Gestural codas pave the way to the understanding of verbal irony

***Santiago González-Fuente[1], Mª Victoria Escandell-Vidal[2], Pilar Prieto[3,1]**

[1] Dept. of Translation and Language Sciences, Universitat Pompeu Fabra, Spain.
[2] Dept. of Spanish Language and General Linguistics, Universidad Española de Educación a Distancia.
[3] Institució Catalana de Recerca i Estudis Avançats, ICREA.

* **Correspondence:**
Santiago González Fuente
Dept. of Translation and Language Sciences
Universitat Pompeu Fabra
Barcelona, 08018, Spain
santiago.gonzalez@upf.edu

Abstract

It is well known that speakers rely on prosodic and gestural features at the time of producing and understanding verbal irony. Yet little research has examined (a) how gestures manifest themselves in spontaneous speech, both during and after ironic utterances; and (b) how the presence of the so-called 'gestural codas' (audiovisual cues produced after the ironic utterance) influences irony detection. In Experiment 1, spontaneously produced verbal irony utterances generated between pairs of friends in conversational dyads were analyzed for semantic, prosodic and visual contrasts. Results show that ironic utterances contrast with immediately preceding non-ironic utterances, both in terms of prosody and gesture. Experiment 2 tested the contribution of the presence vs. absence of such 'gestural codas' to the perception of verbal irony. An irony rating task was conducted in which participants were audiovisually presented with a set of ambiguous discourse contexts followed by a set of matching ironic and non-ironic utterances presented in two conditions, namely without coda and with coda. Results show that subjects detected the speaker's ironic intent significantly better when post-utterance codas were present (88%) than when they were not (56%), thus confirming the hypothesis that visual information produced after ironic sentences is a key factor in the identification of the speaker's ironic intent.

1. Introduction

From Classical times to the present, language philosophers, psycholinguists and pragmaticians have investigated verbal irony, a complex but common phenomenon whereby (in its most archetypal case) an individual chooses to say "Oh, great!" when he/she actually means "Oh, damn!" Classical accounts, as well as more current cognitive-pragmatic approaches, have stressed the fact that one of the key factors in understanding verbal irony consists of the recognition of some kind of contrast or 'incongruence' between two contradictory propositional forms involved in the whole speech act (i.e. between the expected proposition "Oh, damn!" and the actual proposition "Oh, great!") (Curcó 1995). This simple but critical assumption is contained, in some form or another, in the majority of the accounts of verbal irony proposed so far (e.g. Searle 1969, Grice 1975, Clark & Gerrig 1984, Gibbs 1994, Sperber & Wilson 1986/1995). In the Classical account of rhetorics,[1] irony is regarded as involving the replacement of a literal meaning with a figurative meaning, where this figurative meaning is in fact the opposite of the literal meaning. Thus, traditional approaches to verbal irony propose that we understand an ironic remark when we detect the contradiction between what has been said and what it is really meant. Similarly, conventional/logical approaches to verbal irony (e.g.

[1] See e.g. Quintilian's *Institutio Oratoria*.

Grice 1975) propose that the key to understanding an ironic remark relies on the detection of the incompatibility between its literal meaning and the pragmatic implicature inferred by the listener. Yet there are some cases that classical and conventional/logical accounts cannot explain, namely those in which speakers may mean what they are saying literally and yet still intend to be ironic. These ironic remarks cannot be evaluated in terms of truth conditions: the contrast that triggers the ironic interpretation is not produced by an incompatibility between the literal and figurative meanings of the ironic remark (i.e. when someone who loves surfing says "I love surfing" when confronted with a placid, waveless sea). To explain these cases, current cognitive-pragmatic approaches to irony propose a more complex vision of irony which is based on the human ability to simultaneously process contrasting information belonging to different levels. Thus, Gibbs (1994) claims that irony is a common form of thought through which humans juxtapose their expectations on reality. He adds that one of the internal functioning mechanisms of the phenomenon of irony consists in highlighting a discrepancy between expectations and reality (Gibbs 2012). One of the current cognitive-pragmatic accounts of irony is formulated within the framework of Relevance Theory (Sperber & Wilson 1986/1995, among others), which proposes that the cognitive Principle of Relevance assists us during the inferential processes. Within the relevance-theoretic approach, irony is understood as a pragmatic phenomenon that "consists in echoing a thought attributed to an individual, a group or to people in general, and expressing a mocking, skeptical or critical attitude to this thought" (Sperber & Wilson 1986/1995:125). Thus, what the speaker intends when he/she utters an ironic utterance is not "to provide information about the content of an attributed thought, but to convey her/his own attitude or reaction to that thought" (Wilson & Sperber 2012:128-129). When using verbal irony, speakers are simultaneously communicating propositional information as well as a critical attitude toward that proposition, together with their own disassociation from that attitude (Sperber & Wilson 1986/1995).

In natural conversation, speakers use a variety of linguistic strategies to mark their ironic intent, some of them being syntactic and discursive (e.g., Escandell & Leonetti 2014, Ruiz Gurillo 2008). Among these strategies, prosody has been analyzed very extensively. It has long been noted that speakers rely on prosodic signals when producing and perceiving verbal irony (see Bryant & Fox Tree 2002, 2005, Bryant 2010, 2011). Several studies have analyzed the prosodic properties of ironic utterances by comparing them to non-ironic ones (e.g. Gibbs 2000, Nakassis & Snedeker 2002, Anolli et al. 2002, Attardo et al. 2003, Laval & Bert-Erboul, 2005, Cheang & Pell 2009, Bryant & Fox Tree 2002, 2005, Bryant 2010, Scharrer et al. 2011, Padilla 2012). In general, ironic utterances have been reported to contrast with non-ironic utterances in their use of pitch modulations (e.g. lower or higher F0 mean and higher F0 variability values than their non-ironic counterparts), as well as intensity modulations (e.g. higher intensity values and variability) and duration changes (e.g. slower syllable durations, as well as more pauses). Other non-F0 features like non-modal voice quality have also been claimed to signal irony or sarcasm[2] (e.g. Van Lancker et al. 1981, Cheang & Pell 2008, 2009). Though some of these studies are based on read data produced with a purposeful stereotypic 'ironic tone', research has also shown that in spontaneous speech, verbal irony is not produced in a systematic fashion (Attardo et al. 2003, 2013, Bryant & Fox Tree 2005, 2010). In fact, it has been shown that irony does not necessarily have to be cued with overt linguistic marking and can be successfully interpreted by relying only on contextual cues. Despite this lack of systematicity, it is clear that speakers employ prosodic modulations when being ironic and that these modulations help listeners to infer irony by detecting a certain 'incongruence' between the coded meaning and the attitude (i.e. the 'actual intention') of the speaker. The complex nature of the phenomenon seems to indicate that speakers can signal the presence of verbal irony by combining and contrasting a variety of prosodic marks, this is, that "because of the inextricable relations between intentions and

[2] While some authors understand sarcasm as a subtype of verbal irony that is characterized by an explicit negative and critical attitude towards an event or person (Kreuz & Glucksberg 1989, Kumon-Nakamura, Glucksberg & Brown 1995, Cheang & Pell 2008), other authors use both terms (e.g. sarcasm and irony) interchangeably (Amenta et al. 2008, Attardo et al. 2003). For practical reasons, the literature reviewed in the present article includes studies dealing with the production and perception of both ironic and sarcastic utterances.

emotional tones of voice", prosodic signals specifically employed to highlight (i.e. to make 'relevant') an ironic remark overlap with the affective prosody embedded in the ironic utterances (Bryant 2010:546).

Within Relevance Theory, researchers have proposed that prosodic modulations encode procedural instructions that guide the inferential process by constraining the range of possible interpretations (Sperber & Wilson, 1986/1995, House 1990, 2006, Clark & Lyndsey 1990, Fretheim 2002, Wilson & Wharton 2006, Escandell-Vidal 1998, 2011a, 2011b, and Prieto et al. 2013, among others). In the case of irony, prosodic signals have been proposed to serve as guidance to help a listener understand a speaker's critical or ironic attitude with respect to the proposition expressed. Interestingly, recent research has shown the importance of gestural patterns in the detection of different types of pragmatic inferences (see e.g. Borràs-Comes et al. 2011, Goldin-Meadow 2003, Holler & Wilkin 2009, Prieto et al. 2011, Prieto et al. 2013, Krahmer & Swerts 2004, Swerts & Krahmer 2005). Thus it seems reasonable to hypothesize that visual cues might be as relevant as prosodic features in the production of ironic speech.

At this juncture, a relevant area of research is the study of the visual correlates of verbal irony. In conversation, speakers often use the so-called 'ironic gesture' (ironic winks, facial expressions involving specific eye and eyebrow configurations, laughter and smiles, etc.; see e.g. Gibbs 2000, Bryant 2011). Several studies have documented the presence of specific facial expressions during the production of verbal irony (Attardo et al. 2003, 2011, Bryant 2011, 2012, Haiman 1998, Hancock 2004, Kreuz 1996, Caucci & Kreuz 2012, Gibbs 2000). Bryant (2011), Attardo (2011) and Smoski and Bachorowski (2003) observed that laughter is typically used by speakers to indicate the presence of an ironic statement, as well as by listeners to mark the understanding of the ironic intention of the speaker (both in response laughter, as well as in laughter that occurs during or immediately after a social partner's laugh, e.g. the so-called 'antiphonal' laughter). These features have been claimed to express a positive stance between social partners and reinforce a shared positive affective experience (Smoski & Bachorowski 2003). Caucci and Kreuz (2012) recently found that one of the largest differences in facial cues between a set of 66 sarcastic and literal English utterances was the greater amount of smiling that occurred in sarcastic utterances. By contrast, other studies such as Attardo et al. (2003) reported that the most common visual cue to irony was in fact the absence of any facial expression, i.e. a sort of expressionless face produced after the ironic target pronunciation (i.e. during the coda following an ironic utterance), characterized as a "blank face" (Attardo et al. 2003:243).

The gestural marks mentioned above (smiles, facial expressions) can be understood as social signals that provide relevant communicative information about the ironic intent of the speaker. Another social signal of intentional meaning is gaze behaviour and recent work has found that gaze deviation is used by speakers when producing sarcastic utterances. Williams et al.'s (2009) experiments found that speakers deviated their gaze when being sarcastic in conversations with an unknown interlocutor. They measured eye contact between pairs of strangers when uttering sincere and sarcastic utterances and found statistically significant differences between the duration of eye contact occurring during sincere statements (63.9%) and sarcastic statements (52.7%). To our knowledge, no systematic studies have been performed on how gestural features (and gaze patterns) manifest themselves in spontaneous speech, both during and after the production of ironic utterances. Do ironic gestures appear more often during the pronunciation of ironic statements or after those statements? Moreover, to our knowledge, there have been no attempts to assess the role of visual cues (including the visual cues included in the codas produced after ironic sentences) in the production and successful understanding of ironic utterances.

The present study was designed to investigate (a) how consistently speakers used the abovementioned gestural cues both during and after the production of ironic statements in spontaneous discourse; and (b) the extent to which gestural codas affect the detection of irony. Experiment 1 was designed to collect spontaneous interactive data that favoured irony production. The rates of prosodic and gestural patterns were assessed as indicators of irony in spontaneous speech, both during and after the production of ironic utterances. It was predicted that we would

encounter higher rates of specific auditory and visual markers in ironic utterances than in their preceding non-ironic utterances, as well as a higher presence of gestural codas after ironic comments. Following up on the findings in Experiment 1, Experiment 2 was aimed at testing the potential effects of the presence of gestural codas on irony detection. Participants had to rate the presence of irony in a set of target utterances presented in an ambiguous context, in two coda conditions (the presence vs. absence of codas). It was expected that listeners would rely on the visual cues produced after the ironic utterance (i.e. gestural codas) for the detection of the ironic intent.

2. Experiment 1

2.1 Methods

2.1.1 Participants

A total of 22 Central Catalan speakers (19 women and 3 men; mean age = 22.24; stdev = 3.354) from the Barcelona area (mainly students at the Universitat Pompeu Fabra) participated in the study. They participated in pairs (11 pairs in total). It was a requirement that all pairs of participants knew each other previously, as other studies had suggested that ironic utterances occur considerably more often among friends or members of a family (e.g. Gibbs 2000). All participants were native speakers of Catalan, and they all considered Catalan to be their dominant language (relative to Spanish). Catalan dominance was 82.37% (stdev = 13.873), according to their own reports about the amount of time per day they spoke Catalan. All subjects participated voluntarily and gave informed consent to being audiovisually recorded, and all granted permission for usage of their data for research and educational purposes. They were each paid a small stipend (€5) for their participation.

2.1.2 Materials

The *stimulus* materials consisted of (a) two video sequences (henceforth named Video A and Video B; see two stills of each video in Figure 1) presented in an audiovisual mode and (b) a set of 8 sentences related to the videos (4 sentences per video), which were presented on two cards (see Example 1). The video sequences and sentences were selected in order to prompt incongruent contextual situations that would lead to spontaneous ironic responses (see 2.3 Procedure from the participants. Taking into account what Curcó (2000) and Morreall (1989) point out about the close relationship that exists between cognitive processes involved in producing and detecting both humorous and ironic utterances (where the perception of *incongruity* is the central element in achieving the humorous or ironic interpretation), two video clips related to the same situation (singing a song) were selected. First, Video A (2′ 45″) showed a group of amateurs performing an atrocious rendition of a song; and second, Video B (3′ 37″) showed a group of professional singers performing *a capella* with good vocal technique. While Video A conflicted with the expected situation of a singing performance, Video B showed a typical professional one (see the two panels in Figure 1).

Figure 1. Still images of Video A (left panel) and Video B (right panel).

The 8 prompt sentences consisted of a set of comments on the performances in Video A and Video B,

and they were given to the participants in written form after they had watched both videos in order to elicit their reactions (see Example 1 below for an example). For each video, there were four prompt sentences, two of them general comments by a commentator and two of them ostensibly comments made by singers in the respective group depicted. The sentences were designed to create a potential set of incongruities between the comment and the contextual situation, which would hopefully trigger the production of ironic utterances (in the case of Video A) and non-ironic utterances (in the case of Video B). Thus, while the contents of video and sentences were incongruent for Video A, they were reasonably congruent in Video B.

Example 1. Example of prompt sentence

Aquests cantants tenen un futur esplèndid al món de la música
'These singers have a splendid future in the world of music'

2.1.3. Procedure

The recordings took place in a quiet room at the Universitat Pompeu Fabra in Barcelona. Participants signed up for the experiment in pairs, with the understanding that they should have a relationship of friendship or family ties with the other person (this was a precondition for participation). Upon arrival, they were randomly designated as "Speaker A" and "Speaker B". As can be seen in Figure 2, the two participants sat in designated chairs facing each other about 4.5 ft. apart. In front of each participant was a laptop computer equipped with earphones, and next to the computer there was a card containing the 4 prompt sentences (Speaker A had the four Video A sentences and Speaker B the four Video B sentences). Three video cameras (a Panasonic *3MOS* HD-AVCCAM, a Sony *Handycam* HDR-CX115E and a Toshiba *Camileo* S20) were set up, two aimed at the two speakers, and the third one recording a wide shot of the full scene. Also, the experiment was audio recorded using a PMD660 Marantz professional portable digital recorder and a Rode NTG2 condenser microphone, which was situated on the table between the laptop computers.

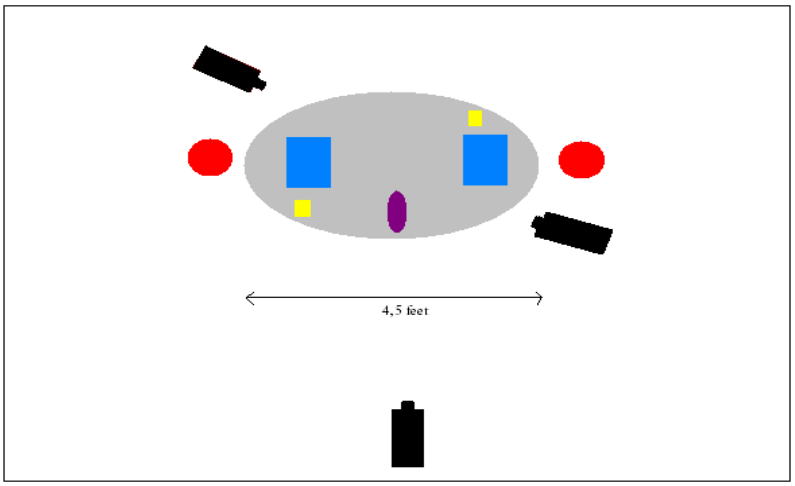
Participants were unaware of the real purpose of the study, and they were given no explicit instructions on to how interact. They were told that the goal of the experiment was to explore issues generically related to communication. To make the conversational interaction as natural as possible, no instructions about seating height, body posture or gestures were given.

The video stimuli were presented in a counterbalanced order alternatively for each pair of participants. They were both given the following written instructions: "You have two video files on the desktop of your laptop. Watch them simultaneously, discussing what you see. Your task will not be to describe their content, but rather to evaluate what you see, commenting freely, criticizing, praising, or even joking. You will listen to the audio track using only one earphone, so you can hear what your partner says and share impressions with him/her. When you finish watching Video A, close the lid of the laptop and do two things. First, exchange general impressions about the video. Second, the participant who has the card corresponding to Video A should read aloud the set of sentences on the card; as each sentence is read out, you must both react to it and make comments When you have finished, repeat this procedure with Video B." The participants were then left alone in the room, having been instructed to call the experimenter (the first author of the study) back into the room when they had completed the task.

All conversations were audiovisually recorded by the three cameras and the audio recorder. The video recordings were digitized at 25 frames per second, with a resolution of 720×576 pixels. The sample rate of the sound was 44,100 Hz using 16-bit quantization.

The total duration of the 11 recording sessions was 3 hours 26 minutes, with a mean duration of 19 minutes 38 seconds per experimental session.

Figure 2. Experimental setup. Laptops are represented as rectangles and the microphone as an oval figure on top of the oval table. Participants are represented as circles facing each other across the table, and the three video cameras as black rectangular shapes. The cards containing sentences are represented as small light-shaded squares next to the two laptops.



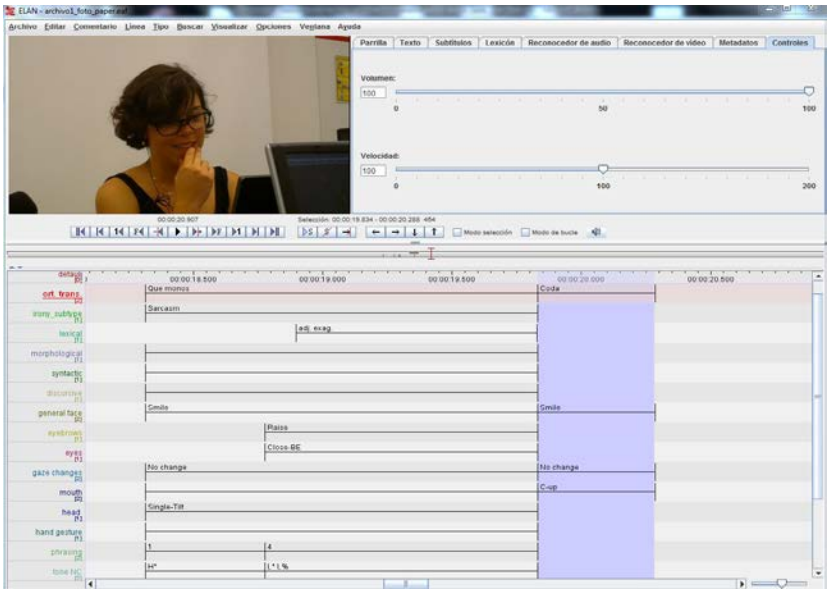
2.1.4 Data coding

First, the first author identified and extracted the ironic utterances from the 11 conversations (coming from both spontaneous exchanges and responses to the readings of ironic prompts). Whenever possible, any utterances that immediately preceded the ironic utterances (henceforth, baseline utterances) were also extracted. The selection of ironic utterances was made following the wide definition proposed by Gibbs (2000:13): "Each form of irony minimally reflects the idea of a speaker providing some contrast between expectations and reality."

The baseline and ironic target utterances were transcribed orthographically and a number of visual and auditory cues were manually annotated by the first author using ELAN (Lausberg & Sloetjes 2009).[3] All the pragmatic strategies (irony subtypes) and the lexico-syntactic and visual cues observed were annotated in different ELAN tiers, as is illustrated in Figure 3. Also, the prosodic characteristics of the target utterances were coded using Praat (Boersma & Weenink 2008) and automatically imported into ELAN.

Figure 3. Example of labelling with the target ironic sentence "*Que monos!*" ("How cute!").

[3] ELAN is an open source tool used for annotating and aligning transcriptions with video data.



A brief explanation of the coding used for every tier is presented below.

Orthographic transcription and presence of gestural codas (tier 1). The first tier was used to (a) perform an orthographic transcription of the target sentences and (b) code the presence or absence of visual cues after a sentence had been pronounced (labelled 'Coda' vs. 'No coda').

Coding of irony subtypes (tier 2). We followed Gibbs' (2000) proposal and labelled the following five irony subtypes: 'sarcasm', where the speakers spoke positively to convey a more negative intent; 'hyperbole', where the speakers expressed their non-literal meaning by exaggerating the reality of the situation; 'understatement', where the speakers conveyed their ironic messages by stating far less than was obviously the case; 'jocularity', where ironic speech was intended to tease or poke fun; and rhetorical questions, where speakers asked questions implying a critical or humorous intention.

Lexico-syntactic coding (tiers 3-6). Tier 3 was used to annotate exaggerated words and expressions (e.g. '*molt*' ['very'], '*meravellós*' ['wonderful'], '*m'encanta*' ['I love']), as well as mitigation words and expressions (e.g. '*una mica*' ['a little'], '*potser*' ['maybe']; see Scharrer et al. 2011). Tier 4 was used to annotate the presence of superlative or diminutive suffixes (e.g. '*moltíssim*' ['very much'], '*miqueta*' ['a little bit']). Tier 5 was used to annotate left dislocations (topicalizations) (e.g. *Entusiasmadíssima, estava* ["Very excited, she was"]; see Escandell-Vidal et al. (2014). Finally, tier 6 was used to annotate the use of code-switching and code-mixing, as well as direct speech in Spanish (e.g. '*I deia, "¡Guau! ¡Me están animando!"*' ['And he said, "Wow! They are cheering me on!"' —the framing is Catalan while the direct quote is in Spanish) and discourse markers such as '*bueno*', '*clar*', '*no*?' ['well', 'of course', 'right?'] (see Ruiz Gurillo 2008 and Muñoa-Barredo 1997).

Visual coding (tiers 7-13). Following Allwood et al.'s (2005) gestures coding proposal and McNeill (1992), the following gestural cues produced during and after the utterance of sentences were annotated:

General face (tier 7), i.e. the general impression that the coder received from the facial expression of the subject, taking 'Smile', 'Laugh', 'Scowl' or 'Neutral' values (see Table 1 for labelling of these gestures); *eyebrow movements (tier 8)*, i.e. when one or both eyebrows departed from neutral position; *eyes (tier 9)*, i.e. eyelid movements; *gaze changes (tier 10)*; *mouth (tier 11)*, i.e. mouth expressions in terms of lip shape; *head (tier 12)*, i.e. head movements; and *hand gestures (tier 13)*, i.e. arm and hand gestures.

Table 1 show picture stills of the facial and body gestures that were annotated most frequently in the corpus.

Table 1. Examples of facial and body gestures.

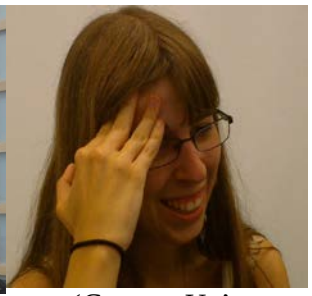
Tier	*Labelling examples*		
General Face	 ‘Smile’	 ‘Laugh’	 ‘Scowl’
Eyebrow Movements	 ‘Raise’	 ‘Frown’	
Eyes	 ‘Close both’	 ‘Squint’	 ‘Exaggerated Opening’
Gaze changes	 ‘Towards interlocutor’	 ‘Gaze Deviation’[4]	 ‘Towards materials’

[4] By “gaze deviation” we refer to some kind of brief and transitory shifting of gaze away from the interlocutor.

Mouth

'Stretched lips'

'Protruded lips'



'Corners-Up'

'Corners-Down'

Head

'Single Nod'

'Single Tilt'

'Shake'

Hand gestures

'Beat'

'Other' [5]

Prosodic coding (tiers 14 - 21).

Phrasing (tier 14). Following the Cat_ToBI proposal (Prieto 2014), the following break indices were annotated: prosodic units composed of clitics with content words were labelled '0'; word sequences '1-2'; end of intermediate phrases '3'; and end of intonational phrases '4'.

Tone nuclear configurations (tier 15). Again following the Cat_ToBI proposal (Prieto 2014), boundary tones (those associated with intonational boundaries) and pitch accents (those associated

[5] The 'Other' value includes metaphorical, deictic and iconic gestures (McNeill 1992).

with accented syllables) were labelled.

Voice quality (tier 16). In this tier voice quality features (labelled 'Creaky', 'Falsetto' or 'Breathy') were perceptually annotated and confirmed by examining their acoustic correlates using Praat (Boersma & Weenink~~,~~ 2008).

Finally, following Bryant (2010), the following values were extracted both in the baseline and ironic target conditions and annotated in tiers 17 to 20: *average pitch (tier 17)* and *pitch variability (tier 18)* (or standard deviation values) in Hz, *average loudness (tier 19)* in dBs, and *mean syllable duration (MSD) (tier 20)* in ms. To correct for between-speaker variability in F0 measurements, F0 values were converted to semitones (relative to 1 Hz). MSD was taken as a measure of speech rate and was calculated by dividing the total duration of the target utterance (in ms.) by the number of syllables.

Inter-rater reliability

To test the reliability of (a) the detection of ironic utterances and (b) the pragmatic, prosodic and gestural coding of target ironic utterances described above, an inter-transcriber reliability test was conducted with a subset of 20% of the data. Three independent coders labelled a random selection of the data following the guidelines described in the previous section (see 2.4). Since the total duration of the recordings amounted to 3 hours and 30 minutes, the reliability test involved 40 minutes of video (20% of the total play time). For the pragmatic and audiovisual coding, a random selection of 15 ironic target and baseline utterances was coded (specifically 6 ironic target utterances + 6 baseline utterances + 3 ironic utterances without previous baseline utterance), again constituting a total of 20% of the data.

The Kappa statistic (Randolph 2008) was obtained. This measure calculates the degree of agreement in classification over that which would be expected by chance and is scored as a number between -1.0 and 1.0, with -1.0 indicating perfect disagreement below chance, 0.0 indicating agreement equal to chance and 1.0 indicating perfect disagreement above chance. Since three raters were involved in our study, the Fleiss fixed marginal kappa statistical measure was used (Grassmann & Tomasello 2009, Iverson & Goldin-Meadow 2005). Fleiss's (1981:214) equally arbitrary guidelines characterize kappas over 0.75 as excellent, 0.40 to 0.75 as fair to good and below 0.40 as poor. The Fleiss fixed marginal kappa statistic obtained for the detection and classification of ironic utterances was 0.64 and 0.71 respectively; for verbal cues (considered overall), it was 0.81; for prosodic cues, it was 0.53 in tone nuclear configurations, 0.87 in phrasing and 0.84 in voice quality; for visual cues (also considered overall), it was 0.85; and, finally, for the annotation of laughter and response values, it was 0.86 and 0.92 respectively. The fact that the Fleiss kappa statistical measure was lower for tone nuclear configuration annotation than for the rest of the annotations might be due to the fact that raters had to choose among a considerably higher number of categories or because of the high level of experience that this type of phonological annotation requires (Escudero et al. 2012). We think that these scores reveal a substantial agreement among raters, especially in visual cues, and thus validate the annotations made in the corpus.

2.2 Results

A total of 47 ironic utterances were extracted from the database. Of these, 33 ironic targets had baseline utterances available for analysis (i.e. without *overlapping* issues). In this section we report the results of our analysis of the semantic and audiovisual data.[6]

[6] The distribution of ironic productions during the experimental session was as follows: 72% of the ironic

One of the most important results of Experiment 1 (it was in fact what led us to design perception Experiment 2) was the presence of gestural codas in 70% of ironic utterances, as compared to 27% in baseline utterances. Nonetheless, we present in this section an exhaustive report of all the variables examined in order to characterize the corpus that we obtained and also to compare our results with those previously reported in literature.

Irony subtypes

The most common irony subtype found in the 47 ironic utterances was 'jocularity' (34%), followed by 'hyperbole' (23%), 'understatement' (19%), 'sarcasm' (13%) and 'rhetorical question' (9%).

2.2.1 Lexico-syntactic cues

In this section, as well as in the following sections, results will be presented by comparing the target ironic utterances to the baseline utterances. As expected, lexical, morphological, syntactic or discourse verbal irony markers appeared more often in ironic target than in baseline utterances. A set of chi-square tests revealed that only the rate of appearance of lexical markers was significantly different in ironic vs. baseline utterances ($\chi^2(1) = 4.02$, at $p < 0.05$). Thus, no significant differences between morphological, syntactic or discursive cues were found between baseline and ironic utterances. However, though we find a similar percentage of utterances with discursive cues in both conditions, this is due to the fact that utterances in both conditions used a wide array of discourse markers. Yet when we analyze specific types of discursive cues, it is important to highlight the fact that 4 ironic utterances (12%) used code-switching or code-mixing (e.g. "*Estan una mica* colocadillos"[7] ["They're a little bit *stoned*."]), and 4 of them (12%) used direct speech in Spanish (e.g. (7.5) "I deia: '*guau, me están animando*'" ["And he said, 'Wow, they're cheering me on.'"]). By contrast, only one baseline utterance used code-switching or code-mixing and none used direct speech.

2.2.2 Prosodic cues

Tonal nuclear configurations

As expected, the typical tonal configuration of a broad-focus statement (e.g. L* L%) was more frequently found in baseline utterances than in ironic targets (81% and 67%, respectively). By contrast, ironic utterances were produced with more prominent configuration of emphatic meanings (such as L+H* L%, L* HL%, L* !H%, or L+H* L!H%). Similarly, ironic utterances were produced with interrogative nuclear configurations, as in the case of L* H% and L+H* H%. We did not observe any correlation between the nuclear configuration type and the irony subtype of the utterance.

Phrasing

Ironic utterances contained higher rates of prosodic breaks (e.g. those with a '3' or a '4' break index value) than baseline utterances (18% in baseline utterances vs. 45% in ironic utterances). A chi-square test showed that the difference between the two groups was statistically significant ($\chi^2(1) = 5.65$, at $p < 0.05$).

Pitch, intensity and duration measurements

---

target utterances were produced while watching and commenting on Video A (60% of them in response to trigger sentences, 40% during spontaneous interaction), and 28% while watching and commenting on Video B (38% of them in response to trigger sentences, and 62% in a spontaneous way).

[7] 'Colocadillos' is a Spanish word, not Catalan, in this example of code-mixing.

Table 2 shows the mean and standard deviation values of the four acoustic measures (namely, F0 mean and F0 variability, intensity mean and MSD [mean syllable duration]) across the two conditions, namely baseline utterances and ironic utterances.

T-tests were used to determine the independence of the means with 'utterance type' as the independent variable and the four acoustic dimensions as dependent variables. They showed that only MSD values were significantly different between baseline and ironic target utterances at $p < 0.05$.

Table 2. Mean and standard deviation values of the four acoustic measures across the 33 baseline and ironic target utterances. F0 and 'F0 variability' values are in semitones, intensity values are given in decibels, and MSD values are given in milliseconds.

	Baseline utterances		Ironic utterances	
Acoustic dimension	*M*	*SD*	*M*	*SD*
F0 (st.)	90,14	3,06	90,76	3,40
F0 variabilty (st.)	27,76	18,10	37,29	23,75
Intensity (dB)	63,39	3,75	64,38	4,98
MSD (ms.)	167	29	185*	45

Note. F0 = fundamental frequency (pitch); F0 variability = F0 standard deviation (pitch variation respect to F0 mean); intensity = amplitude; MSD = mean syllable duration. All semitone values are relative to 1 Hz. Significance '*' = $p < 0.05$.

To check for the potential effects of irony subtype on the prosodic measurements of ironic utterances, a repeated-measures MANOVA was used, with 'irony subtype' as independent variable and the four acoustic measures as dependent variables. As expected, we did not find any effect of 'irony subtype' on any of the four acoustic dimensions of ironic utterances. The overall model was not significant, $F = 1.12$, $p = 0.34$ ($\eta^2 = 0.19$).

Voice quality

The results of the voice quality analysis showed that, whereas 45% of the ironic utterances were produced with a non-modal voice quality, only 18% of the baseline utterances were produced with a falsetto or creaky voice. The results of a chi-square test showed that the presence of voice quality features was significantly different between baseline and ironic utterances ($\chi^2(1) = 8.05$, $p < 0.05$).

2.2.3 Visual strategies

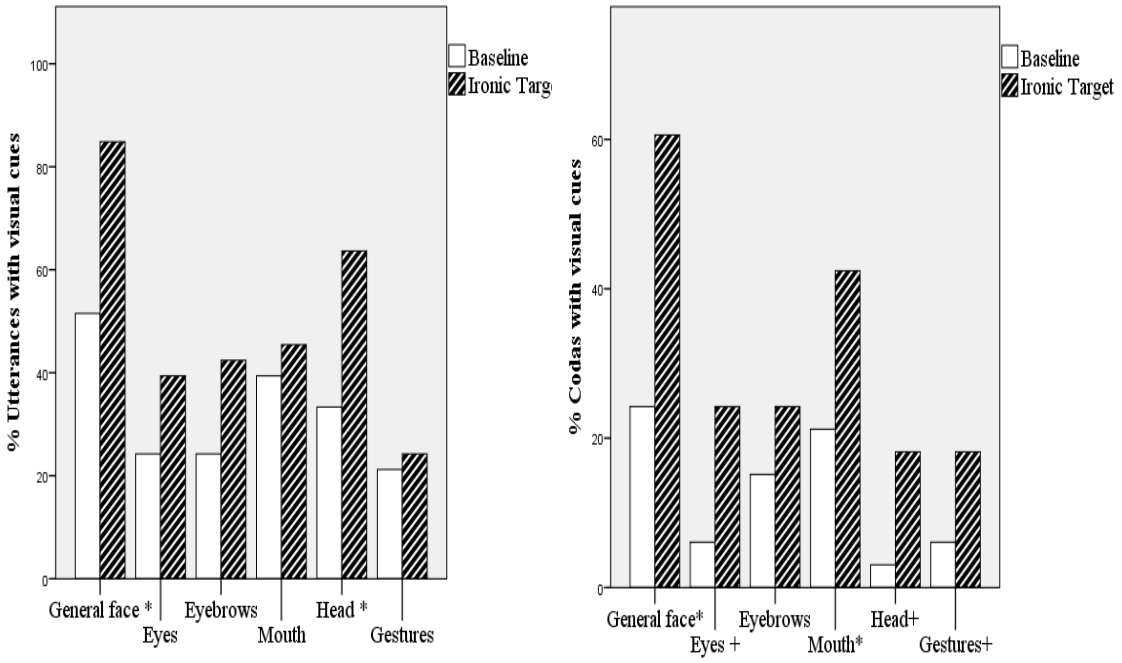
First of all, it is important to mention that a total of 70% of the ironic utterances were followed by a gestural coda, as compared to 27% in baseline utterances. The results of a chi-square test showed that utterance type (ironic vs. baseline) had a significant effect on the number of gestural codas ($\chi^2(1) = 10.24$, $p < 0.01$). For this reason the results in this section will be presented by separating the visual cues found into two conditions, namely 'During sentence pronunciation' and 'During utterance coda'.

General face, eyes, eyebrows, mouth, head and hand gestures

Figure 4 show the percentage of baseline and ironic target utterances in which 'General Face', 'Eyes', 'Eyebrows', 'Mouth', 'Head' and 'Gestures' values differ from 'None' both during the

utterance (left panel) and after, i.e. during the coda (right panel). These results General Face' and 'Head' for the non-coda condition and 'General Face' and 'Mouth' for the coda condition, as can be seen in Figure 5. Interestingly, speakers seem to mark the presence of irony quite systematically through the use of general facial expressions, either during the production of target utterances (85% of the cases) or during the codas (61%).

Figure 4. Percentage of utterances in which visual cues take a value different from 'None' during sentence utterances (left panel) and during codas (right panel) (y-axis). The results are broken down by visual cue ('General Face', 'Eyes', 'Eyebrows', 'Mouth', 'Head' and 'Gestures') and baseline (white solid columns) or ironic target (striped columns) condition (x-axis).



Note. In Figure 5, '*' indicates that $p < 0.05$ and '+' indicates that chi-square test not performed because the expected frequency was less than 5 in more than 20% of the cells.

Gaze

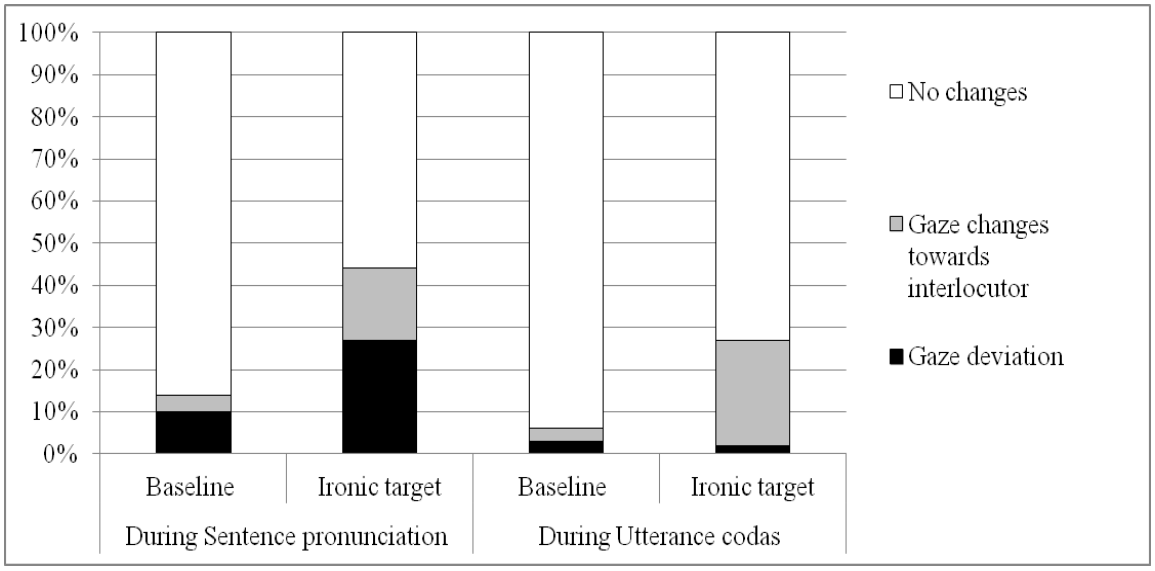
Figure 5 shows the results of gaze changes in two different conditions: (a) produced during baseline or ironic sentences; and (b) produced during baseline or ironic codas. It can be seen that speakers changed their gaze behaviour more often during the pronunciation of ironic utterances (in 44% of cases) than during baseline utterances (14% of cases).[8] In some instances these gaze changes involved a redirection of gaze from the experimental materials towards the interlocutor (grey-shaded part of columns) while in others it was redirected from the interlocutor to elsewhere ("Gaze deviation"—black-shaded part of columns). These results are in agreement with those described by Williams et al. (2009), who concluded that speakers tended to deviate their gaze from the interlocutor when being ironic.

The same pattern can be observed during the production of gestural codas, that is, speakers change their gaze pattern more frequently during ironic utterance codas (27% out of the 70% ironic sentences containing a gestural coda) than during baseline utterance codas (6% out of the 27% of baseline utterances containing a gestural coda). Interestingly, from a total of 27% of gaze changes that occurred during ironic utterance codas, 25% were gazes changes towards the interlocutor,

[8] This difference between utterances' type related to gaze patterns has been found to be significant ($\chi^2(1)= 6.08$, $p < 0.01$).

something which is consistent with the 'information-seeking' function of gaze proposed by Argyle and Cook (1976), among others.

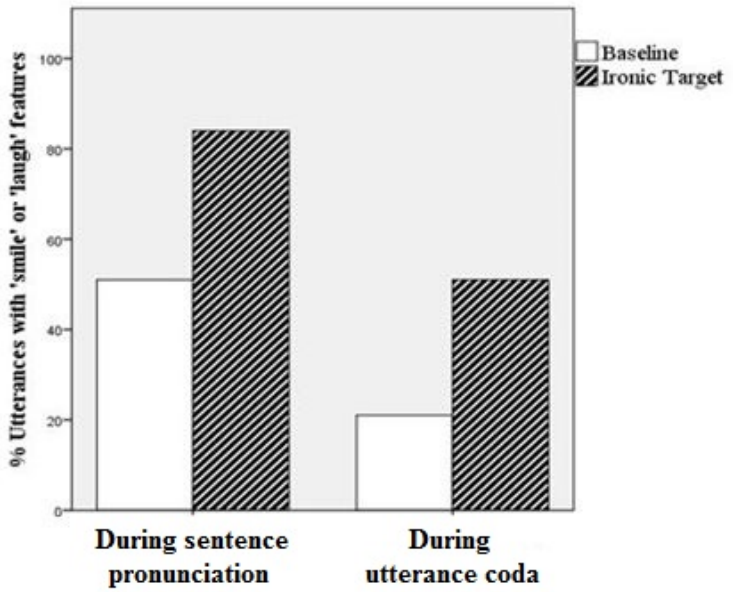
Figure 5. Percentages of utterances with 'No changes' (white part of columns); gaze change 'Towards interlocutor' (grey-shaded part of columns) and 'Gaze deviation' (black-shaded part of columns) values of 'Gaze' variable (y-axis). The results are broken down by location of appearance in target utterances (i.e. during sentence utterance—left columns—vs. during post-utterance codas—right columns), and baseline (columns 1 and 3) or ironic target utterance (columns 2 and 4) (x-axis).



'Laugh' and 'Smile' values of the 'General face' variable

The presence of laughter and smiling has been shown to play a strong communicative role in the expression of irony (Smoski & Bachorowski 2003, Bryant 2010, 2011). Figure 6 shows the percentage of utterances in which 'Laugh' or 'Smile' values of the 'General face' variable appear, both during the pronunciation of the sentence and during the coda. The results show that while speakers smiled or laughed (or did both) 84% of the time during the pronunciation of ironic utterances, they did so 51% of the time in the baseline condition. With respect to post-utterance codas, speakers produced higher rates of smiling or laughter (or both) during the production of ironic codas (51%) than during the production of non-ironic baseline codas (21%). The results of two chi-square tests showed that the utterance type variable was significantly related to the presence or absence of 'laugh' or 'smile', both in the case of sentence utterance ($\chi^2(1) = 6.08$, $p < 0.01$) and in the case of coda ($\chi^2(1) = 6.54$, $p < 0.05$).

Figure 6. Percentages of utterances with 'Smile' or 'Laugh' values of the 'General face' variable (y-axis). The results are broken down into baseline or ironic targets (solid white columns = baseline; striped columns = ironic targets) and the location of appearance of target utterance (e.g. during sentence or during coda) (x-axis).



2.2.4 Summary of lexico-syntactic, prosodic and visual cue results

In order to summarize these results, we compared the mean absolute number of lexico-syntactic, prosodic and visual markers appearing in ironic utterances with the mean number of marks appearing in baseline utterances. Multiple t-test analyses revealed that the absolute number of such cues were significant across ironic and baseline utterances (lexico-syntactic cues: $t(32) = 2.43$, $p < 0.5$); prosodic cues: $t(32) = 2.24$, $p < 0.5$); visual cues: $t(32) = 1.87$, $p < 0.5$). Interestingly, in our corpus ironic utterances showed a mean of 8.63 prosodic and visual cues (vs. 4.48 in baseline utterances), regardless of the pragmatic strategy (i.e. the irony subtype) employed by the speaker. In practical terms, this means that utterances were consistently marked with multimodal (prosodic and gestural) cues, with a combination of at least five audiovisual strategies. By contrast, the mean absolute number of lexico-syntactic cues was 0.53 for non-ironic utterances and 1.62 for ironic target utterances. If we compare the mean absolute number of visual cues to the number of prosodic cues, the concentration of visual signals is higher than prosodic signals. That is, while the mean absolute number of prosodic cues was 2.03 for baseline vs. 4.21 for ironic utterances, the mean absolute number of visual cues was 2.45 for baseline and 4.42 for ironic utterances. Interestingly, a mean of 1.93 visual cues (out of the total mean number of 4.42) appeared during gestural codas.

In sum, the results of Experiment 1 show that (a) speakers signal verbal irony through a varied set of prosodic and gestural cues; and (b) the presence of gestural codas is a consistent marker of verbal irony in this corpus (70% of the ironic utterances had some type of gestural coda containing visual markers). Such gestural codas contain visual cues that help the listener to understand the speaker's ironic intent by (1) conveying her/his attitude or emotion (through facial expressions, smiling/ laughter or head movements) and also (2) making explicit the speaker's desire to check for understanding of the ironic remark (through directing his/her gaze towards the listener). Though the results obtained for gestural codas in Experiment 1 were of great interest, the number of utterances obtained (33 ironic target and 33 baseline utterances) only allow us to make qualitative and not quantitative analyses. We therefore decided to run a perception experiment to specifically test the perceptual relevance of gestural codas for the understanding of irony.

3. Experiment 2

An irony rating task was designed to test the contribution of the presence vs. absence of gestural codas to the detection of verbal irony in ambiguous discourse contexts.

3.1 Methods

3.1.1 Participants

A total of 24 Catalan speakers (15 women and 9 men; mean age = 27.4; stdev = 7.7) participated in the experiment. They considered themselves to be Catalan-dominant and reported speaking in Catalan an average of 82% of the time (stdev = 7.23).

3.1.2 Audiovisual materials

In order to obtain the audiovisual materials to be used in Experiments 2 (i.e. the ironic and non-ironic performances of the target sentences), three native Catalan speakers participated in a Discourse Completion Task (henceforth DCT; Blum-Kulka 1989, Billmyer et al. 2000, Félix-Brasdefer et al. 2010). The DCT methodology consists of a semi-spontaneous elicitation task in which a given situational prompt is presented to the speaker, who is then asked to produce a given follow-up sentence in accordance with the stipulated context. A set of 4 discourse contexts were created by the authors, each one divided into 2 conditions, namely the 'non-ironic' condition, which was intended to trigger a non-ironic interpretation, and the 'ironic' condition, intended to trigger an ironic interpretation, as seen in Example 2 below. Crucially, the 4 follow-up sentences were created such that they were equally credible responses in both ironic and non-ironic discourse contexts (e.g. the follow-up comment "*Sembla que farà bo, avui!*" ["It looks like we're going to have great weather today!"] is equally apt in both (1a) and (1b)).

Example 2. Discourse context with two alternative contextual paths eliciting the same follow-up utterance:

> John and you are roommates and you are having breakfast in the kitchen, which is an interior room of the house with no windows.
>
> **(a)** Non-ironic condition
>
> You go outside to the balcony for a moment, see that is a sunny day, and when you go back to the kitchen, you say to John:
> "It looks like we're going to have great weather today!" (target follow-up utterance).
>
> **(b)** Ironic condition
>
> You go outside to the balcony for a moment, see that it is raining cats and dogs, and when you go back to the kitchen, you say to John:
> "It looks like we're going to have great weather today!" (target follow-up utterance).

Importantly, to prevent the participants' biases from affecting their interpretation of the scenario (and thus their rendering of the sentence), information related with social class, job and the particular interests of the characters in the scenario was not presented. Most importantly, the discourse context (which would prompt either an ironic or a non-ironic utterance performance) was designed to affect the two characters in the same way.

Participants read the prompt contexts and were then recorded producing the stipulated follow-up sentences in a quiet room at the Universitat Pompeu Fabra with a Panasonic AG-HMC41 professional digital video camera. Since head movements and facial expressions were relevant for our research purposes, speakers were asked to face the camera and were filmed against a white

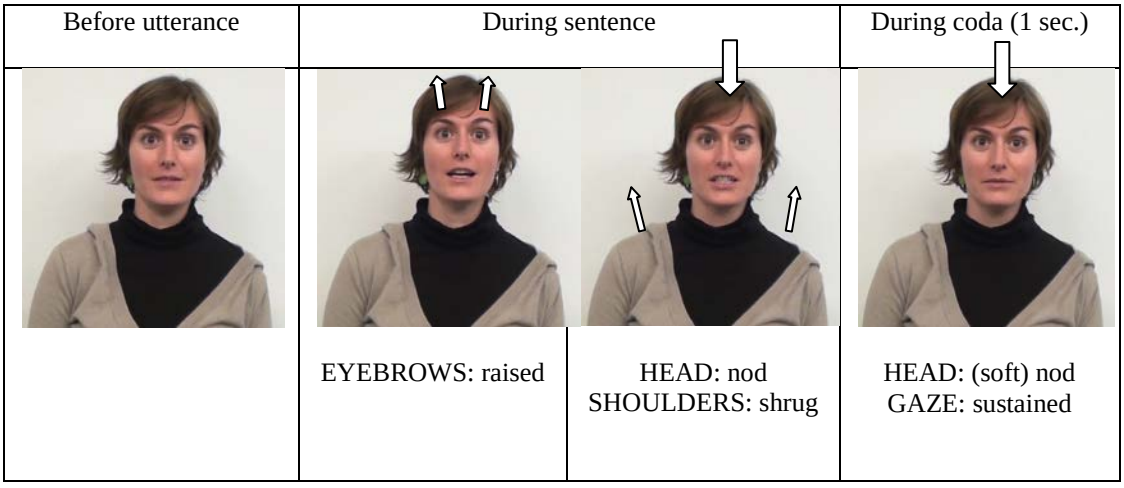
backdrop, with heads and upper bodies fully included within the video frame. The video recordings were digitized at 25 frames per second, with a resolution of 720 x 576 pixels. The sound was sampled at 44,100 Hz using 16-bit quantization. A total of 24 utterances were obtained, that is, 12 ironic utterances and 12 non-ironic utterances (3 speakers x 4 discourse contexts x 2 conditions—non-ironic vs. ironic).

In order to assess the prosodic cues to non-ironic and ironic utterances, the 24 target follow-up sentences were acoustically analyzed with Praat (Boersma & Weenink 2008) and coded prosodically following the Cat_ToBI system (Prieto 2014). In general, the most systematic differences between sincere and ironic utterances were (a) their nuclear tone configuration pattern and (b) their duration. Sincere utterances were performed with a L*L% nuclear configuration pattern (91% of sentences) and ironic utterances with a L+H* L% pattern (82% of sentences). Ironic utterances were also produced at a slower tempo in 65% of cases.

With respect to gestural cues, the 24 target follow-up utterances were analyzed with ELAN (Lausberg & Sloetjes 2009) following the criteria used in Experiment 1 (see Figure 7 for a set of examples).

Figure 7. Examples of the typical gestures produced during non-ironic and ironic performances of target sentences.

(a) Non-ironic performance of target sentence "*Sembla que farà bo, avui!*" ("It looks like we're going to have great weather today!")



(b) Ironic performance of target sentence "Sembla que farà bo, avui!" ["It looks like we're going to have great weather today!"]

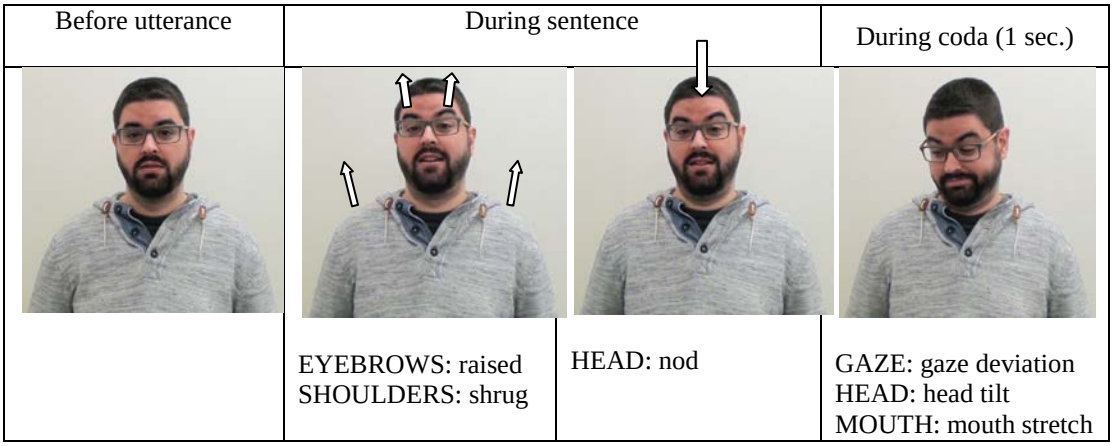


Table 3 shows the percentages of gesture types occurring in ironic and non-ironic utterances broken down by moment of occurrence (i.e. during sentence utterance or during post-utterance codas). First, it will be noted that a different range of gesture types appears in ironic utterances relative to non-ironic utterances. In non-ironic utterances, the most common visual cue was head nodding (83%), which might be indicating some kind of 'agreement' with the literal meaning of the sentence. By contrast, head movements such as shaking and tilting, as well as mouth stretching (all of them suggesting some kind of contradiction) appeared only in ironic utterances and to a higher degree during ironic utterance codas than during ironic sentences.[9] Though eyebrow raising and shrugging are present in both ironic and non-ironic utterances, ironic utterance codas show a higher rate of eyebrow raising (66%) than non-ironic codas (8%). Second, ironic utterance codas presented a higher rate of gestures than non-ironic utterances codas.

Table 3. Percentage of gesture types occurring in ironic and non-ironic utterances, broken down by moment of occurrence (during sentence or during coda).

Utterance type	Gestures	During sentence	During coda (1 sec.)
Ironic utterances	Mouth smile	16%	16%
	Mouth stretching	8%	33%
	Eyebrow raising	25%	66%
	Eye squinting	17%	8%
	Head shaking/tilting	17%	50%
	Head nodding	8%	0%
	Shrugging	17%	8%
Non-ironic utterances	Mouth smile	33%	41%
	Mouth stretching	0%	0%
	Eyebrow raising	25%	8%
	Eye squinting	0%	0%
	Head shaking/tilting	0%	0%
	Head nodding	83%	50%
	Shrugging	33%	0%

Table 4 shows the percentages of occurrence of sustained gaze vs. deviated gaze in ironic and non-ironic utterances, broken down by moment of occurrence (during sentence or during coda). First, 100% of non-ironic performances were produced with a sustained gaze towards the camera, both during the sentence utterance and during the coda (even in the 5 cases in which non-ironic utterance codas did not present any gestural cues). By contrast, ironic performances showed some gaze deviations during the sentence utterance (33%) but only in one case during the coda (8%).

Table 4. Percentage of occurrence of sustained gaze vs. deviated gaze in ironic and non-ironic utterances,

[9] With respect to 'smiles', contrarily to the results of Experiment 1, smiles appeared more frequently in non-ironic sentences (33% during sentence and 41% during codas) than in ironic ones (16% and 16%), which can be explained by the differing experimental conditions in the two experiments: in Experiment 1 non-ironic and ironic utterances were produced consecutively in the context of a conversation among friends, and in Experiment 2 the non-ironic and ironic target sentences were elicited by means of a DCT task in which participants produced both types of sentences as if addressing the camera, so the communicative function of using smiles to convey humour may have been affected by the absence of a real interlocutor.

broken down by its moment of occurrence (during sentence or during coda).

Utterance type	Gaze	During sentence	During coda (1 sec.)
Ironic utterances	Sustained gaze	67%	92%
	Deviated gaze	33%	8%
Non-ironic utterances	Sustained gaze	100%	100%
	Deviated gaze	0%	0%

In general, the gestural and eye gaze characteristics of the ironic vs. non-ironic performances of target sentences are consistent with the results of Experiment 1, in terms of both gestural and eye gaze patterns with specific gestures and patterns of gaze deviation characterizing ironic productions. In the case of ironic gestural codas, the gaze behaviour that characterizes them is sustained gaze.

3.1.3 Materials

The 24 recorded utterances obtained from the DCT materials were digitally edited using Adobe Premiere CS5 to obtain two sets of materials. For the 'Coda' condition, the 24 videos contained the pronunciation of the target sentence plus 1 second of the utterance coda. For the 'No- coda' condition, these same 24 videos were edited and the coda was deleted (i.e. they only contained the pronunciation of the target sentence). The resulting 48 videos were used as stimuli for Experiment 2.

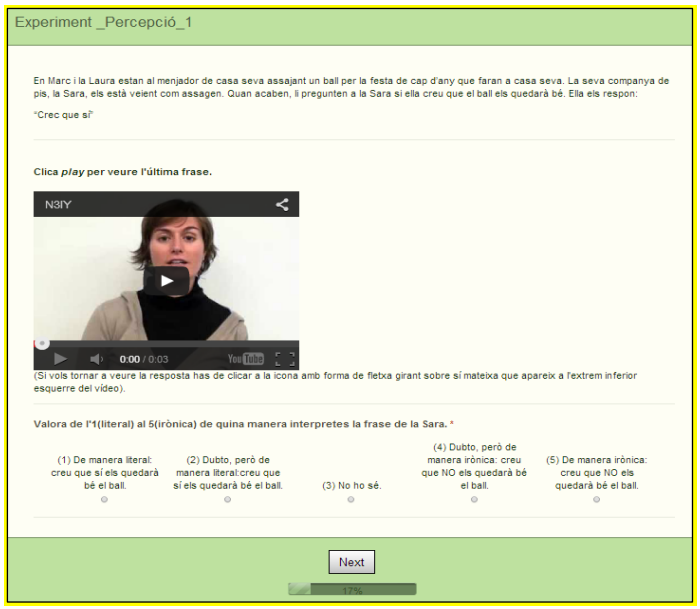
The discourse contexts used in the DCT task (see section 3.1.2) were adapted in such a way that they would be ambiguous and would not offer any clue about the ironic vs. non-ironic interpretation of the follow-up utterance (see Example 3). The ambiguity of the context was intended to ensure that the interpretation of the follow-up utterance as ironic or non-ironic would depend exclusively on how it was performed, that is, on auditory and visual cues.

Example 3. Ambiguous discourse context.

> John and Peter are roommates and are having breakfast in the kitchen, which is an interior room of the house with no windows.
>
> John goes outside to the balcony for a moment, and when he comes back to the kitchen he says to Peter:
>
> "It looks like we're going to have great weather today!" (target follow-up utterance).

The experimental materials were prepared using SurveyGizmo (Vanek & McDaniel 2006) (open-source software for generating and administering online questionnaires) (see Figure 8).

Figure 8. Survey Gizmo screenshot



Because the recordings selected involved the same speaker performing both ironic and non-ironic target utterances, two separate sets of experimental materials were designed in order to avoid subjects having to assess the same speaker producing both ironic and non-ironic interpretations of the utterance. Each set of materials consisted of 24 ambiguous discourse contexts followed by recorded responses presented in one of two coda conditions (i.e. 'with coda' or 'without coda'), and in one of the two utterance performance conditions (i.e. 12 in 'non-ironic' condition and 12 in 'ironic' condition).

3.1.4 Procedure

Participants completed one of the two versions of the two online audiovisual questionnaries. After reading each discourse context, they were asked to listen to an audiovisual recording of someone responding to the context. For each recording, listeners were asked to rate the degree of perceived irony expressed by the speaker on a 5-point Likert scale (from 1 = non-ironic to 5 = ironic). They could read the context and listen to/watch the recording as many times as they wanted.

A total of 576 responses were obtained (24 participants (12 for questionnaire 1 + 12 for questionnaire 2) x 24 questions). The mean duration of this experiment per participant was 16 minutes.

3.1.5 Measures and statistical analyses

The 576 responses were analyzed with a Generalized Linear Mixed Model (GLMM) using IBM SPSS Statistics 20.0 (IBM Corporation 2011), with 'Perceived degree of irony' as a dependent variable. The fixed factors were 'Utterance performance' (2 levels: 'Non-ironic' intonation/gesture vs. 'Ironic' intonation/gesture), and 'Presence of coda' (2 levels: 'With coda' and 'Without coda'). Subject and Item (a random combination of 'Speaker of the utterance' and 'Discourse context') were set as random factors.

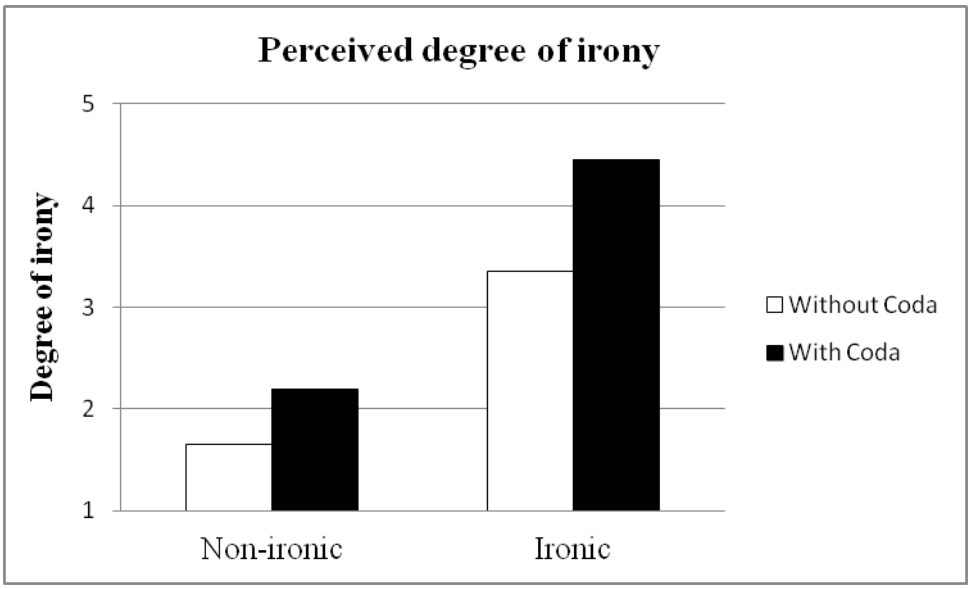
3.2 Results

A GLMM analysis was run with Perceived Degree of Irony as dependent variable, with 'Utterance performance' (2 levels: 'Non-ironic' intonation/gesture vs. 'Ironic' intonation/gesture), and 'Presence of Coda' (2 levels: 'With Coda' and 'Without Coda') as fixed factors. A main effect of 'Utterance performance' was found ($F(1,572) = 350.46$, $p < .001$), as well as a main effect of 'Presence of coda'

($F(1,572) = 52.94$, $p < .001$). Post-hoc analyses revealed a significant interaction between 'Utterance performance' x 'Presence of coda' ($F(1,572) = 6.52$, $p < .005$), indicating that the effect of 'With coda or 'Without coda' presentation on the 'Perceived degree of irony' variable differed depending on whether the target sentence had been produced with an ironic or a non-ironic intent.

Figure 9 shows the mean irony ratings (from 1 'Non-ironic' to 5 'Ironic', y-axes) as a function of two conditions: (a) 'Non-ironic' (left columns) and 'Ironic' (right columns) utterance performance conditions (x-axes) and (b) 'Without coda' (white columns) and 'With coda' (black columns) conditions. These results show that irony ratings were higher in the conditions where utterances were followed with codas, both for ironic and non-ironic utterances. As expected, gestural codas increased irony detection after the production of ironic sentences. Yet even more interestingly, even in the 'Non-ironic' condition the presence of a visual coda (with no clear ironic visual cues; see Figure 7 and Table 3) had the effect of increasing the irony detection.

Figure 9. Average irony scores (from 1 'Non-ironic' to 5 'Ironic', y-axes) as a function of two conditions: (a) 'Non-ironic' (left columns) and 'Ironic' (right columns) utterance performance conditions (x-axes) and (b) 'Without coda' (white columns) and 'With coda' (black columns) conditions.



The results of Experiment 2 clearly show that visual codas produced after ironic utterances help listeners to understand the speaker's ironic intent. Interestingly, this boosting effect was also present when utterances were performed non-ironically.

4. Discussion and conclusions

It is well known that prosodic and visual cues are important ingredients of communication. Recent work has convincingly shown that speech and gestures form a unique and unified system (McNeill 1992, Cartmill et al. 2012) and that prosodic and gestural patterns are key in the detection of different types of pragmatic inferences (see e.g. Borràs-Comes et al. 2011, Goldin-Meadow 2003, Holler & Wilkin 2009, Prieto et al. 2011, Krahmer & Swerts 2004, Swerts & Krahmer 2005). In the domain of the expression and detection of irony, this article has examined two main questions, namely (a) how prosodic and gestural features manifest themselves in spontaneous non-scripted speech, both during and after ironic utterances; and (b) how the presence of the so-called 'gestural codas' (audiovisual cues produced after the ironic utterance) influences irony detection.

In Experiment 1, spontaneously produced ironic utterances were analyzed for semantic, prosodic and visual contrasts and compared with their preceding baseline utterances. Results showed that speakers contrast ironic utterances with immediately preceding non-ironic utterances, in terms of both prosody and gesture. With respect to prosody, results show that relative to baseline utterances ironic speech is characterized by a significantly higher rates of emphatic tone nuclear configurations (20% incidence of L+H* L%, L+H* L!H% and L!H% in ironic target utterances vs. 3% in baseline utterances) and a more frequent presence of higher-level prosodic phrases (45% in ironic target utterances vs. 18% in baseline utterances). The phrasing results agree with Potts (2005), who claims that ironic speech is characterized by multiple intonational phrases that tend to highlight each word of the target sentence. Of the four acoustic dimensions analyzed (namely, F0 mean and F0 standard deviation, mean syllable duration and mean intensity), only mean syllable duration (a measure of speech rate) was found to be significant. Speakers produced ironic utterances at a significantly slower speech tempo than baseline utterances. A decrease in speech rate has been documented as one of the prosodic regularities that signal the presence of ironic intent across languages (Anolli et al. 2002, Bryant 2010, Laval & Bert-Erboul 2005, Scharrer et al. 2011). Bryant (2010:556) suggests a cognitive explanation for this pattern, as follows: "Slowing down speech gives the listener more time to process the relatively higher propositional load often contained in verbal irony, compared to literal interpretations of the same utterances." With respect to the behaviour of pitch variability (F0 variability values) as well as mean pitch and intensity, results showed no directional tendencies for ironic speech, being higher or lower in ironic utterances than in their baseline counterparts. Previous results have also showed inconsistent prosodic patterns across studies and across languages. While mean F0 values have been shown to increase in Italian and Cantonese sarcastic irony (Anolli et al. 2002, Cheang & Pell 2009), as well as in French sarcastic requests (Laval & Bert-Erboul 2005) and English sarcasm (Bryant & Fox Tree 2005), a decrease in mean F0 has been found in English sarcastic utterances (Attardo et al. 2003, Cheang & Pell 2008) and German ironic criticism (Scharrer et al. 2011). Similarly, regarding pitch variability, while F0 variability has been found to be higher in English and French sarcastic utterances (Attardo et al. 2003, Laval & Bert-Erboul 2005), a reduced F0 range has been reported for Cantonese sarcastic irony (Cheang & Pell 2009). Bryant (2011) suggests that these discrepancies between studies might be explained partly by potential crosslinguistic differences or by the fact that different studies have focused on different types of verbal irony.

In general, the results agree with previous studies in that there is no particular 'ironic tone of voice' that is specific to the marking of irony, and that speakers can indicate the presence of verbal irony by combining and contrasting a variety of prosodic modulations that are not special to verbal irony (Attardo et al. 2003, Bryant 2010, 2011). In normal conversation, speakers are inclined to use a varied set of prosodic modulations which will help listeners to infer irony by detecting certain 'incongruence' between the coded meaning and the attitude (i.e. the 'actual intention') of the speaker. The complex nature of the phenomenon seems to indicate that speakers can signal the presence of verbal irony by combining and contrasting a variety of prosodic marks, this is, "because of the inextricable relations between intentions and emotional tones of voice", prosodic signals specifically employed to highlight (i.e. to make 'relevant') an ironic remark can overlap with the affective prosody embedded in the ironic utterance (Bryant 2010:546).

Verbal irony can also be signalled with speech-accompanying gestures which can be produced both during and after ironic speech (e.g. ironic winks, facial expressions involving specific eye and eyebrow configurations, laughter and smiles, etc.; Caucci & Kreuz 2012, Attardo et al. 2011). To our knowledge, this is the first gestural study of the spontaneous use of gestures during ironic speech. Our results have revealed that speakers produce ironic utterances with higher rates of facial expressions, smiles, laughter and/or gaze changes towards the interlocutor, both during and after ironic utterances. Specifically, results show that occurrences of 'Smile/Laughter', 'Eyes', 'Eyebrows', 'Mouth', 'Head' and 'Gestures' are more frequent in ironic target than in baseline conditions, both during utterance pronunciation and their codas. Social-communicative function cues like 'laughter' and 'smile' (jointly considered) have been found to systematically appear in ironic utterances, not only during the production of the actual utterance (84% ironic target vs. 51%

baseline), but also during post-utterance codas (51% ironic target vs. 21% baseline), which is consistent with the experimental results obtained by Bryant (2011), Eisterhold et al. (2006) and Caucci and Kreuz (2012), who claim that laughter is a meta-communicative cue often used as a signal of ironic intent. Regarding gaze behaviour, the results show that speakers deviated their gaze significantly more often when producing ironic utterances (44%) than in baseline utterances (14%). In the case of codas, ironic codas tended to display gaze directed at the interlocutor. Interestingly, gaze changes seem to have two different functions in this context. While deviated gaze with no specific destination (i.e. fleeting deviations) seems to mark the ironic intent of the speaker (which is consistent with Williams et al.'s 2009 study), gaze directed at the interlocutor seems to have the function of checking the interlocutor's understanding of the ironic intent. This result highlights the fact that gaze features should be regarded as important informative cues in spoken interaction and deserve to be studied in greater depth in the context of the comprehension of irony (Argyle & Cook 1976, Gale & Monk 2000, Griffin 2001, Glenberg et al. 1998),

In sum, results from Experiment 1 show evidence that, in conversational contexts, speakers show the need to provide a good amount of prosodic and gestural information (including gaze) to indicate their ironic intent. In the corpus, ironic utterances were marked by 8.63 prosodic and visual cues on average (vs. 4.48 in baseline utterances), regardless of the lexico-syntactic cues and the pragmatic strategy (i.e. the irony subtype) employed by the speaker. This concentration of prosodic and gestural marks was consistent across ironic utterances, and showed higher rates of occurrence than lexico-syntactic marking. As pointed out above, an interesting result of Experiment 1 is that 70% of the ironic utterances were followed by what we called a "gestural coda" (as opposed to 29% of baseline utterances). Experiment 2 tested the relevance of these gestural codas through an irony detection task.

The results of Experiment 2 showed that, in absence of contextual cues, the presence of explicit codas (codas that are fulfilled with ironic facial expressions and/or gaze patterns) helped listeners to significantly increase their irony ratings, both for ironic and non-ironic utterances. While the increased ratings for ironic utterances were entirely expected, the increased ratings for non-ironic utterances surprising, given that 5 of the 12 non-ironic utterances used in Experiment 2 did not contain gestural cues in their codas with the exception of sustained gaze. This result is consistent with Attardo's et al. (2003) study, in which a sustained gaze directed towards the interlocutor showing no specific emotion (what they called 'blank face') was described as the most common cue to irony in their corpus. The unexpected results of the perception of non-ironic sentences in the coda as more ironic demonstrate that utterance codas filled with sustained gaze trigger the listeners' inferential processes. This finding agrees with Argyle and Cook (1976), Stivers et al. (2010) and Rossano (2010), who claim that the primary function of the eyes is to gather sensory input, especially when feedback —often smiling or laughter— is expected (see Argyle & Cook 1976, Vilhjalmsson 1997, Rossano 2010, Cosnier's 1991, Stivers et al. 2010). All these studies agree in considering eye gaze directed at the interlocutor one of the most important gestural signals characterizing the search for information and general response from the interlocutor. In general, the results show that the presence of gestural codas produced after speech utterances constitute an important cue that favours the interpretation of irony, regardless of whether they are produced with a smile, laughter, head or eyebrow movements or simply with sustained gaze directed at the listener.

In sum, from an audiovisual perspective, the findings presented in this study suggest that various verbal and non-verbal (i.e. prosody and gesture) components of communicative acts are important in the production and detection of ironic intent. These results agree with recent work on the relevance of prosodic and gestural patterns in the detection of prosodic meaning (e.g. Goldin-Meadow 2003, Krahmer & Swerts 2004, Swerts & Krahmer 2005, Borràs-Comes et al. 2011, Prieto et al. 2011). In the case of ironic speech, both prosodic and gestural markers are presumably used in order to reduce the processing effort of the interlocutor until the speaker ensures that the ironic understanding process has been completed, as House (1990, 2006), Clark and Lyndsey (1990), Fretheim (2002), Wilson and Wharton (2006) and Escandell-Vidal (1998, 2011a, 2011b) have proposed for prosody within Relevance Theory. Our results agree with the claims of Relevance Theory regarding verbal

irony: given the existing gap between the content of the utterance and its final interpretation in ironic contexts (and the extra cognitive effort required on the part of speakers and listeners), conversational participants use act-accompanying features such as prosody and gesture (and especially gestural codas), in order to help the interlocutor to achieve the ironic interpretation. Thus, we conclude that the presence of prosodic and gestural codas help in guiding the hearer in the interpretation of an utterance by means of providing overt clues about the assumptions and attitudes held by the speaker (or, in relevance-theoretic terms, for identifying high-level explicatures). We thus suggest that the results of both experiments constitute empirical evidence for the extension of Wilson and Wharton (2006) and Escandell-Vidal (2011a, 2011b)'s claims on the role of prosody, namely, that both prosody and gesture can act as active procedural instructions for pragmatic inferencing.

Aknowledgements

Parts of this paper were presented at UR-Ling Workshop 2014 (Barcelona 13 May 2014) and From Sound to Gesture (S2G) Conference (Padova 21-23 May 2014). We are grateful to the audience at these conferences for their helpful comments and discussion. We are also grateful to Alfonso Igualada, Joan Borràs-Comes and Núria Esteve-Gibert, the three independent coders that participated in the interreliability task. This research has been funded by two research grants awarded by the Spanish Ministry of Economy and Competitiveness (BFU2012-31995 "Gestures, prosody and linguistic structure" and FFI2012-31785 "Semántica procedimental y contenido explícito III"), and by a grant awarded by the Generalitat de Catalunya (2014 SGR-925) to the Prosodic Studies Group. Also, the research work of the first author is funded by the FPU2012/05893 grant (awarded by the Spanish Ministry of Economy and Competitiveness).

References

Allwood, Jens, Cerrato, Loredana, Dybjær, Laila, Jokinen, Kristiina, Navaretta, Constanza, Paggio, Patrizia, 2005. "The MUMIN Multimodal Coding Scheme". NorFA Yearbook. Available at http://cst.dk/mumin/workshop200406/MUMIN-coding-scheme-v1.3.doc. Retrieved May 13th 2014.

Amenta, Simona, Balconi, Michela, 2008. Understanding irony: An ERP analysis on the elaboration of acoustic ironic statements. Neuropsychological Trends 3, 7-27.

Anolli, Luigi, Ciceri, Rita, Infantino, Maria Giaele, 2002. "From 'blame by praise' to 'praise by blame': Analysis of vocal patterns in ironic communication". International Journal of Psychology 37, 266–276.

Argyle, Michael, Cook, Mark, 1976. Gaze and mutual gaze. Cambridge: Cambridge University Press.

Attardo, Salvatore, Eisterhold, Jodi, Hay, Jennifer, Poggi, Isabella, 2003. Multimodal markers of irony and sarcasm. International Journal of Humor Research 16, 243-260.

Attardo, S., Pickering L., Baker A. 2011. Prosodic and multimodal markers of humor in conversation. Pragmatics & Cognition, 19(2), 224-247.

Attardo, Salvatore, Wagner, Manuela Maria, Urios-Aparisi, Eduardo. 2013. Prosody and Humour. In: Attardo, Salvatore, Manuela Maria Wagner and Eduardo Urios-Aparisi (eds.), Prosody and Humor, John Benjamins, 189-201.

Blum-Kulka, Shoshana, House, Juliane, Kasper, Gabriele, 1989. Investigating cross-cultural pragmatics: An introductory overview. In Blum-Kulka, S., House, J., Kasper, G. (eds.), Cross-cultural pragmatics: Requests and apologies, Ablex, Norwood (NJ), 1-34.

Boersma, Paul, Weenink, David, 2008. Praat: doing phonetics by computer. Version 5.3.23. Computer program. Available at http://www.praat.org/. Retrieved July 28th 2012.

Borràs-Comes, Joan, Roseano, Paolo, Vanrell, Maria del Mar, Chen, Aoju, Prieto, Pilar, 2011. Perceiving uncertainty: facial gestures, intonation, and lexical choice. In: Proceedings of GESPIN 2011, Bielefeld (Germany).

Billmyer, Kristine, Varghese, Marika, 2000. Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. Applied Linguistics, 21(4), 517-552.

Bryant, Gregory A., Fox Tree, Jean E., 2005. Is there an ironic tone of voice? Lang Speech 48 (3), 257-277.

Bryant, Gregory A., 2010. Prosodic contrasts in ironic speech. Discourse Processes 47 (7), 545-566.

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

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

Caucci, Gina M., Kreuz, Roger J., 2012. Social and paralinguistic cues to sarcasm. Humor: International Journal of Humor Research, 25, 1-22.

Cartmill, Erica A., Ece Demir, Özlem, Goldin-Meadow Susan, 2012. Studying gesture. Research Methods in Child Language: A Practical Guide. Blackwell Publishing Ltd.

Cheang, Henry S., Pell, Mark D., 2008. The sound of sarcasm. Speech Communication 50, 366-381.

Cheang, Henry S., Pell, Mark D., 2009. Acoustic markers of sarcasm in Cantonese and English. Journal of the Acoustical Society of America 126 (3), 1394–1405.

Clark, Herbert H., Gerrig, Richard J., 1984. On the pretense theory of irony. Journal of Experimental Psychology: General 113 (1), 121-126.

Clark, Billy, Lyndsey, Geoff, 1990. Intonation, grammar and utterance interpretation. UCL Working Papers in Linguistics 2, 32-51.

Cosnier, Jacques, 1991. Les gestes de la question. In C. Kerbrat-Orecchioni (ed.), La question, Presses Universitaires de Lyon, Lyon, 163–171.

Curcó, Carmen, 1995. Some observations on the pragmatics of humorous interpretations: a relevance theoretic approach. Working Papers in Linguistics 7, 27-47.

Curcó, Carmen, 2000. Irony: Negation, echo and metarepresentation. Lingua 110, 257-280.

Eisterhold, Jodi, Attardo, Salvatore, Boxer, Diana, 2006. Reactions to irony in discourse: Evidence for the Least Disruption Principle. Journal of Pragmatics 38 (8), 1239-1256.

Escandell-Vidal, Maria Victoria, Leonetti, Manuel, 2014. Fronting and Irony in Spanish. In A. Dufter & A. Octavio de Toledo (eds.), Left Sentence Peripheries in Spanish: Diachronic, Variationist and Typological Perspectives, vol. viii. John Benjamins, Amsterdam, 309–342.

Escandell-Vidal, Maria Victoria, 2011a. Prosodia y pragmática. Studies in Hispanic and Lusophone Linguistics 4 (1), 193-208.

Escandell-Vidal, Maria Victoria. 2011b. *Verum focus* y prosodia: cuando la duración (sí que) importa. Oralia 14, 181-202.

Escandell-Vidal, Maria Victoria, 1998. Intonation and procedural encoding: the case of Spanish interrogatives. In Rouchota, Villy, Jucker & Andreas (Eds.), Current Issues in Relevance Theory. John Benjamins, Amsterdam, 169-203.

Escudero, David, Aguilar, Lourdes, Vanrell, Maria del Mar, Prieto, Pilar, 2012. Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labelling system, *Speech Communication* 54(4), 566-582.

Félix-Brasdefer, Julio C., 2010. Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor and E. Usó-Juan (eds.), Speech act performance: Theoretical, Empirical, and methodological issues, John Benjamins, Amsterdam/Philadelphia, 41-56.

Fleiss, Joseph L., 1981. Statistical methods for rates and proportions. John Wiley, New York.

Fretheim, Thorstein, 2002. Intonation as a constraint on inferential processing. Presented in Speech Prosody International Conference, Aix-en-Provence, France. Available at http://www.lpl.univ-aix.fr/sp2002/pdf/fretheim.pdf. Retrieved September 2nd 2012.

Gale, Caroline, Monk, Andrew, 2000. Where am I looking? The accuracy of video-mediated gaze awareness. Perception & Psychophysics 62, 586–595.

Gibbs, Raymond W., 1994. The Poetics of Mind: Figurative thought and Figurative Language. Academic Press, San Diego.

Gibbs, Raymond W., 2000. Irony in talk among friends. Metaphor and Symbol 15, 5-27.

Gibbs, Raymond W., 2012. Are ironic acts deliberate? Journal of Pragmatics 44, 104-115.

Glenberg, Arthur M., Schroeder, Jennifer L., Robertson, David A., 1998. Averting gaze disengages the environment and facilitates remembering. Memory & Cognition 26, 651–658.

Goldin-Meadow, Susan, 2003. Hearing Gesture: How our hands help us think. Harvard University Press, Cambridge.

Grassmann, Susanne, Tomasello, Michael, 2009. Young children follow pointing over words in interpreting acts of reference. Developmental Science 13, 252-263.

Grice, Paul H., 1975, Logic and Conversation. In Grice, Paul, 1989, Studies in the Way of Words, Harvard University Press, 22-40.

Griffin, Zenzi M., 2001. Gaze durations during speech reflect word selection and phonological encoding. Cognition 82, 1–14.

Hancock, Jeffrey T., 2004. Ironic use in face-to-face and computer-mediated conversation. Journal of Language and Social Psychology, 23, 447– 463.

Holler, Judith, Wilkin, Katie, 2009. Communicating common ground: how mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task. Language and Cognitive Processes 24, 267-289.

House, Jill, 1990. Intonation structures and pragmatic interpretation. In Ramsaran, S. (ed.), Studies in the pronunciation of English. Routledge, London, pp. 38-57.

House, Jill, 2006. Constructing a context with intonation. Journal of Pragmatics 38, 1542-1558.

IBM Corp. Released, 2011. IBM SPSS Statistics for Windows, Version 20.0. Computer program. IBM Corp, Armonk, New York.

Iverson, Jana M., Goldin-Meadow, Susan, 2005. Gesture paves the way for language development. Psychological Science 16 (5), 367-371.

Krahmer, Emiel, Swerts, Marc, 2004. More about brows. In Z. Ruttkay & C. Pelachaud (eds.). From brows to trust: Evaluating embodied conversational agents, Kluwer Academic Press, Dordrecht, pp. 191-216.

Kumon-Nakamura, Sachi, Glucksberg, Sam, Brown, Mary, 1995. How about another piece of pie: The allusional pretense theory of discourse irony. Journal of Experimental Psychology 124, 3-21.

Lausberg, Hedda, Sloetjes, Han, 2009. Coding gestural behavior with the NEUROGES-ELAN system. Behavior research methods, Instruments, & Computers 41 (3), 841-849. Computer program. Software available at http://www.lat-mpi.eu/tools/elan/. Retrieved July 2nd 2012.

Laval, Virginie, Bert-Erboul, Alain, 2005. French-speaking children's understanding of sarcasm: The role of intonation and context. Journal of Speech, Language, & Hearing Research 48, 610–620.

McNeill, David, 1992. Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago.

Morreall, John, 1989. Enjoying Incongruity. Humor 2 (1), 1-18.

Muñoa Barredo, Inma, 1997. Pragmatic Functions of Code-Switching among Basque-Spanish Bilinguals. University of Urbana-Champaign, Chicago. Available at http://webs.uvigo.es/ssl/actas1997/04/Munhoa.pdf. Retrieved April 28th, 2012.

Nakassis, Constantine y Jesse Snedeker. 2002. "Beyond Sarcasm: Intonation and Context as Relational Cues in Childrens's Reconigtion of Irony". Procedings of the Tweenty-sixth Boston University Conference on Language Development. A. Greenhyll, M. Hughs, H. Little Field, y H. Walsh (eds.). Cascadilla Press, Summerville, MA.

Nonhebel, Annika, Crasborn, Onno, van der Kooij, Els, 2004. Sign language transcription conventions for the ECHO project. Available at http://www.let.kun.nl/sign-lang/echo/docs/transcr_conv.pdf. Retrieved October 25th, 2012.

Padilla, Xose, (2012). ¿Existen rasgos prosódicos objetivos en los enunciados irónicos? Oralia 14, 203-224.

Potts, Christopher, 2005. Lexicalized intonational meaning. In Shigeto Kawahara (ed.), *University of Massachusetts Occasional Papers 30* (UMOP 30), 129-146. Amherst, MA: GLSA.

Prieto, Pilar, Borràs-Comes, Joan, Tubau, Susagna, Espinal, Teresa, 2013. Prosody and gesture constrain the interpretation of double negation. Lingua 131, 136-150.

Prieto, Pilar, 2014. The Intonational Phonology of Catalan. In: Sun-Ah Jun (ed.), Prosodic Typology 2. Oxford University Press, Oxford, 43-80.

Prieto, Pilar, Pugliesi, Cecilia, Borràs-Comes, Joan, Arroyo Ernesto, Blat, Josep, 2011. Crossmodal prosodic and gestural contribution to the perception of contrastive focus. In: Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011), Florence.

Randolph, Justus J., 2008. Online Kappa Calculator. Online computer program. Available at http://justus.randolph.name/kappa. Retrieved September 26th, 2012.

Rossano, F. 2010. Questioning and responding in Italian. Journal of Pragmatics, 42(10), 2756–2771.

Ruiz Gurillo, Leonor, 2008. Las metarrepresentaciones en el español hablado. Spanish in Context 5 (1), 40-63.

Searle, John. 1979. Expression and meaning: Studies in the theory of speech acts. Cambridge, England: Cambridge University.

Scharrer, Lisa, Christmann, Ursula, Knoll, Monja, 2011. Voice Modulations in German Ironic Speech. Language and Speech 54 (4), 435-465.

Smoski, Moria J., Bachorowski, Jo-Anne, 2003. Antiphonal laughter between friends and strangers. Cognition & Emotion 17, 327-340.

Sperber, Dan, Wilson, Deirdre, 1986/1995. Relevance: Communication and Cognition. Blackwell, Oxford. Sperber, Dan, Wilson, Deirdre, 1993. Linguistic form and relevance. Lingua 90, 1-25.

Stivers, T., & Rossano, F. 2010. Mobilizing response. Research on Language and Social Interaction, 43(1), 1–31.

Swerts, Marc, Krahmer, Emiel, 2005. Audiovisual prosody and feeling of knowing. Journal of Memory and Language 53, 81-94.

Vanek, Christian, McDaniel, Scott. 2006. SurveyGizmo. Online Survey Software. Available from: http://www.surveygizmo.com/. Retrieved 14 December 2012.

Van Lancker, Diana, Canter, Gerald J., Terbeek, Dale, 1981. Disambiguation of ditropic sentences: acoustic and phonetic cues. Journal of Speech and Hearing Research 24, 64-69.

Vilhjalmsson, Hannes H., 1997. Autonomous communicative behaviors in avatars. Unpublished Master's thesis. Massachusetts Institute of Technology, Cambridge, MA.

Williams, Jason A., Burns, Erin L., Harmon, Elisabeth A., 2009. Insincere utterances and gaze: Eye contact during sarcastic statements. Perceptual and Motor Skills 108 (2), 565-572.

Wilson, Deirdre, 2000. Metarepresentation in linguistic communication. In: Sperber, D. (ed.), Metarepresentations. A multidisciplinary perspective. Oxford University Press, New York, pp.411–448.

Wilson, Deirdre, Wharton, Tim, 2006. Relevance and prosody. Journal of Pragmatics 38, 1557-1579.

Wilson, Deirdre, Sperber, Dan, 2012. Explaining Irony. In: Wilson, D., Sperber, D. (eds.), Meaning and Relevance. Cambridge University Press, Cambridge, pp. 123-145.