Recognizing irony through gestural cues in spoken and signed language

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

[1] University of Milan – Bicocca   [2] Ca' Foscari University of Venice

Introduction. When people communicate, they rely on verbal and nonverbal cues, i.e., acoustical and visual markers, to transmit and interpret the message. Prosodic features encode both linguistic (e.g., recognition of the intended speech act – statement, question or order; disambiguation of focused structures, see Prieto, 2015 and references therein) and affective (signaling the speaker's attitude towards a proposition, Hammerschmidt & Jürgens, 2007) information. Visual markers, that is facial expressions and bodily movements, on the other hand 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 language, facial expressions can encode also linguistic information. 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 these visual cues are parasitic on acoustical data (as suggested by Scarborough et al., 2009; Krahmer et al., 2002; Srinivasan & Massaro, 2003; House, 2002), 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 sign languages, on the other hand, it is well known that specific facial expressions and bodily movements may encode grammatical functions, signaling different lexical and syntactic structures, such as relative clauses, questions, conditionals, adverbials, and topics (Pfau & Quer, 2010). Non-manual markers (facial expressions, head and body movements) are thus viewed as ideal candidates for an intonational analogue in sign languages (Sandler & Lillo-Martin, 2006). We aimed at further exploring the questions of the role played by visual cues in disambiguating sentences, and of the linguistic or affective status of non-manual markers, focusing on the recognition of irony. A sentence like "What a great job" can be uttered sincerely, and thus be interpreted literally, or ironically, signaling the speaker's derogatory attitude towards that message. 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), still, 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), and suggested that specific non-manual markers signaled the signer's meaning and attitude; nevertheless the linguistic or affective status of these markers was not investigated. We thus conducted two experiments to compare (i) the rate of recognition of ironic statements using only visual with respect to only acoustic cues, and (ii) the rate of recognition of ironic remarks in Italian with only visual cues and in LIS by hearing participants who know LIS and by participants who do not know LIS.

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 native speakers 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, after a context that supported its ironic interpretation. The 20 remarks produced by Italian actors were then edited to obtain three versions: the original version with video and audio, one with only the video and

one with only the audio. With this material, we prepared two web-based experiments. Study 1 focused on Italian only, presenting 20 minimal pairs of remarks (the sincere and the ironic version of the same comment) in three different conditions, displayed in this order: only video (V), but with the remark written on the screen, only audio (A), and video with audio (VA). The participants were 81 adults, who had to recognize the ironic remark. In Study 2, we compared the recognition of ironic remarks in LIS and in Italian, hypothesizing that if non-manual markers of irony have a linguistic status, ironic remarks in LIS will be recognized at a higher rate by participants who know LIS. Again, participants were presented with 40 pairs of videos (20 in LIS and 20 in Italian, without audio, with the remark written on the screen) of the actors uttering the same remark in a sincere and in an ironic manner and were asked to recognize the ironic one. The participants were 141 Italian hearing adults, 68 of which who know LIS, and 73 who do not know LIS.

Results. In Study 1, we found that participants recognized ironic remarks with an accuracy rate of 84% in the V condition; of 79% in the A condition; and of 91% in the VA condition. We ran a Generalized Linear Mixed-Effects Model and subsequent post-hoc comparisons and found that accuracy in the VA condition was higher than in the V and in the A conditions (both ps <.0001); also accuracy in the V condition is higher than in the A condition (p = 0.0091).
In Study 2, hearing participants who know LIS had an accuracy rate of 80% in Italia, and of 70% in LIS; those who do not know LIS had an accuracy rate of 78% in Italian, and of 73% in LIS. We ran a a Generalized Linear Mixed-Effects Model and found that there was no difference in the recognition of ironic remarks in Italian between the two groups of participants, but that ironic remarks in LIS were recognized better by hearing participants who know LIS.

Discussion. In Study 1, we found that purely visual cues do enable interlocutors to correctly recognize ironic remarks, even better than only acoustical markers. These results highlight the contribution of facial expressions and bodily movements in facilitating the correct interpretation of the intended meaning. The results of the Study 2 indicate that the knowledge of LIS does improve accuracy in recognizing ironic remarks in LIS. This suggests that signer's facial expressions and bodily movements do not solve a purely affective function, but are grammaticalized, at least to some degree (Brentari et al., 2018). Benitez-Quiroz and colleagues (2016) claimed that the facial expressions of negative moral judgment (anger, disgust and contempt) have evolved into a facial expression of negation, which has then been grammaticalized as a non-manual marker in American Sign Language. Further research should explore this possibility also for the facial expressions that are associated to ironic criticisms, in spoken and signed languages.

Benitez-Quiroz, Wilbur & Martinez (2016). The not face: A grammaticalization of facial expressions of emotion. Cognition, 150; **Borràs-Comes & Prieto (2011)**. 'Seeing tunes.' The role of visual gestures in tune interpretation. Laboratory Phonology, 2(2); **Bryant & Fox Tree (2002)**. Recognizing verbal irony in spontaneous speech. Metaphor and symbol, 17(2); **Domaneschi, Passarelli & Chiorri (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); **Hammerschmidt & Jürgens (2007)**. Acoustical correlates of affective prosody. Journal of Voice, 21(5); **House, D. (2002)**. Perception of question intonation and facial gestures. Fonetik 44(1); **Krahmer, Ruttkay, Swerts & Wesselink (2002)**. Pitch, eyebrows and the perception of focus, in First International Conference on Speech Prosody; **McCullough, Emmorey & Sereno (2005).** Neural organization for recognition of grammatical and emotional facial expressions in deaf ASL signers and hearing nonsigners. Cognitive Brain Research, 22(2); **Pfau & Quer (2010)**. Nonmanuals: Their grammatical and prosodic roles. In Brentari, D. (ed.), Sign languages; **Prieto (2015)**. Intonational meaning. Cognitive Science, 6(4); **Prieto, Borràs-Comes, Tubau & Espinal (2013)**. Prosody and gesture constrain the interpretation of double negation. Lingua, 131; **Sandler & Lillo-Martin (2006)**. Sign language and linguistic universals. Cambridge UP; **Scarborough, Keating, Mattys, Cho & Alwan (2009)**. Optical phonetics and visual perception of lexical and phrasal stress in English. Language and Speech, 52(2-3); **Srinivasan & Massaro (2003)**. Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. Language and Speech, 46(1).