giustolisi & Panzeri (2021), Study 4

As already anticipated, our second aim was to test the hypothesis that participants might rely on the perceived 'jocularity' of speakers to attribute them an ironic intent. In particular, in the Study 4 of Giustolisi & Panzeri (2021), which presented pairs of muted videos of speakers uttering sincere and ironic comments, participants often misinterpreted sincere compliments for ironic remarks. Specifically, 93 Italian speakers (76 females, 17 males, mean age = 23, SD = 4) evaluated the same pairs of sincere/ironic remarks for which we collected the friendliness ratings, with the task of detecting the ironic one, relying on visual cues

only. In general, the performance was higher in detecting ironic compliments (versus sincere criticisms, in pairs with negative content) than ironic criticisms (versus sincere compliments, in pairs with positive content). Since we collected the friendliness ratings for each of the videos used in that study, we can directly test the prediction that those videos that had higher friendliness ratings compared to their competing counterparts are more likely interpreted as ironic. To do so, for each pair, we calculated the difference between the mean z-score corresponding to the sincere and to the ironic comment. We called this difference "z-difference". When sincere items received higher friendliness ratings than their ironic counterparts the z-difference is positive; otherwise, it is negative. Moreover, the higher the absolute value of the z-difference, the bigger the gap between ratings.

We used the z-difference as predictor in the generalized linear mixed model analysis of participants' accuracy in Giustolisi & Panzeri (2021) Study 4.

2.4.1 Hypothesis

If participants associate the speaker's positive stance (i.e., friendliness) to jocularity, and rely on it to decide whether a comment is ironic, we expect accuracy rates to be influenced by z-difference. Specifically, we predict that accuracy in irony detection will be higher in case of negative z-difference, and lower for positive z-difference. In both cases, the higher/the lower accuracy will be dependent upon the z-difference absolute value. We will analyze this factor separately for comments with negative content (sincere criticisms and ironic compliments) and comments with positive content (sincere compliments and ironic criticisms) to see if the parameter of friendliness might explain the difference in detecting the ironic comment in the two types of pairs.

2.5 Results and analysis

Firstly, we visualized item accuracy as a function of item z-difference (Figure 2). As predicted, accuracy decreased as z-difference increased, with a steeper slope for comments with positive content (sincere compliments / ironic criticisms).

We ran a generalized linear mixed model analysis with content (negative coded as -0.5 and positive coded as 0.5) and z-difference (mean-centered) as fixed factor, with by participant and by item random intercepts and by participant random slopes for the effects of comment and z-difference and we estimated simple slopes.

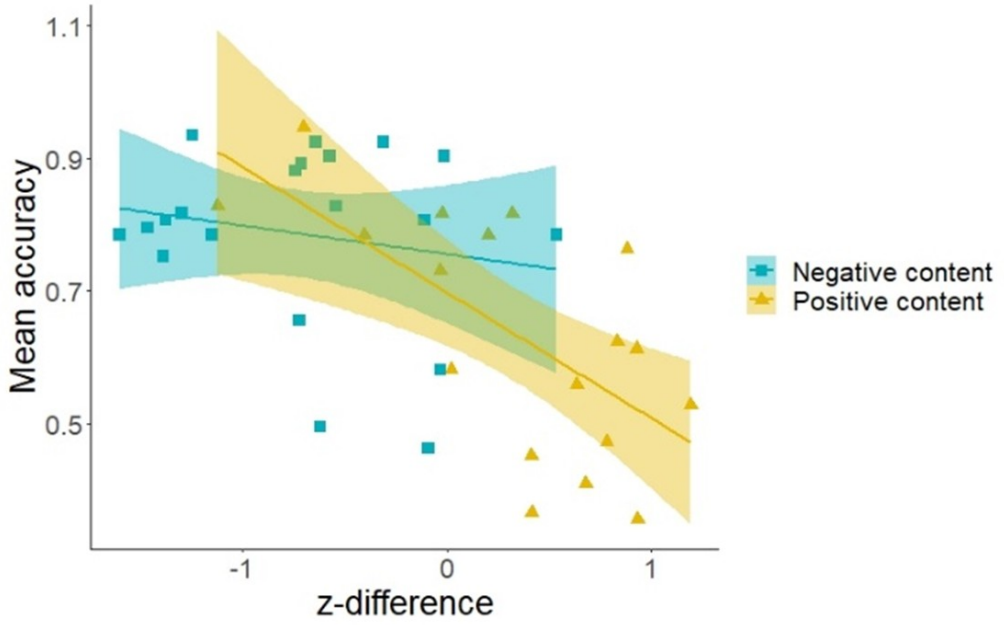

Figure 2: By item mean accuracy as a function of item z-difference for comments with negative/positive content. Points indicate items (composed by pairs of comments) and regression lines are plotted with 95% CI.

Considering comments with positive content, accuracy significantly decreased as z-difference increased ($\beta = -1.11, SE = 0.34, t = -3.24, p = .001$); for comments with negative content accuracy decreased as z-difference increased, but not significantly ($\beta = -0.59, SE = 0.34, t = -1.73, p = .085$).

3 General discussion

Our general goal was to investigate the nature of the visual cues connected to the production of irony and their role in irony detection. Inspired by Mauchand, Vergis & Pell (2020), who focused on acoustic irony markers and found evidence that interlocutors are sensitive to the speakers' affective stance, we asked 103 raters to evaluate the friendliness level of a series of sincere/ironic comments presented as muted videos, and we confirmed that speakers uttering compliments are perceived more friendly than those uttering criticisms, and that speakers uttering a comment with positive content (e.g. "It was a beautiful hike!") are perceived more friendly than speaker producing a comment with negative content (e.g. "You surely are bad at drawing!"). If we had to put friendliness ratings on a scale, the two extremes would be occupied by sincere comments, sincere criticisms on the lower step, and sincere compliments on

the higher step. Ratings for ironic compliments and criticisms would be in the middle, with close mean values. This pattern of results provides evidence that the muting function of irony (that softens the meanness of sincere criticisms but diminishes the benevolence of sincere compliments as well) is also conveyed through the ironist face, and not only acoustically.

In a previous set of studies (Giustolisi & Panzeri 2021) we showed that interlocutors can recognize ironic comments relying on speakers' facial expressions, at least when they know the content of the remark. When no linguistic cues were available, though, accuracy dropped, and participants mistook sincere compliments for ironic criticisms. Hypothesizing that interlocutors might attribute jocularity to irony, and thus use the perceived positive affective stance as a proxy to decide whether a speaker is ironic, our second purpose was to check whether the friendliness ratings we collected could account for the results of Study 4 of Giustolisi & Panzeri (2021). We showed that when presented with pairs of videos in which a person was uttering the same remark, once sincerely and once ironically, participants tend to attribute irony to the friendlier video of the pair, and, in general, the higher the difference in friendliness between the sincere and the ironic video, the higher the probability that the friendlier video is deemed as ironic. Specifically, in the case of remarks with positive content, i.e. sincere compliments and ironic criticisms, we found that the z-difference was a significant predictor of accuracy: the more the z-difference was below 0, the higher accuracy, the more the z-difference was above 0, the lower accuracy. Since sincere compliments were in general rated as friendlier than ironic criticisms (Figure 1), this can account for the high number of errors made by the participants of Giustolisi & Panzeri (2021)'s Study 4, mistaking sincere praises for ironic comments, and this fact can account also for the findings of Deliens et al. (2018)'s Experiment 1, in which sincere compliments received higher irony ratings than sincere criticisms. In the case of remarks with negative content, the trend was similar, but it did not reach statistical significance. In particular, the friendliness ratings of sincere criticisms were extremely low (Figure 1), determining a difference in affective stance between ironic compliments and sincere criticisms that was rather high. Probably, the lack of significance could be due to the extremely low polarity of z-difference in this condition. This, in turn, could explain why participants of Giustolisi & Panzeri (2021)'s Study 4 were so good in detecting irony among sincere criticisms and ironic compliments.

Our data thus support the hypothesis that irony is attributed relying on the perceived affective stance of the speaker, even when this strategy leads to misattributions if only visual cues are considered. Since irony

is associated with the friendliness / jocularity of a speaker, the positive affective stance that accompanies sincere praises might be mistaken for irony, compared to the more negative attitude associated with ironic criticisms. Indeed, ironic blames, which constitute the most common form of irony, are accompanied by a negative attitude of the speaker towards the thought echoed by the remark (Wilson & Sperber 2012) or towards the person who would be so foolish to entertain that thought (Clark & Gerrig 1984), especially in the case of sarcasm (irony directed towards a victim). Even if some scholars view irony as an umbrella term for figuratively conveyed meanings, covering both sarcasm and jocularity (and rhetorical questions, hyperbole and understatement as well, Gibbs 2000), Grice (1978) actually claimed that irony necessarily reflects a hostile or derogatory judgment or a feeling such as indignation or contempt, and he maintained that speakers who say something literally negative to convey the opposite meaning would have a playful attitude, but they should not be considered ironic. We interpreted the results of our studies as evidence that participants relied on the speaker's perceived friendliness / jocularity to infer ironic intents: This hypothesis, in line with the Asymmetry of affect and Tinge hypothesis, assumes, contra Grice (1978), that irony is a unitary phenomenon that permits to blame by praising, and to praise by blaming. Still, it could be objected that our results could be task-dependent: in Giustolisi & Panzeri (2021) participants performed a discrimination task, in which they were asked to detect the ironic comment, basing their answers on visual and/or prosodic cues. As Deliens et al. (2018) argue, though, this is an off-line, forced-choice, categorization task, and it does not necessarily pinpoint the actual comprehension of ironic speakers' communicative goals. It would be interesting to see whether our results could be replicated with a more ecological task, in which participants have to infer speakers' ironic intent.

Acknowledgements

Databases and scripts for the analyses are stored on OSF (https://osf.io/k3ade/).

References

Aguert, M. 2022. Paraverbal expression of verbal irony: Vocal cues matter and facial cues even more. *Journal of Nonverbal Behavior* 46(1). 45–70.

Attardo, S., J. Eisterhold, J. Hay & I. Poggi. 2003. Multimodal markers of irony and sarcasm. *Humor* 16(2). 243–260.

Bates, D., M. Mächler, B. Bolker & S. Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bryant, G. A. & J. E. Fox Tree. 2002. Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol* 17(2). 99–119.

Bryant, G. A. & J. E. Fox Tree. 2005. Is there an ironic tone of voice? *Language and Speech* 48(3). 257–277.

Burgers, C. & M. Van Mulken. 2017. Humor markers. In *The Routledge Handbook of Language and Humor*, 385–399. New York: Routledge.

Clark, H. H. & R. J. Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General* 113(1). 121–126.

Deliens, G., K. Antoniou, E. Clin, E. Ostashchenko & M. Kissine. 2018. Context, facial expression and prosody in irony processing. *Journal of Memory and Language* 99. 35–48.

Dews, S., J. Kaplan & E. Winner. 1995. Why not say it directly? The social functions of irony. *Discourse Processes* 19(3). 347–367.

Félix-Brasdefer, J. C. 2010. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Uso-Juan (eds.), *Speech Act Performance: Theoretical, Empirical, and Methodological Issues*, 41–56. Amsterdam/Philadelphia: John Benjamins.

Gibbs, Raymond W. 2000. Irony in talk among friends. *Metaphor and Symbol* 15(1–2). 5–27.

Giustolisi, B. & F. Panzeri. 2021. The role of visual cues in detecting irony. In *Proceedings of Sinn und Bedeutung*, vol. 25, 292–306.

Grice, H. P. 1978. Further notes on logic and conversation. In P. Cole (ed.), *Pragmatics: Syntax and Semantics*, vol. 9, 41–57. New York: Academic Press.

Holler, J. & S. C. Levinson. 2019. Multimodal language processing in human communication. *Trends in Cognitive Sciences* 23(8). 639–652.

Kotthoff, H. 2003. Responding to irony in different contexts: On cognition in conversation. *Journal of Pragmatics* 35(9). 1387–1411.

Li, S., A. Chen, Y. Chen & P. Tang. 2022. The role of auditory and visual cues in the interpretation of Mandarin ironic speech. *Journal of Pragmatics* 201. 3–14.

Mantovan, L., B. Giustolisi & F. Panzeri. 2019. Signing something while meaning its opposite: The expression of irony in Italian Sign Language (LIS). *Journal of Pragmatics* 142. 47–61.

Matthews, J. K., J. T. Hancock & P. J. Dunham. 2006. The roles of politeness and humor in the asymmetry of affect in verbal irony. *Discourse Processes* 41(1). 3–24.

Mauchand, M., N. Vergis & M. D. Pell. 2020. Irony, prosody, and social impressions of affective stance. *Discourse Processes* 57(2). 141–157.

McKinnon, S. & P. Prieto. 2014. The role of prosody and gesture in the perception of mock impoliteness. *Journal of Politeness Research* 10(2). 185–219.

Muecke, D. C. 1978. Irony markers. *Poetics* 7(4). 363–375.

Pexman, P. M. & K. M. Olineck. 2002. Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Processes* 33(3). 199–217.

R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. `https://www.R-project.org/`.

Wilson, D. & D. Sperber. 2012. Explaining irony. In *Meaning and Relevance*, 123–145. Cambridge University Press.