Joker face. Recognizing irony in the visual mode in spoken and signed language

Francesca Panzeri[1], Beatrice Giustolisi[1], Lara Mantovan[2] and Chiara Calderone[2]

[1] University of Milan-Bicocca, Italy
[2] Ca' Foscari University of Venice, Italy
francesca.panzeri@unimib.it, beatrice.giustolisi@unimib.it,
lara.mantovan@unive.it, chiara.calderone@unive.it

Abstract

A remark such as "The party was really funny!" can be interpreted literally, as a praise, or ironically, as a criticism. To avoid misunderstandings, the ironist can pronounce the comment with a particular intonational contour, the so called ironic tone of voice, and accompany it with specific facial expressions. Also in Italian Sign Language (LIS) non-manual markers can signal the ironic intent of the signer, and thus facilitate the detection of the intended meaning. We report the results of two experimental studies that aim at identifying the role played by purely visual cues in the detection of irony in spoken Italian, and at verifying whether ironic remarks in LIS are recognizable also by those who are not competent in that language. We discuss the findings in the light of the debate concerning the affective or grammatical status of facial expressions, in spoken and signed languages.

1   Introduction

When people communicate, they rely on verbal and nonverbal cues, that is, acoustical and visual markers, to transmit and interpret the message. Prosodic features may encode linguistic or affective information (but see Seddoh, 2002 for a critique of this dichotomy). As for linguistic prosody, different speech acts, such as statements, questions, and orders, can be disambiguated through specific intonation, and the interpretation of focused structures relies on the placement of stress on particular speech units (see Prieto, 2015 and references therein). In these cases, the contribution of prosody is viewed as linguistic because it has an effect on the interpretation of statements. Speakers may also use specific acoustic profiles to signal their own emotional attitude (e.g., anger, joy, contempt) towards a proposition, with prosody transmitting purely affective information (Hammerschmidt & Jürgens, 2007). Visual markers, that is facial expressions and bodily movements, are less studied. Humans have evolved an ability to infer from the speaker's facial expressions her emotions and other socially relevant states (McCullough et al., 2005 and references therein); still, it is not clear whether, in spoken languages, facial expressions transmit only the speaker's attitude, or whether they can also serve linguistic functions.

Specific facial expressions typically accompany the intonational contour associated to questions, orders, and to the marking of prominent or new information; it is still debated whether visual cues are simply parasitic on acoustical data, offering weak and redundant information with respect to the stronger auditory signals (as suggested by Scarborough et al., 2009; Krahmer et al., 2002; Srinivasan & Massaro, 2003), or whether they can provide independent evidence for the interpretation of the sentence. Recent studies found that hearers do rely on visual information to disambiguate between contrastive focus statements and echo questions (Borràs-Comes & Prieto, 2011), between double or simple negation interpretations (Prieto et al., 2013), and to identify the illocutionary force of a speech act (Domaneschi et al., 2017).

In signed languages, facial expressions can encode both linguistic and affective information, and they have thus been considered the analogue of intonation (Sandler & Lillo-Martin, 2006). The so-called non-manual markers (NMMs, facial expressions, head and body movements) may signal different lexical and syntactic structures, such as relative clauses, questions, conditionals, adverbials, and topics (Pfau & Quer, 2010). It is noteworthy that the very same facial movements can be used to express both linguistic/grammatical meanings and paralinguistic/affective information: furrowed eyebrows, for instance, signal content questions in different sign languages, but they are also associated to the expression of anger in general human communication (de Vos et al., 2009). Once that specific facial expressions have been related to linguistic constructions in a signed language (e.g., eyes wide in commands, head nods and tilts in advices for American Sign Language, Brentari et al., 2018), it can be hard to recognize whether they solve an affective or linguistic function.

Different diagnostics have been proposed to distinguish the grammatical from the affective status of NMMs: linguistic NMMs are more conventionalized, and exhibit less individual variation (Brentari et al., 2018); they are less gradient, because they give rise to a categorical interpretation (a sentence is either a question or not, whereas a person can be angry at various degrees, de Voos et al., 2009); they exhibit an abrupt onset and offset (Baker-Shenk, 1983); and they are aligned with manual signs and thus have a systematic distribution (Dachkovsky, 2005, de Vos et al., 2009). Besides these articulatory features, linguistic and affective NMMs differ in that the former, by definition, concern the interpretation of sentences, and thus have an influence on their acceptability, and are required by grammar (Reilly & Bellugi, 1996). Focusing on their accessibility, affective NMMs are potentially recognizable also by non-signers, since at least some basic facial expressions are claimed to be universal or at least cross-culturally identifiable (Darwin, 1872; Ekman, 1973/2006), whereas grammatical NMMs are claimed to be more arbitrary and thus inaccessible to non-signers (Hermann & Pendzich, 2014), with some exceptions (for instance, head nods for assertion and headshake for negation).

In the present work, we aimed at further exploring the questions of the role played by visual cues in disambiguating ironic from sincere statements, in spoken and in signed languages. A sentence like "What a great job" can be uttered sincerely, and thus be interpreted literally, or ironically, with the speaker signaling that she does not believe that the job was well done. To avoid misunderstandings, speakers can make their ironic intent manifest using irony markers (Attardo et al., 2003). Besides choosing particular linguistic expressions, such as extreme adjectives, or rhetorical questions, the ironist can pronounce the comment with the so-called ironic tone of voice (Cheang & Pell, 2008), and accompany it with particular facial expressions (e.g., winks, nudges, Muecke, 1978). Acoustical and visual cues can then help the hearers recognizing the intended meaning, disambiguating thus between a literal and an ironic interpretation. It has been found that hearers can correctly recognize ironic remarks, even in the absence of a preceding context, relying only on acoustical cues (Bryant & Fox Tree, 2002). To our knowledge, the contribution provided by only corporal movements has not been tested yet.

In a preceding study (Mantovan et al., 2019), our research group investigated the expression of irony in Italian Sign Language (LIS): having compared minimal pairs of literal compliments / ironic criticisms (such as "That's very beautiful!") and of literal compliments / ironic criticisms (such as "What an awful house!"), we noticed that specific NMMs signaled the signer's meaning (with more raised eyebrows, head nods and lateral tilts in ironic remarks than in their literal counterparts) and also the signer's attitude (with mouth corners up accompanying both literal and ironic compliments, and corners down co-occurring with sincere and sarcastic criticisms). Nevertheless, the linguistic or affective status

of these markers was not questioned, nor it was tested whether the presence of these markers could effectively facilitate the recognition of irony.

2   The studies

We discussed how in spoken languages ironic remarks are typically accompanied by a specific intonation and by particular facial expressions. Interlocutors are able to recognize ironic remarks relying only on the ironic tone of voice, whereas the contribution of purely visual information in the disambiguation of sincere/ironic comments have not been tested yet. Our first study aims at filling this gap, by comparing the rate of recognition of ironic statements in three different conditions: when interlocutors can rely on both visual and acoustical information; when they can only hear the remark; and when they can only see the facial expression and bodily movements of the speaker. Our goal is to determine whether irony can be detected with purely visual information.

As for signed languages, in LIS ironic remarks tend to be marked with particular NMMs. As we saw in the previous section, NMMs can serve a linguistic or an affective function. One of the diagnostics criteria to distinguish these roles questioned their accessibility: grammatical NMMs can be considered as more arbitrary, and thus recognizable at a higher rate by signers; affective NMMs on the other hand express the signer's emotional state, and thus they should be recognizable independently from linguistic knowledge, at least within the same cultural group. The second study aims at gathering some evidence in favor, or against, the grammatical status of ironic NMMs in LIS: a group of hearing participants who know LIS (yes-LIS), and another group who does not know LIS (no-LIS), were asked to detect ironic comments, in LIS and in Italian; we compared their rate of recognition of irony in the two modalities. We hypothesized that if ironic NMMs in LIS had a linguistic status, they would facilitate the recognition of irony in the yes-LIS group more that in the no-LIS group.

2.1 Methods

We used a Discourse Completion Task (Félix-Brasdefer, 2010) to obtain a semi-spontaneous elicitation of five minimal pairs of remarks by two groups of Italian adults: four deaf native signers and four Italian speakers who were videotaped while uttering the remarks. The same remark (e.g., "The party was really funny!") was elicited in one session after a context that triggered its literal interpretation and, in a different session, that took place at least two weeks after the first session, after a context that supported its ironic interpretation. All the remarks had a literal positive interpretation, and thus they corresponded to compliments in their sincere reading, and to ironic criticisms in their ironic interpretation (see Mantovan et al., 2019 for further details on the elicitation procedure for native signers: the same procedure was used also for Italian speakers). The 20 remarks produced by Italian speakers were then edited to obtain three versions: the original version with video and audio (VA), one with only the video (V), and one with only the audio (A). With this material, we prepared two web-based experiments.

Study 1 focused on spoken Italian only. Participants were shown the two utterances of the same remark, and were asked to indicate when the speaker was being ironical. The 20 remarks (the 5 minimal pairs of literal compliments / ironic criticisms uttered by four Italian native speakers) were presented in three different conditions: the first block comprised only pairs of videos, without audio (with the remark written on the screen); the second block only pairs of audio tracks; the final block displayed pairs of videos with audio. Every participant, then, was presented with a total of 60 items (20 pairs for each condition).

Study 2 included instances of ironic remarks in the two modalities, spoken Italian, and LIS. For this study, we only presented pairs of videos (without audio, but again with the uttered remark written on the screen) of the 8 actors (four native signers, and four Italian speakers) uttering the same remark in a sincere and in an ironic way, and partecipants had to recognize the ironical remark. The total number of items for this study was 40 (20 pairs in Italian, and 20 pairs in LIS).

The two studies were implemented on Qualtrics and distributed on-line.

The participants of Study 1 were 81 Italian adults (61 female), with a mean age of 23 years. The participants of Study 2 were 141 Italian hearing adults, 73 with no exposure to LIS (58 female, mean age: 26 years), and 68 who know LIS (63 female, mean age: 30 years).

2.2 Results

Since in Study 1 our goal was to compare the relative strength of visual and auditory information in the detection of irony, we first verified whether ironic remarks were indeed recognizable when both these cues were present. We thus looked at accuracy rates in the third condition, VA, and found that for two items accuracy was at chance level, and we thus eliminated these items. Moreover, since the experiment was web-based, we also wanted to exclude the possibility that participants were overall responding at chance level. We thus controlled that each participant responded correctly, always in the last VA block, to at least 12 items out of the remaining 18 (66% accuracy). Seven participants out of 81 did not meet this requirement, and were eliminated. The following analyses were then ran on 74 participants who responded to 18 items in the three different modalities. The rate of recognition of irony in the only video condition (V) was 83.71% (SD = 36.94), in the only audio condition (A) was 79.43% (SD = 40.44), and in the last condition, where both visual and auditory cues were present (VA), was 90.72% (SD = 29.02). How mean accuracy varied across modalities is depicted in Figure 1.
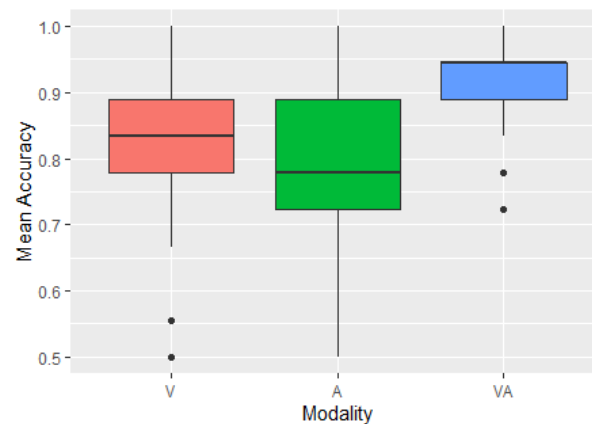

Figure 1: Boxplots depicting accuracy in the recognition of ironic remarks across modalities in Study 1.

Accuracy was analyzed using generalized mixed-effects models. Participant and items were entered as random intercepts in the null model. Adding modality as fixed factor significantly increased the goodness of fit of the model ($\chi 2$ = 72.41, p > 0.001). Post-hoc comparisons indicated that accuracy in the VA condition was higher than accuracy in both the V condition and the A condition (both ps < 0.0001); and also that accuracy in the V condition was higher than accuracy in the A condition (p = 0.009).

The aim of Study 2 was to verify whether the rate of recognition of irony markers in LIS was related to the knowledge of the language. We first checked whether the 40 critical items (5 remarks uttered by 4 Italian speakers and by 4 native LIS signers) were recognizable above chance level for at least one of the group of participants. One item had an accuracy around 35% for both groups and was thus eliminated from further analyses.

We then calculated the accuracy of irony recognition in the two languages (LIS and Italian), by the two groups of hearing participants, those who know LIS (yes-LIS), and those who do not know it (no-LIS). We found that the yes-LIS participants had an accuracy of 79.68% (SD = 40.25) in Italian, and of 79.23% (SD = 40.58) in LIS; the no-LIS group had an accuracy of 77.94% (SD = 41.48) in Italian and of 73.01% (SD = 44.40) in LIS. The results are illustrated in Figure 2.
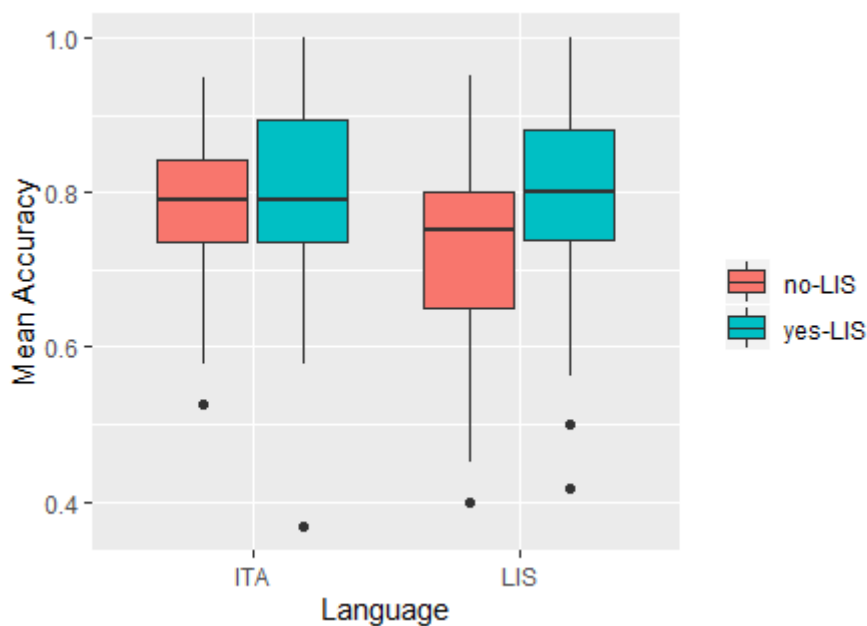

Figure 2: Boxplots representing accuracy in the recognition of ironic remarks in Italian (ITA) and in LIS by hearing participants who know LIS (yes-LIS) and who do not know LIS (no-LIS) in Study 2.

Accuracy was analyzed using generalized mixed-effects models. Participant and items were entered as random intercepts, whereas language (Italian vs. LIS) and knowledge of LIS (yes-LIS vs. no-LIS) as fixed factors. The two main effects were not significant (language: p = 0.33; knowledge of LIS: p = 0.43), whereas the interaction was significant ($\beta = 0.31$, SE = 0.14, z = 2.18, p = 0.03): ironic remarks in LIS were recognized better by the hearing participants who know LIS compared to those who do not know LIS. The group (no-LIS and yes-LIS) per language (LIS and Italian) effect is plotted in Figure 3.
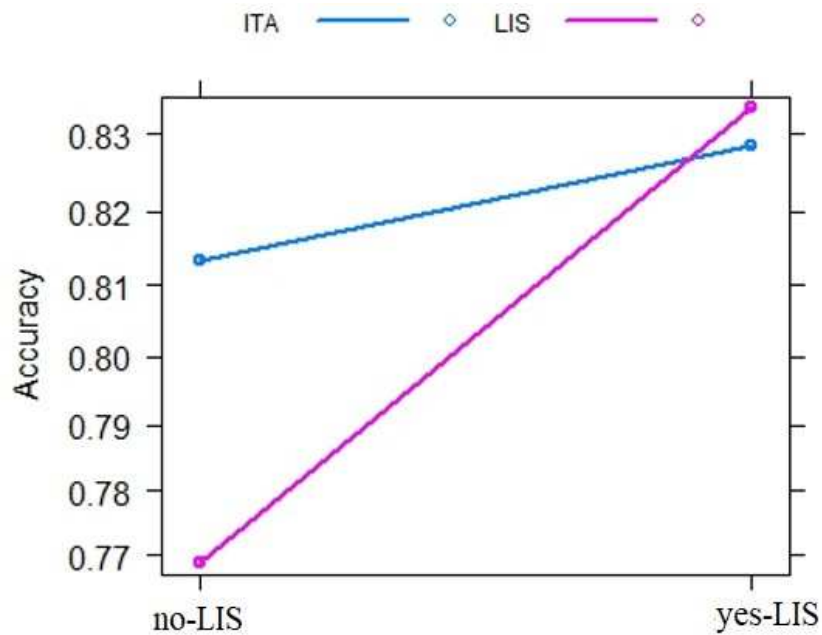
Figure 3: Participants (no-LIS and yes-LIS) by language (Italian and LIS) effect plot.

3 Discussion

In Study 1, we found that purely visual cues did enable interlocutors to correctly recognize ironic remarks, with an accuracy rate of 84%, well beyond chance level. Moreover, when participants had to distinguish ironic and sincere utterances of the same remark, they performed better when they were relying on purely visual cues than when their judgment was based on acoustical information. These results highlight the contribution of facial expressions and bodily movements in facilitating the correct interpretation of the intended meaning, and question the superiority of prosodic features in helping interlocutors in the interpretive process.

As for the debate concerning the affective or linguistic status of acoustical and visual cues in the detection of irony in spoken languages, our first study cannot settle the issue. Indeed, the recognition of the speaker's sincere or ironic intent has an effect on the interpretation of the sentence: the comment "The party was really funny!" constitutes a literally positive compliment under one reading, and ends up being a negative criticism on the other reading, with the speaker signaling that she does not believe that the party was funny. Traditionalist approaches to irony claim that with this type of figurative language the comment gets a reversed interpretation, and within this perspective, then, prosody and facial expressions would play a linguistic role, since they disambiguate the intended meaning. Nevertheless, both the Echoic account (Wilson & Sperber, 2012) and the Pretence account (Clark & Gerrig, 1984) to irony claim that the ironic remark is simply mentioned, preserving its literal meaning, and the ironist manifests to her interlocutors her scornful and negative attitude towards the proposition expressed, or towards the person who would hold such a view. The intonational contour would then have an affective nature, because the speaker would be

expressing acoustically her contempt towards the proposition she is echoing, or she would be exaggeratedly imitating the person she is pretending to be. Also facial expressions can be used to express the speaker's negative attitude, and thus they would only indirectly have an effect on the interpretation of the sentence.

Our second study, on the other hand, aimed at addressing the question of the affective or linguistic status of ironic NMMs in LIS. Following Brentari et al. (2018), we hypothesized that if the expression of irony were accompanied by grammaticalized NMMs, the detection of irony would be easier for those participants who know the language, and harder for those who are not familiar with it. We found that the participants who know LIS could detect ironic remark at a higher rate than those who do not know it. Still, this latter group of hearing individuals who are not acquainted with LIS could recognize irony with a 73% accuracy, which was lower than their accuracy for spoken Italian (78%), but still above chance level.

Taken together, the results of our two studies might suggest a more nuanced picture of the linguistic and affective status of prosody and facial expressions, both in spoken and in signed languages. Crasborn (2006) explicitly claimed, for signed languages, that "It is a commonly accepted intuition that linguistic non-manual signals have grammaticalized from emotional facial expressions, but few studies have explicitly addressed this hypothesis". Along the same lines, Benitez-Quiroz and colleagues (2016) argued that the facial expressions of negative moral judgment (anger, disgust and contempt) have evolved into a facial expression of ("linguistic") negation, which has then been grammaticalized as a non-manual marker in American Sign Language. This line of reasoning hypothesizes that some components of human languages, at least those related to prosody in spoken and signed language (i.e., NMMs), have evolved from facial expressions that encode the speaker/signer's emotions. If this is correct, then it should not be surprising that even facial expressions, and not only prosody, permit the detection of the speaker's ironic intent in spoken languages, and that even non-signers can recognize the expression of irony in LIS above chance level.

3.1 Limits and future directions

The studies we presented have important limitations that need to be acknowledged. First of all, participants were presented with two different utterances of the same remark, one sincere and one ironic, and were asked to recognize the sarcastic one. This is not an ecological setting, and the high accuracy in this task could come from a direct comparison of the two utterances. We are currently running a follow up experiment, in which the ironic and sincere remarks are presented one at a time, again in the three modalities (only video, only audio, and with both visual and acoustical cues). With this new design, we expect the overall accuracy to be lower, and that the contribution of purely acoustical information could override the role of visual cues, since the ironic intonation could be more conventionalized than facial expressions in the manifestation of irony.

Moreover, in our studies, we presented only remarks that were literally positive (such as "The party was really funny!"), and thus constituted compliments when the speaker was sincere, and criticisms when she was being ironical. As a consequence of this, we cannot be sure whether our participants were recognizing the speaker's ironical meaning or her scornful attitude, as manifested by the expression of her negative emotional state. It would be interesting to test also literally negative remarks, such as "This house is really awful!", that amount to criticisms when the speaker is sincere, and to compliments when she is being ironical. In our previous study on the expression of irony in LIS, we found that the position of the corners of the mouth was signaling the signer's attitude: when the signer had corners-up (a sort of smile), she was making a compliment, when the mouth corners were down she

was conveying a criticism. We hypothesized that these NMMs could be indirect signals of irony: when corners-up were co-occurring with a sign with a negative meaning (e.g., with AWFUL), or when corners-down accompanied a positive sign (e.g., FUNNY), the interlocutor could understand that the intended meaning should be reversed, and thus that the signer was being ironical. Testing the recognition of ironic compliments / literal criticisms in the three modalities could help us gaining a better understanding of the role played by visual and acoustical cues.

As for the second study we ran, on the recognition of irony in LIS by hearing participants who know and who do not know LIS, some weaknesses should be discussed. We found that the yes-LIS groups had an accuracy in the recognition of irony in LIS (but not in Italian) that was significantly higher than the no-LIS group, and we interpreted this result as possible evidence of the (partially) grammatical status of irony NMMs in LIS. Nevertheless, as pointed out by one reviewer, our participants were hearing persons studying LIS as second language (L2), and the mastery of linguistic NMMs is a late acquisition for L2 learners, at least for their correct production. We are currently running more fine-grained analyses that take into account the level of competence in LIS for each yes-LIS participant, to test whether knowledge of LIS is related to the accuracy of irony recognition.

Another analysis we are currently doing aims at disentangling the contribution of specific NMMs and other prosodic characteristics (such as the prolonged articulation of the evaluative sign) that are possibly related to the expression of irony in LIS. We plan to annotate their presence in each critical remark, and check whether there are cues that facilitate the detection of irony in LIS, and further verify whether these cues play a different role for the yes-LIS and the no-LIS participants.

References

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

Baker-Shenk, C. (1983). *A Microanalysis of the Nonmanual Components of Questions in American Sign Language.* Dissertation, University of California, Berkeley, CA.

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, 150, 77-84.

Borràs-Comes, J., & Prieto, P. (2011). 'Seeing tunes.' The role of visual gestures in tune interpretation. *Laboratory Phonology*, 2(2), 355-380.

Brentari, D., Falk, J., Giannakidou, A., Herrmann, A., Volk, E., & Steinbach, M. (2018). Production and comprehension of prosodic markers in sign language imperatives. *Frontiers in psychology*, 9, 770.

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

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech communication*, 50(5), 366-381.

Clark, H.H. & Gerrig, R.J. (1984). On the Pretense Theory of Irony. *Journal of Experimental Psychology*, 1, 121–126.

Crasborn, O.A. (2006). Nonmanual structures in sign language. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics (2nd edition)*, Oxford: Elsevier, 668-672.

Dachkovsky, S. (2005) *Facial expression as intonation in ISL: The case of conditionals.* Dissertation, University of Haifa.

Darwin, C. (1872). *The expression of emotions in animals and man.* London: Murray.

De Vos, C., Van der Kooij, E., & Crasborn, O. (2009). Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and speech*, 52(2-3), 315-339.

Domaneschi, F., Passarelli, M., & Chiorri, C. (2017). Facial expressions and speech acts: experimental evidences on the role of the upper face as an illocutionary force indicating device in language comprehension. *Cognitive processing*, 18(3), 285-306.

Ekman, P. (1973/2006). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review*, Los Altos (CA): Malor Books.

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*, Amsterdam/Philadelphia: John Benjamins, 41-56.

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice*, 21(5), 531-540.

Herrmann, A. & Pendzich, N.K. (2014). Nonmanual Gestures in Sign Languages. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D.McNeill & J. Bressem (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2, Berlin: de Gruyter Mouton, 2147–2160.

Krahmer, E., Ruttkay, Z., Swerts, M., & Wesselink, W. (2002). Pitch, eyebrows and the perception of focus, in *Speech Prosody 2002 Conference*.

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

McCullough, S., Emmorey, K., & Sereno, M. (2005). Neural organization for recognition of grammatical and emotional facial expressions in deaf ASL signers and hearing nonsigners. *Cognitive Brain Research*, 22(2), 193-203.

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

Pfau, R., & Quer, J. (2010). Nonmanuals: Their grammatical and prosodic roles. In D. Brentari (Ed.), *Sign languages*, Cambridge: Cambridge UP.

Prieto, P. (2015). Intonational meaning. *Cognitive Science*, 6(4), 371-381.

Prieto, P., Borràs-Comes, J., Tubau, S., & Espinal, M. T. (2013). Prosody and gesture constrain the interpretation of double negation. *Lingua*, 131, 136-150.

Reilly, J. S., & Bellugi, U. (1996). Competition on the face: Affect and language in ASL motherese. *Journal of child language*, 23(1), 219-239.

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*, Cambridge: Cambridge UP.

Seddoh, S. A. (2002). How discrete or independent are "affective prosody" and "linguistic prosody"'? *Aphasiology*, 16(7), 683-692.

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, 52(2-3), 135-175.

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46(1), 1-22.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*, Cambridge: Cambridge UP.