

AUDIOVISUAL PROSODY AND VERBAL IRONY

Santiago González Fuente

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI
Dra. Pilar Prieto Vives

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL
LLENGUATGE



A mis padres,

— *Why did you name the movie Bananas?*
— *Because there are no bananas in it.*

Reporter and Woody Allen

Necesito ovejas, ¿tienes ovejas?
Te cambio una arcilla por una oveja.

Ernesto Rodríguez Pérez

I'd rather be dead than singing
'Satisfaction' when I'm forty-five.

Mick Jagger

Todos los bades están allí.
Bares.

Pedro Lomo

¡Aquí!

Bruno Zaragoza Arias

Acknowledgments

I am indebted to many people, in one way or another, for their guidance, help and support during the last four—almost five—years. I apologize if I have forgotten to mention someone in this section. In any case, mentioned or not, no words would be sufficient to express my gratitude to all the people that have contributed to making this PhD dissertation possible.

First of all, I owe my deepest gratitude to my supervisor Pilar Prieto. I am very grateful to you, Pilar. First, for giving me the opportunity to be a member of the GrEP Research Group; it has been an amazing and very enriching experience for me to be part of GrEP, both from the academic and also the personal point of view. I have learned from you so many things related not only to academics, but also about collaborative working, and personal self-improvement. Second, I want to thank you for your patience, comprehension, and support in overcoming numerous obstacles I have faced through my PhD. Even in the most difficult moments, your always positive, careful and encouraging attitude is one of the main reasons why I have achieved this goal. So many thanks for everything, Pilar.

Secondly, I would like to thank the members of my dissertation committee, Professors Ana M^a Fernández Planas, Xose A. Padilla García and Isabella Poggi for accepting to review this thesis and for

taking the time out of their busy schedules to offer their insights for this work.

I also want to express my special gratitude to Ira Noveck for hosting and guiding my research at the *Institut des Sciences Cognitives* of Lyon, as well as to his family, Monica, Noemí, Isaac—and Cookie—for being so kind with me during my stay in Lyon. Thank you, Ira. I will remember those three months as one of the high points of my PhD. Many thanks also to all the members of the *Institut des Sciences Cognitives* for make my stay in Lyon as productive as it was happy. Thank you Thomas Castelain, Ludivine Dupuy, Quentin Moreau, Gustavo Estivalet, Agustin Richard, Emmanuel Trouche, Maël Garnotel, Romain Mathieu, Auriane Coderc, Robert Reinecke, Anne Reboul, Anne Cheylus, Arthur Lefevre, Miguel Pedroza, Jonathan Faure, Jeremy Yeaton, and, especially, many thanks to Yang Hu for your kindness and friendship.

I also would like to express my gratitude to the researchers that collaborated in the papers that are part of this dissertation, namely Victoria Escandell-Vidal, from the Universidad Nacional de Educación a Distancia, and Patrick Zabalbeascoa, from the Universitat Pompeu Fabra. It has been a pleasure to work with you. Also many thanks to M^a Teresa Espinal, Susagna Tubau and Feifei Li, from the Universitat Autònoma de Barcelona, with whom I collaborated in different projects during these five years. I have learned a lot working with you, many thanks for everything.

Also want to thank all the members of the *Grup d'Estudis de Prosòdia*, namely Meghan Armstrong, Florence Baills, Marco Barone, Joan Borràs, Núria Esteve, Martina Garufi, Iris Hübscher, Alfonso Igualada, Evi Kiagia, Olga Kushch, Judith Llanes, Paolo Roseano, Cristina Sánchez, Rafèu Sichel, Jill Thorson, Maria del Mar Vanrell, and Ingrid Vilà. And also many thanks to my other PhD colleagues at Universitat Pompeu Fabra: Celia Alba, Toni Bassaganyas, Robert Bailey, Sara Cañas, Marta García, Mihajlo Ignjatovic, Alexandra Navarrete, Alba Milà, Aina Obis, Veronika Richtarcikova, Alexandra Spalek, Lieke van Maastricht, Eugenio Vigo, Blanca Arias, Tamara Vorovyeba, Kata Wohlmuth, Chenjie Yuan, and Giorgia Zorzi. Thank you for everything, guys, you made my time at the Universitat Pompeu Fabra great, and I will miss you.

I'm also indebted to many people in the Department of Translation and Language Sciences at the Universitat Pompeu Fabra. Many thanks to Sergi Torner, Toni Badia, and Àlex Alsina for their support and comprehension and—especially—during the final stages of the PhD. Moreover, I must also express my gratitude to the members of the Secretary of the Translation and Language Sciences Department at the UPF: Susi Bolós, Yolanda Bejarano, Yolanda Vicente, Núria Abad, Rafa Ordóñez, Nayat Chourak, and Eulàlia Palet. Thank your for your support and patient solving all kinds of bureaucratic messes, I really appreciate it.

I also owe a great debt of gratitude to other professors and colleagues from the Universitat Pompeu Fabra from whom I have had the opportunity to learn and to work with. Many thanks to Carmen López, Irene Renau, Sergi Torner, Laura Borràs, Àlex Alsina, Josep M^a Fontana, Lourdes Díaz, Clara Lorda, Andrés Chandía, Mertixell Uriel, María Sanz, Sara Silvente, Juan María Garrido, Laura Acosta, Carmen González, Marta Molina, Carmen Pérez, Aurora Bel, Louise McNally, Gemma Barberà, and Laia Mayol.

This Ph.D. was funded by an FPU grant (FPU2012-05893) and by a project (BFU2012-31995), both awarded from the Spanish MINECO, the first to my self and the latter to Pilar Prieto. Many thanks to all the participants in the experiments, both people that participated in the elaboration and creation of the experimental materials—especially to Celia Alba, Esther Arias Valor, Marina Blasco, Joan Borràs-Comes, Oriol Borrega, Núria Esteve-Gibert, Lindes Farré González, Bernat Grau Dilla, Patricia Mallén, Alba Milà, Maria Porta Serramià, Ernesto Rodríguez Pérez, Paolo Roseano, Cristina Sobreviela, Marina Vilà, and Rita Zaragoza—and also to those which actively participated in the production and perception experiments. Thank you very much to the staff of the Escola Sant Martí in Arenys de Munt, Institució “La Miranda” in Sant Just Desvern, Escola La Farigola del Clot in Barcelona, and Escola Pública Dr. Estalella Graells in Vilafranca del Penedès for granting me access to the meetings with children. Also many thanks

to the children and their families for their participation in the experiments.

I am also in debt with the organizers and professors of the Màster de Estudios Fònics from CSIC-UIMP, Marianela Fernández, Patricia Infante, José María Lahoz, Joaquim Llisterri, Maria Machuca, Eugenia San Segundo, and especially to the director, Juana Gil Fernández. Doing this master's degree was one of the greatest educational experiences of my life. I will always be in debt with you.

Also many thanks to Mariona Taulé, Montse Nofre, Santi Reig, Glòria Valdívia, Marta Vila, and especially to M^a Antonia Martí, from the Centre de Computació i Llenguatge at the Universitat de Barcelona, for giving me the opportunity to work in the field of linguistics for the first time in my life.

I also feel special gratitude towards the members of the *Laboratori de Fonètica* at the Universitat de Barcelona, Eugenio Celdrán, Ana M^a Fernández Planas, Ramon Cerdà, Paolo Roseano, Wendy Elvira García, Lourdes Romera, Josefina Carrera, and Valeria Salcioli. It was by studying Linguistics at the Universitat de Barcelona when I took my first steps in the field of phonetics, so I feel the Laboratori de Fonètica and their members are an important part of me and my career. Many thanks for that. At this point I would like to give special thanks to Dr. Elvira García for creating

and generously sharing the Praat script that I used to create some of the pictures—the nicest ones, in fact—of this thesis.

Finally, I want to thank my friends for their eternal support and patience. You know who you are. I love you, and I want to apologize if these last five years I have not been as present in your lives as I would have like. I was working on irony, you know. Still am.

My last words of thanks are for my family. There are no sufficient words of gratitude for my parents, Josefina and Augusto, who have so generously worked hard for their family and from whom I have learned everything important I know. Thank you for your constant unconditional support, you are always an example for me. I want also to express my gratitude to my brothers, cousins, aunts, and uncles and their wives, husbands, sisters and sons. I love you all. Thank you for your support and for putting up with me.

And last, but not least, I want to thank Marta for her constant and lovely support during these last years. I admire you and I love you. You have stayed strong and given me strength too. Thank you for waiting for me—I'll be coming home soon.

Abstract

This dissertation takes an integrated approach to the study of audiovisual cues to verbal irony. While pragmatic studies have mainly focused on the role of the discourse context in irony detection, little is known about the role of prosodic and gestural cues in this process.

The thesis includes four experimental studies—each one described in a separate chapter—addressing a set of questions using a variety of experimental designs. The first one is a case study of a professional comedian and reveals (a) that ironic utterances display a higher density of prosodic and gestural markers than non-ironic utterances; and (b) that gestural markers can appear both temporally aligned with prosodic prominence but can also appear independently, as gestural codas. The second study includes two experiments: (a) a production experiment eliciting spontaneous ironic speech which reveals that in non-professional spontaneous speech, too, speakers employ a higher density of prosodic and gestural markers in ironic compared to non-ironic utterances; and (b) a perception experiment on the contribution of gestural codas to the detection of verbal irony, which shows that speakers detect ironic intent significantly better when post-utterance gestural codas are present than when they are not. Following up on this idea, the third study presents three perception experiments on the relative contribution of contextual vs. prosodic vs. gestural cues to verbal irony understanding. Overall, results of the three experiments

emphasize the role of contrast effects in irony perception. The first experiment shows that (a) listeners detect irony more accurately when they have access to both prosodic and gestural cues than when they just rely on prosodic information, (b) that listeners rely more strongly on gestural information than on prosodic information, and (c) that listeners rely more heavily on gestural cues than on prosodic or contextual ones for detecting irony. Finally, the fourth study addresses the contribution of prosodic and gestural cues to children's early understanding of verbal irony, showing that mismatched multimodal cues of emotion facilitate the detection of irony by 5-year-old children.

Altogether, this dissertation shows that both prosodic and gestural markers of irony aid in guiding the hearer in the interpretation of an utterance by providing overt clues about the assumptions, emotions and attitudes held by the speaker. Together with recent studies on the general pragmatic effects of prosody and gesture, the claim is that audiovisual markers of irony are strong triggers of implicature strength which help decode speech intentions in interaction. In addition, the dissertation presents novel empirical evidence of the stronger effects of multimodal—and especially gestural—cues in comparison with contextual cues, both in adult and child populations. This crucial finding leads us to claim that the study of prosodic and gestural cues to verbal irony should be at the core of any pragmatic or psycholinguistic account of verbal irony production and comprehension.

Resumen

Esta tesis aborda el estudio de las marcas audiovisuales de la ironía verbal desde una perspectiva integral. Los estudios pragmáticos se han centrado principalmente en investigar el papel del contexto discursivo en la detección de la ironía, pero poco se sabe sobre el rol que desempeñan las marcas prosódicas y gestuales en este proceso.

La presente tesis incluye cuatro estudios experimentales —cada uno de ellos incluido en un capítulo separado— que abordan diferentes preguntas de investigación utilizando varios diseños experimentales. El primero es un estudio de caso sobre un cómico profesional y muestra, en primer lugar, que los enunciados irónicos presentan una mayor densidad de marcadores prosódicos y gestuales que los enunciados no irónicos y, segundo, que los marcadores gestuales pueden aparecer alineados temporalmente con la prominencia prosódica, pero también de forma independiente, como codas gestuales. El segundo estudio incluye dos experimentos. Uno de producción —diseñado para obtener discurso irónico espontáneo—, cuyos resultados confirman que también en habla espontánea los hablantes no profesionales emplean una mayor densidad de marcadores prosódicos y gestuales cuando son irónicos en comparación con cuando no lo son; y, en segundo lugar, un experimento de percepción que investiga la contribución de las codas gestuales a la detección de la ironía verbal y cuyos resultados muestran claramente cómo la intención irónica se detecta significativamente mejor cuando los hablantes tienen acceso a codas

gestuales que cuando no la tienen. El tercer estudio de esta tesis contiene tres experimentos de percepción que examinan cómo el contexto, las marcas prosódicas y las marcas gestuales contribuyen a la comprensión de la ironía verbal. En general, los resultados de los tres experimentos subrayan la importancia que los “efectos de contraste” tienen en el proceso de detección de la ironía. El primero muestra que los oyentes detectan la ironía con más precisión cuando tienen acceso a las marcas prosódicas y gestuales de manera conjunta que cuando solo tienen acceso a la información prosódica; el segundo, que la información visual resulta más convincente que la información prosódica a la hora de detectar la ironía; y, por último, el tercer experimento muestra que los oyentes emplean preferentemente las marcas gestuales por encima de las prosódicas e incluso de las contextuales a la hora de detectar la ironía. Finalmente, el cuarto estudio investiga cómo los niños desarrollan la capacidad de detectar la ironía verbal a través de las marcas prosódicas y gestuales, y los resultados muestran que las marcas multimodales facilitan la detección de la ironía en niños desde los 5 años de edad.

En conjunto, esta tesis muestra que tanto los marcadores prosódicos como los gestuales contribuyen de manera significativa a la comprensión de la ironía verbal, guiando al oyente en la interpretación del enunciado mediante el suministro de pistas sobre las suposiciones, las emociones y las actitudes del ironizador. Siguiendo la línea de algunos estudios recientes sobre los efectos pragmáticos de la prosodia y el gesto, los resultados de los experimentos de esta tesis muestran que los marcadores

audiovisuales de la ironía son potentes desencadenadores del proceso inferencial necesario para decodificar las intenciones del hablante en las interacciones comunicativas. Además, esta tesis presenta evidencia empírica de la gran incidencia que tienen las marcas multimodales —y especialmente de las gestuales— en la detección de la ironía verbal en comparación con las marcas contextuales, tanto en la población adulta como en la infantil. Este hallazgo fundamental nos lleva a afirmar que el estudio de las señales prosódicas y gestuales de la ironía debería considerarse una parte integral fundamental de cualquier explicación pragmática o psicolingüística sobre la producción y comprensión de la ironía verbal.

Resum

Aquesta tesi adopta una perspectiva integral a l'estudi de les marques audiovisuals en la ironia verbal. Els estudis pragmàtics s'han centrat principalment a investigar el paper del context discursiu en la detecció de la ironia, però se sap poc sobre el rol que juguen les marques prosòdiques i gestuals en aquest procés.

La tesi inclou quatre estudis experimentals —cadascun d'ells descrit en un capítol separat— que aborden diferents preguntes de recerca i fan servir diversos dissenys experimentals. El primer és un estudi de cas sobre que analitza el discurs d'un còmic professional i mostra, en primer lloc, que els enunciats irònics presenten una major densitat de marcadors prosòdics i gestuals que els enunciats no irònics; i, en segon lloc, que els marcadors gestuals poden aparèixer temporalment alineats amb la prominència prosòdica, però també de forma independent, en el que anomenem “codes gestuals”. El segon estudi inclou dos experiments. Un de producció, dissenyat per obtenir discurs irònic espontani i els resultats del qual confirmen que en parla espontània els parlants no professionals també empenen una major densitat de marcadors prosòdics i gestuals quan són irònics en comparació a quan no ho són; i, en segon lloc, un experiment de percepció sobre la contribució de les codes gestuals a la detecció de la ironia verbal, els resultats del qual demostren que la ironia es detecta millor quan els parlants tenen accés a codes gestuals que quan no en tenen. El tercer estudi presenta tres experiments de percepció que examinen com el context, les marques prosòdiques i les marques gestuals contribueixen a la

comprensió de la ironia verbal. En general, els resultats dels tres experiments subratllen la importància dels efectes de contrast en el procés de detecció de la ironia. El primer experiment mostra que els oients detecten la ironia amb més precisió quan tenen accés a les marques prosòdiques i gestuals alhora, comparat amb quan només tenen accés a la informació prosòdica; el segon, que la informació visual és més poderosa que la informació prosòdica a l'hora de detectar la ironia, i, finalment, el tercer experiment mostra que els oients empren preferentment les marques gestuals per sobre de les prosòdiques o les contextuals a l'hora de detectar la ironia. Finalment, el quart estudi investiga com els infants aprenen a comprendre la ironia verbal a través de les marques prosòdiques i gestuals, i els resultats mostren que les marques multimodals faciliten la detecció de la ironia des dels 5 anys d'edat.

En conjunt, aquesta tesi mostra que tant els marcadors prosòdics com els gestuals contribueixen a la comprensió de la ironia verbal, tot guiant l'oient en la interpretació de l'enunciat mitjançant el subministrament de pistes sobre els supòsits, les emocions i les actituds del parlant irònic. Seguint la línia d'estudis recents sobre els efectes pragmàtics de la prosòdia i el gest, els resultats dels experiments d'aquesta tesi mostren que els marcadors audiovisuals de la ironia són potents factors que desencadenen el procés inferencial necessari per a descodificar les intencions del parlant en les interaccions comunicatives. A més, aquesta tesi presenta evidència empírica de la gran incidència que tenen les marques multimodals —i especialment de les gestuals— en la detecció de la ironia verbal, en comparació amb les marques contextuals, tant en

població adulta com en població infantil. Aquesta troballa fonamental reforça la idea que l'estudi dels aspectes prosòdics i gestuals hauria de ser una part integral fonamental de qualsevol explicació pragmàtica o psicolingüística sobre la producció i comprensió de la ironia verbal.

List of original publications

CHAPTER 2

González-Fuente S. (2016). “La prosodia audiovisual de la ironía verbal: un estudio de caso”. *Revista de la Sociedad Española de Lingüística* 45(1), pp. 77-104.

CHAPTER 3

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

CHAPTER 4

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

CHAPTER 5

González-Fuente S, and Prieto P. (submitted). “Mismatching prosodic and gestural cues of emotion facilitate the detection of verbal irony in children”. *Journal of Child Language*.

Table of contents

Acknowledgments	v
Abstract.....	xiii
List of original publications.....	xxv
1. INTRODUCTION	31
1.1. What is verbal irony?.....	31
1.1.1 What is verbal irony used for? Functions and classifications	33
1.1.2 How does irony work? The cognitive processing of verbal irony	36
1.2. Prior work	41
1.2.1. Prosodic markers of verbal irony.....	41
1.2.2. Gestural markers of verbal irony	45
1.2.3. The developmental perspective.....	47
1.3. Theoretical Framework.....	50
1.3.1. Multimodal communication: the audiovisual-prosody perspective	50
1.3.2. The Relevance Theory perspective.....	54
1.3.3. Contrast effects in verbal irony comprehension: the Contrast-Assimilation Theory.....	56
1.4. General objectives, research questions and hypotheses	60
2. CHAPTER 2: “La prosodia audiovisual de la ironía verbal: un estudio de caso”	65
2.1. Introducción.....	65

2.1.1. Lenguaje indirecto e ironía verbal	66
2.1.2. Los componentes prosódico y gestual en el estudio de la ironía verbal	68
2.1.3. La Teoría de la Relevancia y la perspectiva de la Prosodia Audiovisual	74
2.2. Metodología.....	79
2.2.1. Análisis cuantitativo	82
2.2.2. Análisis cualitativo	84
2.3. Resultados.....	85
2.3.1. Análisis cuantitativo	85
2.3.2. Análisis cualitativo	89
2.4. Discusión y conclusiones	102
3. CHAPTER 3: “Gestural codas pave the way to verbal irony understanding”	113
3.1. Introduction	113
3.2. Experiment 1	120
3.2.1. Methods	120
3.2.2. Results.....	134
3.3. Experiment 2	145
3.3.1. Methods	145
3.3.2. Results.....	156
3.4. Discussion and conclusions	157
4. CHAPTER 4: “Communicating irony: when gesture cues are more powerful than prosodic and contextual cues”.....	165

4.1. Introduction	165
4.2. Experiment 1	176
4.2.1. Methods	176
4.2.2. Results.....	186
4.3. Experiment 2	187
4.3.1. Methods	188
4.3.2. Results.....	189
4.4. Experiment 3	191
4.4.1. Methods	192
4.4.2. Results.....	194
4.5. Discussion and conclusions	198
 CHAPTER 5: “Mismatching prosodic and gestural cues of emotion facilitate the detection of verbal irony in children”	 205
5.1. Introduction	205
5.2. Methods	213
5.2.1. Preliminary study: Discourse Completion Task	213
5.2.2. Experimental materials	221
5.2.3. Participants.....	225
5.2.4. Procedure	225
5.3. Results	228
5.4. Discussion and Conclusions	232
 6. GENERAL DISCUSSION AND CONCLUSIONS	 237

6.1. Summary of findings	237
6.2. Is there an ironic tone of voice or an ironic gestural pattern?	240
6.3. Gestural cues can appear both aligned and misaligned with prosodic cues: the particularity of gestural codas.....	245
6.4. Prosodic and gestural contrasts signal ironic intent	248
6.5. Visual cues are stronger than prosodic and contextual cues for verbal irony detection in both adult and child populations	251
6.6. Multimodal cues facilitate early detection of irony in children	253
7. References	257
8. Appendices	281
Appendix A.....	281
Appendix B	285
Appendix C.....	287

1. INTRODUCTION

1.1. What is verbal irony?

This dissertation focuses on the role of prosody and gestural cues in verbal irony production, comprehension and development. Within the field of human communication, the phenomenon of irony has generated a vast amount of literature dedicated solely to its study. From classical times to the present, language philosophers, psycholinguists and pragmaticians have shown interest in this complex but common phenomenon whereby (in its most archetypal case) an individual chooses to say “You’re so brilliant, man!” when he/she actually means “What a clumsy oaf!”

Classical rhetorical approaches defined verbal irony as a figure of speech in which what is said is the opposite of what is meant, and for many centuries, the study of verbal irony was circumscribed to the study of the use of rhetorical devices in literature works. In the 1970’s, the Standard Pragmatic Model proposed by Grice (1975) overcame this conception by arguing that verbal irony is not only used in literature, but also in real-life language interaction, being a very common form of communication in everyday conversations. Grice’s Standard Pragmatic Model (1975) introduced the notions of “cooperative principle” (a general implicit assumption which speakers follow in conversations), “conversational maxims” (a set of assumptions underlying the cooperative principle), and “implicatures” (the inferences that speakers draw in conversation

when a conversational maxim is violated). With these new conceptual tools, Grice (1975) proposed an explanation of verbal irony which consists in “an intentional flouting of the maxim of quality” (i.e., “try to make your contribution one that is true”). This flouting of the maxim of quality is assumed to trigger semantic implicatures which are discrepant with the semantics of the sentence. From this perspective, producing verbal irony constitutes a social behavior which affects not only the semantics of the utterance—*what is said*—but also the psychological processes underlying the production and comprehension of the implied meaning—*what is meant*. However, whereas Grice’s account of verbal irony constitutes a step forward in the understanding of verbal irony, it still does not provide an account of two crucial aspects which post-Gricean cognitive accounts identified, namely *what irony is used for* (i.e., why a rational speaker would choose an utterance whose meaning is the opposite of the one he intends to communicate) and *how irony works* (i.e., the cognitive processes involved in verbal irony comprehension), and how they relate to the general architecture of cognition. In order to address these issues, in recent decades post-Gricean cognitive approaches to verbal irony (e.g., Allusional Pretense Theory, Kreuz & Glucksberg, 1989; Clark & Gerrig, 1984; Kumon-Nakamura, Glucksberg & Brown, 1995; Indirect Negation Theory, Giora, 1995; Relevance Theory, Sperber & Wilson, 1986/1995; and the General Theory of Verbal Humor, Attardo, 2000) have widened their focus of interest from the specific study of the semantics of the literal and implied meaning of the ironic utterances and have addressed the study of verbal irony

(a) by considering as a central part of their accounts other aspects involved in the irony communication process such as the attitudes, emotions and communicative goals held by speakers, and, (b) by experimentally testing the nature and effects of the cognitive processes underlying the processing of verbal irony. In the next sections I present a brief summary of the answers that pragmatic and psycholinguistic cognitive approaches have provided to account for these two main and crucial issues of ironic communication.

1.1.1. What is verbal irony used for? Functions and classifications

Some current pragmatic-psychological accounts have focused on answering a question such as *what irony is used for* (e.g., Hutcheon, 1996; Attardo, 2013; Sperber & Wilson, 1986/1995; Dews & Winner, 1995; Colston, 1999). So far many classifications of irony have been elaborated depending on the communicative goals and strategies employed by speakers to convey an ironic intent. It has been shown that by using verbal irony (i.e. by saying something with a meaning opposite to or discrepant with the actual intended meaning) speakers achieve certain communicative goals that warrant its use (Colston, 1999). Whereas the most studied type of verbal irony function has been the expression of some kind of negative or critical evaluation towards an event or an interlocutor (Sperber & Wilson, 1986/1995; Kumon-Nakamura et al., 1995; Cheang & Pell, 2008), it has also been shown that verbal irony can be used to achieve more positive functions (Ruiz Gurillo, 2008)

such as humor (Attardo, 2013), surprise (Colston & Keller, 1998), or politeness (Alvarado & Padilla, 2010). In fact, it has been proposed that speakers may use verbal irony to achieve specific (and overlapping) discourse goals. For example, Alvarado & Padilla (2010) analyzed a set of ironic utterances from a spontaneous speech corpus and found that speakers used apparently critical irony for positive functions such as, for example, strengthening friendship ties and integrating the speaker and addressee into the conversational group, thus showing that irony and politeness are perfectly compatible pragmatic mechanisms.

Due to the complexity of ironic communication, which involves a variety of communicative goals produced in a wide range of situations, the classification of verbal irony in different subtypes has been a controversial point in recent literature. As Gibbs (2000: 342) suggests, variation in the forms of irony presents “an important challenge for cognitive science theories of irony. Is it necessarily the case that a single theory will account for the multiple forms and functions of irony in ordinary speech?” He concludes that “irony is not a single category of figurative language, but includes a variety of types, each of which is motivated by slightly different cognitive, linguistic, and social factors, and conveys somewhat different pragmatic meanings” (Gibbs, 2000: 356). Gibbs (2000) proposed grouping the most representative forms of verbal irony in a classification based on the mixed criteria of the speaker’s communicative goal and the semantic strategy he/she uses to convey the intended meaning (see Table 1). This classification

contains five irony subtypes, namely *sarcasm*, in which the speaker utters a positive sentence to convey the criticism; *hyperbole*, in which the speaker exaggerates the reality of the situation; *understatement*, in which the speaker conveys an ironic meaning by stating far less than was obviously the case; *jocular*, in which the ironic speech is intended to tease or poke fun; and *rhetorical questions*, in which the speaker asks questions implying a critical or humorous intention.

Table 1. Examples of irony subtypes proposed by Gibbs (2000). (Context: *Mark and Peter are friends and are riding a bicycle. Mark crashes his bicycle into a tree. Then, Mark says...*).

Irony subtype	Example
Sarcasm	<i>“Good job!”</i>
Hyperbole	<i>“You are the most amazing bicyclist in the state of North Carolina!”</i>
Understatement	<i>“You are a little bit clumsy.”</i>
Jocular	<i>“I’ll race you to the end of the street!”</i>
Rhetorical question	<i>“Do you have something against trees?”</i>

From these verbal irony subtypes, probably the one most explored in the literature is sarcasm. It is important to point out that traditional work on irony has sometimes used the terms “irony” and

“sarcasm” in a complementary fashion (i.e. “irony” conveying a positive intent vs. “sarcasm” conveying a negative intent) (see Vengalien, 2005, for a complete review on this issue). At this point we want to clarify that, following Gibbs' (2000) classification, in this dissertation we will use the term “irony” as an hyperonym of all forms of ironic communication, including sarcasm, which constitutes a specific irony subtype that is characterized by conveying an explicit negative and critical attitude towards an event or a person, which is the most common conception of “sarcasm” in the literature (e.g. Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995; Gibbs, 2000; Cheang & Pell, 2008; Bryant, 2010).

1.1.2. How does irony work? The cognitive processing of verbal irony

Going a step beyond the Standard Pragmatic Model (Grice, 1975), 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. These pragmatic and psycholinguistic accounts agree with Grice in considering verbal irony as a form of indirect intentional language in which effectively there is some kind of incongruity between what is said (i.e., the propositional content of an utterance) and what is meant (i.e., the implied meaning of that utterance) (e.g., Curcó, 1995; Bryant, 2012). However, while the notion of ‘discrepancy’ or ‘incongruity’ is to some extent contained in all the pragmatic accounts of verbal irony proposed so far,

current cognitive approaches to verbal irony propose that, rather than detecting the contradiction between the semantics of the literal and the implied meanings of a utterance, the internal functioning mechanisms of verbal irony comprehension processes rely on the listeners' recognition of some kind of discrepancy between (1) expectations and reality (Gibbs, 2012); (2) actual and attributed attitudes/beliefs (Sperber & Wilson, 1986/1995); (3) real and imagined discourse acts (Clark & Gerrig, 1984; Kreuz & Glucksberg, 1989); (4) the relevance or inappropriateness of propositional content in a particular context (Attardo, 2000); (5) failed expectations and the attitudes towards those failed expectations (Kumon-Nakamura et al., 1995); or (6) what is negated and what was implicated (Giora, 1995) (see Bryant, 2012, for a review). To our knowledge, the only cognitive pragmatic theory that has explored the role of prosody and gesture in verbal irony comprehension is Relevance Theory (Sperber & Wilson, 1986/1995). In this dissertation we will discuss our results in the light of this cognitive-pragmatic theory, which will be presented in section 1.3.2 below.

Regarding the cognitive processing of ironic utterances, there are two main accounts of this issue, namely the Grade Salience Hypothesis (Giora, 1995) and the Direct Access Model (Gibbs, 1994). The main difference between the two lies in the temporal processing of the ironic statements. While the Grade Salience Hypothesis suggests that the literal interpretation of the statement is examined first, and then is considered in conjunction with the ironic

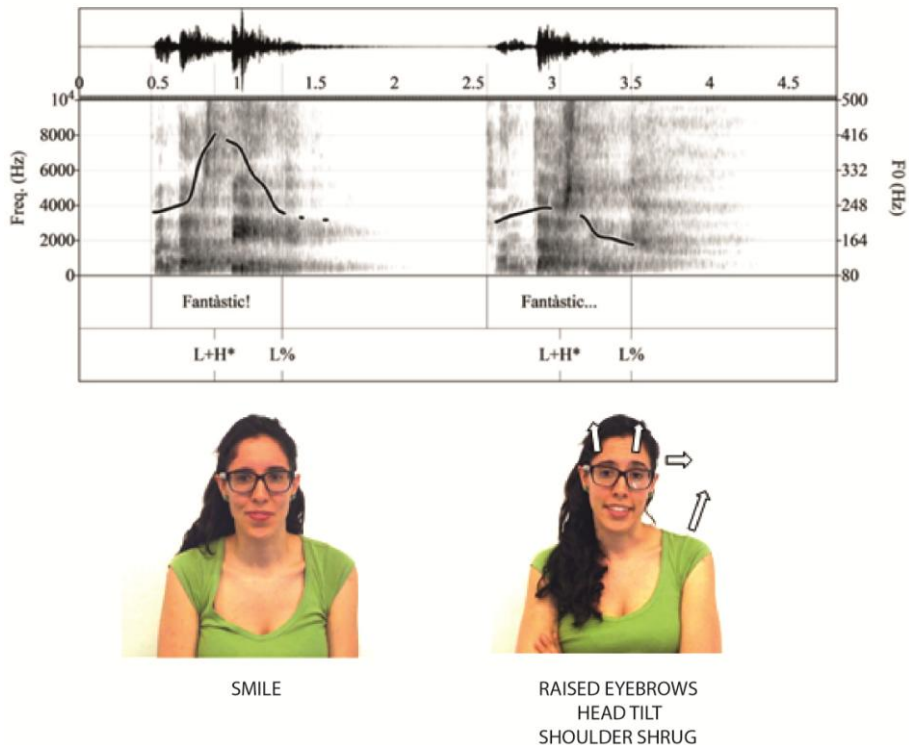
meaning, the Direct Access Model claims that irony can be processed without first activating and assessing the literal interpretation of the statement in appropriate contextual conditions. Interestingly, recent experimental research based on the Contrast and Assimilation Theory proposed by Colston (2002) claims that contrasting information affects the processing of utterances, demonstrates that the degree of contrast between verbal and contextual cues (Ivanko & Pexman, 2003) as well as contrasts between prosodic and contextual cues (Woodland & Voyer, 2011) affect the processing of verbal irony in terms of response time. The results of these experiments are consistent with the Direct Access Model (Gibbs, 1994), as they show that ironic utterances produced with contrasting contextual and prosodic cues can be processed as fast as sincere utterances. Whereas this dissertation does not specifically focus on the discussion of the most appropriate model for verbal irony processing (i.e., Grade Salience Hypothesis vs. Direct Access Model), the studies presented in Chapters 4 and 5 will follow the research line proposed by Colston (2002) by testing the effects that contrasts between multimodal cues (i.e. prosodic and gestural) and contextual cues have in the detection of speaker's ironic intents. A more detailed explanation of the Contrast and Assimilation Theory (Colston, 2002) and the experimental research carried out within this line of research is presented in section 1.3.3 below.

While the abovementioned accounts of verbal irony have emphasized the role of discourse context on verbal irony

comprehension processes, less is known about the role of prosodic and much less of gestural cues in verbal irony detection and comprehension. Even though most approaches to verbal irony agree in considering that understanding an ironic remark involves the evaluation of different cues coming from different sources, namely verbal (i.e. the propositional content of the utterance), contextual (e.g. the specific situation, shared beliefs, communicative goals) and multimodal (i.e. prosodic and gestural) (e.g. Gibbs, 1994; Attardo et al., 2003; Bryant, 2011, 2012), the role of prosodic (and even more so gestural) cues in verbal irony production and comprehension processes has not been fully integrated into most theories.

Research has shown that the expression of irony is a multimodal affair and that ironic sentences can be uttered with a specific set of prosodic and gestural features which hearers can use to identify ironic intent (e.g., Attardo, Eisterhold, Hay & Poggi, 2003; Padilla, 2004, 2009; Poggi, 2007). Figure 1 illustrates an example of a sincere vs. an ironic performance of the Catalan sentence *Fantàstic!* The prosodic features of the ironic rendition of the sentence (i.e., *Fantàstic!* conveying a negative intent) are slower tempo, lower average pitch, lower pitch range, and lower intensity. As for gestures, the ironic utterance displays a set of visual cues such as raised eyebrows, head tilt, shoulder shrug or a smile (mouth corners up).

Figure 1. Waveform, spectrogram and F0 contour, of two versions of the Catalan utterance *Fantàstic!* ‘Fantastic!’, namely sincere (left) and ironic versions (right). A visual display of the speaker's facial gestures appears in the bottom.



In the last two decades, most of the research focusing on the prosodic and gestural cues of verbal irony has tried to delineate the prosodic characteristics of the so-called ‘ironic tone of voice’ or to describe which gestural markers speakers use to accompany an ironic remark. While there are some studies that have investigated the role of prosodic and gestural features in verbal irony production (e.g. Attardo, 2003, 2011; Padilla, 2004, 2009), very few empirical

studies have addressed the question of how prosodic and visual cues interact with verbal and contextual cues from a comprehension point of view. In this dissertation we use an integrated approach to investigate the contribution of prosodic and gestural cues to verbal irony from a production, comprehension and developmental perspective, trying to answer questions such as the following: How are prosodic and gestural cues integrated in ironic speech? What is the contribution of prosodic, gestural, contextual and lexical propositional cues to verbal irony comprehension? What is the relative contribution of prosodic and gestural cues to verbal irony detection? How strong are multimodal cues compared to contextual ones for irony detection? Do multimodal cues facilitate children's appreciation of irony?

1.2. Prior work

1.2.1. Prosodic markers of verbal irony

Several studies have investigated the role of prosody in the expression and recognition of verbal irony (e.g., Gibbs, 2000; Nakassis & Snedeker, 2002; Anolli, Ciceri & Infantino, 2002; Bryant & Fox Tree, 2002, 2005; Attardo et al. 2003; Laval & Bert-Erboul, 2005; Cheang & Pell, 2008, 2009; Bryant, 2010; Attardo, 2011; Scharrer, Christmann & Knoll, 2011; Padilla, 2004, 2009, 2011; Loevenbruck, BenJannet, D'Imperio, Spini & Champagne-Lavau, 2013). Across languages, ironic utterances have been reported to be produced with a slower speech tempo, as well

as wider pitch and intensity modulations, both in spontaneous and in non-spontaneous speech. Most of the studies have reported that ironic sentences are produced with 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), (see e.g. Bryant, 2010; Attardo et al., 2003; Cheang & Pell, 2009; and Rockwell, 2000, for English; Anolli et al., 2002, for Italian; Cheang & Pell, 2009, for Cantonese; Laval & Bert-Erboul, 2005; Loevenbruck et al., 2013; and González-Fuente, Prieto & Noveck, 2016, for French; Scharrer et al., 2011, for German; Padilla, 2004, 2009, 2011, for Spanish). Other non-F0 features like non-modal voice quality have also been claimed to signal irony (e.g. Van Lancker et al., 1981; Cheang & Pell 2008, 2009). While duration cues seem to be consistent across most of the studies in different languages (e.g., duration tends to slow down in ironic speech), other prosodic cues such as average pitch, pitch variability, and intensity do not show a consistent pattern across studies and across languages (see Bryant, 2011, for a review). This lack of consistency may be due to methodological issues such as differences in the irony subtype under analysis, the language-specific implementation of irony and also the intonational phonology of each language, which might privilege either rising or falling pitch accents (Bryant, 2011; Loevenbruck et al., 2013).

In this respect, a few studies have reported on the use of intonational features in different languages (e.g., Attardo, 2001, for English; Padilla, 2004, 2009, 2011, for Spanish; González-Fuente et

al., 2016, for French). Attardo (2001) distinguished a set of intonational tunes that might be considered as “ironic intonation” (2001: 119), namely a flat contour (neither rising nor falling intonation), question intonation, or what he called “exaggerated intonation patterns” (e.g., a singsong melody). On the other hand, Padilla (2004, 2009) claimed that ironic utterances can be marked with specific rising final inflectional patterns (e.g., Padilla, 2004, 2009). As far as we know, González-Fuente et al. (2016) is the only perception study that has specifically investigated the extent to which specific tonal-nuclear configurations influence irony interpretation when compared to other prosodic cues. Interestingly, the results of this study showed that some ironic utterances were produced using a specific intonational contour (H+H!*H%), which has been described in the French_ToBI annotation system as containing a specific pragmatic meaning related to disagreement in the expression of counterfactual statements (or utterances used when the speaker thinks that the listener holds a contrasting view). However, the use of specific tonal-nuclear configurations with verbal irony still remains unexplored.

All in all, the review of the literature presented above on prosodic markers of irony supports the claim by Bryant (2012) that the notion of an “ironic tone of voice” is oversimplified. However, despite the lack of systematicity in the way that speakers prosodically mark an ironic statement, experimental research on the role of prosodic cues to verbal irony detection has shown that specific prosodic modulations (across languages and across studies)

help listeners to infer irony. Perception studies have shown that ironic intent can be successfully extracted from prosodic cues even in the absence of contextual cues (e.g., Loevenbruck et al., 2013, Padilla 2011). For example, Loevenbruck et al. (2013) found through an identification task that the average accuracy score for the 234 pairs of ironic vs. sincere utterances presented without a previous discourse context was 79%. The ironic and sincere utterances for this study were obtained from a production experiment, and the acoustic analysis showed that sarcastic comments were produced with significantly higher pitch levels, wider span, and longer durations as compared to sincere comments. Similarly, Padilla (2011) found that 50 Spanish listeners successfully identified the ironic utterances from a total of 40 ironic and literal utterances in 92% of cases. In this case, the utterances were presented together with the previous context and were extracted from a corpus of spontaneous speech. After the identification task, participants were asked to judge which cue they considered more useful for their decision (i.e., context or tone of voice). A total of 48% of the participants considered that the ‘tone of voice’ was more useful than the previous context for their decision; 50% of the participants considered that both cues were equally important for their decision, and the remaining 2% responded that context was more useful for them than the tone of voice.

In sum, production studies on verbal irony have clearly shown that, while it is not mandatory (Padilla 2004, 2009), speakers actively use

prosodic modulations in their ironic speech. If they do so, it is not in a regular fashion (often using *opposite* markers) so that a consistent ‘ironic tone of voice’ cannot be defended to exist, as claimed by several researchers (Bryant, 2010, 2011; Padilla 2004, 2009). Despite this variation in production, perception experiments have clearly shown that listeners rely on prosodic cues to detect an ironic intent. In this dissertation, we will try to shed some more light on the interplay and relative contribution of prosodic (and gestural) cues to verbal irony production and comprehension, as well as the underlying cognitive processes that can explain this issue.

1.2.2. Gestural markers of verbal irony

A relatively less explored but also relevant area of research is the study of the visual correlates of verbal irony. It has been shown that in conversation, speakers often use the so-called ‘ironic gestures’ (ironic winks, facial expressions involving specific eye and eyebrow configurations, laughter and smiles, etc.; see e.g. Gibbs, 2000; Bryant, 2011, among many others). Several studies have documented the presence of specific gestures and facial expressions during the production of verbal irony (e.g., Attardo et al., 2003, 2011; Bryant, 2011; Haiman, 1998; Hancock, 2004; Kreuz, 1996; Padilla, 2004, 2009; Caucci & Kreuz, 2012; Gibbs, 2000). For example, 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 comprehension 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). In another study, Caucci and Kreuz (2012) 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 and laughter that occurred in sarcastic utterances. These features have been claimed to express a positive stance between social partners and reinforce a shared positive affective experience (Smoski & Bachorowski, 2003). In contrast with the abovementioned studies, other studies such as Attardo et al. (2003) have also reported that a very common visual cue to irony is in fact the absence of any facial expression, i.e. a sort of expressionless face produced after the ironic utterance, characterized as a "blank face" (Attardo et al. 2003: 243).

The gestural markers mentioned above (smiles, facial expressions) can be understood as social signals that provide relevant communicative information about the ironic intent of the speaker (Bryant, 2011). Another social signal of intentional meaning is gaze behavior. It has been shown that gaze deviation is used by speakers when producing sarcastic utterances. Williams, Burns, and Harmon'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 that the duration of eye contact occurring during sincere statements was longer (63.9%) than in sarcastic statements (52.7%).

However, to our knowledge, no empirical studies have been performed assessing the interplay between gestural and prosodic components in ironic speech, and how gestural features manifest themselves in spontaneous speech, both during and after the production of ironic utterances. The production studies included in Chapters 2 and 3 investigated the role and temporal alignment of prosodic and gestural features produced in spontaneous speech by a professional comedian (Chapter 2) and by non-professional speakers (Chapter 3), paying special attention to the prosodic and gestural components produced during and after the ironic remarks.

Moreover, to our knowledge, there have been no attempts at assessing the role of visual cues in the successful understanding of ironic utterances and how strong they are when compared with prosodic and contextual cues. The experimental studies included in Chapter 4 of this dissertation will deal with the question of the strength of gestural and prosodic cues (in relation to context) in verbal irony comprehension.

1.2.3. The developmental perspective

Previous research on the development of irony comprehension has suggested that appreciation of the speaker's intent (for example, understanding whether the speaker is trying to be nice or mean) requires the assessment and integration of multiple cognitive and

intentional cues. This process entails sophisticated inference processes that become more accurate as children grow up (Ackerman, 1983; de Groot, Kaplan, Rosenblatt, Dews & Winner, 1995; Creusere, 2000; Nakassis & Snedeker, 2002; Harris & Pexman, 2003; Filippova & Astington, 2008). While there are some divergences among studies, most of them agree that children begin to understand the ironic intent of the speaker between 5 and 11 years of age (e.g., Milosky & Ford, 2009) and that they do so by means of contextual and prosodic cues.

Developmental studies have shown that facial gestures help children to detect pragmatic meanings such as belief states. For example, Hübscher et al. (2016) performed an uncertainty detection task with 4- to 6-year-old children through a series of materials that controlled for the presence of lexical, intonational, and gestural markers of uncertainty. Their results showed that the presence of gestural cues led children to a better detection of uncertainty. Moreover, they found that younger children were less sensitive to lexical cues of uncertainty (e.g., the use of adverbial forms such as *maybe*) than to gestural and intonational ones, which suggests that in early pragmatic development the intonational and gestural features of communicative interactions may act as bootstrapping mechanisms. These findings agree with Armstrong, Andreu i Barrachina, Esteve-Gibert & Prieto (2014) results, as they found that facial gestures also seemed to scaffold children's performance in detecting incredulity, another type of belief state.

The majority of developmental studies related to the acquisition of verbal irony agree that both context (Ackerman, 1983; Capelli, Nakagawa & Madden, 1990; Winner & Leekman; 1991) and prosody (Ackerman, 1982, 1983; Capelli et al., 1990; Winner & Leekman, 1991; de Groot et al., 1995; Keenan & Quigley, 1999; Nakassis & Snedeker, 2002; Harris & Pexman, 2003; Climie & Pexman, 2008) are useful cues for children to understand sarcastic remarks. However, they disagree on the specific age at which children start to be able to use such cues for this purpose. Regarding contextual cues, whereas some studies have found that they do not play a role in children's detection of sarcastic remarks until they are 11 years of age (Capelli et al., 1990), other studies showed that they are used by children as young as 6 (Ackerman, 1983; Winner & Leekman, 1991). In a similar way, whereas some studies have shown that children can already use prosody as a cue to detect sarcastic intent at age 6 (Keenan & Quigley, 1999), other studies detected no such evidence until children were 8 (Ackerman, 1983; Capelli et al., 1990) or even older (Winner et al., 1987). In any case, as Nakassis and Snedeker (2002) and Laval and Bert-Eboul (2005) have pointed out, some of these discrepancies across experimental results may be attributable to differences in the materials and procedures used, especially those related to the specific prosodic features of the "ironic tone of voice". In this regard, the only study that has controlled for different kinds of ironic intonation patterns is Nakassis & Snedeker (2002), which tested the role of prosodic cues expressing positive and negative intent in adults' and 6-year-old children's comprehension of irony. Their results showed that

prosody acted as a relational cue, that is, that prosodic features facilitated children's comprehension of an ironic remark when the positive or negative valence of the intonational cues agreed with the ironic interpretation of the utterance (for example, prosodic cues reflecting a negative intention led children to understand that the speaker had a critical attitude, which led them to conclude that the speaker was expressing irony).

To our knowledge no studies have investigated the contribution of prosodic and visual cues to intent by performing fine-grained control of the emotional valences conveyed by prosody and gestures in contrast with those of the literal interpretation of the sentence. In order to fill this research gap, Chapter 5 presents a study in which emotional/intentional valence—ranging from positive to negative—conveyed by prosody and by gestures is controlled for in the materials.

1.3. Theoretical Framework

1.3.1. Multimodal communication: the audiovisual prosody perspective

In this thesis we adopt a multimodal communication perspective. Both verbal and non-verbal information are common features in communication exchanges, and research on multimodal communication has shown that gestures and facial expressions continuously co-occur with speech in everyday interaction and

make a significant contribution to our comprehension of speakers' intentions (e.g., Beattie & Shovelton, 1999, McNeill, 2005; Goldin-Meadow, 2003; Poggi, 2006, 2007). Furthermore, as Poggi (2007: 9) claims, "to exchange information about the environment, our mental and affective states, and our identity, we exploit the whole gamut of our sensory modalities—sight, audition, smell, touch, even taste—and several parts of our body". In everyday communication, all these gestures, body movements, and facial expressions are combined with verbal features to construct complex multimodal messages.

In recent years, the study of human communication from a multimodal perspective has contributed to a better understanding of a variety of pragmatic aspects of languages, providing new knowledge about the contribution of the different sensory modalities to the expression of pragmatic meanings (e.g. deception, irony, persuasion), emotions (e.g. pride, compassion, admiration, sadness, guilt) and social relations (e.g. power relations, social emotions) (see Levinson & Holler, 2004, for a review on the origins and evolution of multimodal communication research). These studies take a cognitive science approach and use methods that range from conceptual analysis (e.g. McNeill, 1992; Poggi, 2007) to empirical research (e.g. Attardo et al., 2003, Kraemer & Swerts, 2004), and simulation on embodied agents (e.g. Poggi, 1999, 2000, Pelachaud & Poggi, 2001).

With respect to the study of verbal irony, it has been claimed to be a multimodal affair which involves the assessment and evaluation of different sources of information (e.g. Bryant, 2011, 2012, Poggi, 2007). As mentioned before, some experimental research has shown that speakers actively use multimodal markers to signal the presence of irony (e.g. Attardo et al., 2003, 2011; Bryant, 2011; Haiman, 1998; Hancock, 2004; Kreuz, 1996; Padilla, 2004; Caucci & Kreuz, 2012; Gibbs, 2000). In Poggi's (2007) book *Mind, hands, face and body. A goal and belief view of multimodal communication*, the multimodal nature of irony is viewed as a consequence of the complex nature of the phenomenon, as the ironic speaker communicates something different from what she/he thinks, but, contrary to the case of deception, she/he also wants to be understood by the listener. Poggi (2007: 365) proposes that, in the cases in which contextual cues are not sufficient to direct the listener towards irony, the sender may alert the addressee in two ways, namely, (a) through *meta-communication*, e.g. a communicative act signaling the presence of irony and specifically meaning "I am being ironic" (as, for example, the unexpressive blank face reported by Attardo et al., 2003), or (b) through *paracommunication*, e.g. another communicative act performed either in sequence (if in the same modality) or at the same time (through other modalities), that clearly contradicts it (for example a bored face while uttering an enthusiastic utterance). As Poggi (2007: 365) contends, "in this case, the sender performs two communicative acts with contradictory meanings, X and not-X: in a sense, two overt communicative behaviours to make the addressee

infer backstage communicative thought”. This double function of multimodal cues has also been signaled by Padilla (2009) for prosodic cues. He claims that the pragmatic functions of the ‘ironic tone of voice’ may range from conveying a procedural pragmatic meaning (in which the intonation helps the listener to achieve the ironic interpretation by restraining the possible interpretative options) and a more specific pragmatic meaning (in which the tone of voice would convey a more specific ironic meaning).

The audiovisual prosody perspective

A recent research line within multimodal communication studies is the audiovisual prosody perspective, which claims that prosodic characteristics of speech are complemented by gestural markers and that they can jointly convey a set of pragmatic meanings, such as prominence and focus marking (e.g. Hadar, Steiner, Grant & Clifford, 1983; Cavé, Guaitella, Bertrand, Santi, Harlay & Espesser, 1996; Kraemer & Swerts, 2007; Swerts & Kraemer, 2008; Dohen & Loevenbruck, 2009; Prieto et al., 2015), face-to-face grounding (Nakan, Reinstein, Stocky & Cassell, 2003), and question intonation (Srinivasan & Massaro 2003). Other studies in this line of research (Cvejic et al., 2010, 2012) showed that visual cues to speech prosody are available from a speaker’s face and that interlocutors use this information to correctly identify prosodic content in a statement despite inter- and intrasubject differences. Recent experiments have focused on the study of acceptability of congruent and incongruent combinations of gestural configurations and

intonational patterns (e.g., Borràs-Comes & Prieto, 2011). The results of this study showed that, while certain gestural configurations are more general than others (and therefore more compatible with the various intonational configurations), others are much more specific and therefore less combinable. Recent studies in this direction showed that gestures provide more conclusive evidence than intonation for interpreting the pragmatic content of a statement (Borràs-Comes et al., 2011, Goldin-Meadow, 2003; Holler & Wilkin, 2009; Krahmer & Swerts, 2004; Prieto et al., 2015; Swerts & Krahmer, 2005).

1.3.2. The Relevance Theory perspective

By revising Grice's (1975) account, Relevance Theory (Sperber & Wilson, 1986/1995, among others) defines pragmatics as a "capacity of the mind, a kind of information-processing system, a system for interpreting a particular phenomenon in the world, namely human communicative behavior. It is a proper object of study itself, no longer to be seen as simply an adjunct to natural language semantics. The components of the theory are quite different from those of Gricean and other philosophical descriptions; they include online cognitive processes, input and output representation, processing effort and cognitive effects" (Carston, 2002: 128-129). To our knowledge, Relevance Theory is the only pragmatic theory that has highlighted the role of prosody in communication, and especially in ironic communication (e.g., Wilson & Wharton, 2006; Wilson, 2013). Within this model, 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). This pragmatic account claims that 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).

Within this model, several authors have proposed an explanation about the way in which prosody and gesture interplay with other units and levels of language (e.g., Wilson & Wharton, 2006, and Escandell-Vidal, 2011a, for prosodic cues; Wharton, 2009, for non-verbal cues). Relevance Theory advocates for the existence of different levels of representation, namely conceptual units (i.e., units containing information on representations) and procedural units (i.e., units providing information about how to operate with those representations) (Wilson & Sperber, 1993: 2). Researchers working within this perspective have proposed that prosodic modulations encode procedural instructions, which guide the inferential process by constraining the range of possible interpretations (e.g. 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). As for gestures and facial expressions, only a few studies have specifically addressed this topic (e.g., Wharton, 2009; De Brabanter, 2010; Forceville, 2014), and all of them agree on

highlighting the role of visual cues, since they may even constitute the only mark that specifically manifests the speaker's ironic attitude/emotion.

In this regard, recent contributions within the relevance-theoretic account of irony suggest that in order to understand an ironic remark it is necessary to identify not only the critical attitude towards the proposition but also the affective/emotional attitude of the speaker towards the utterance (Yus, 2016), thus emphasizing the potential role of prosodic and gestural cues conveying less conventional cues (i.e. emotional/affective) in the comprehension of verbal irony.

1.3.3. Contrast effects in verbal irony comprehension: the Contrast-Assimilation Theory

The Contrast-Assimilation Theory constitutes a recent line of psycholinguistic research which has theoretically introduced and empirically verified a predictive relationship between perceptual contrast effects and the pragmatic functions of verbal irony (Colston, 2002; Colston & O'Brien, 2000). These studies demonstrate that “the pattern of contrast versus assimilation effects found in many psychological research literatures enables prediction of the pragmatic functions interpreted from a speaker's use of figurative language, specifically, the degree of criticism expressed by a speaker using a form of verbal irony” (Colston, 2002: 111).

Research on verbal irony comprehension has investigated the effects of a variety of contrasting patterns. First, studies have mainly focused on examining the role of contextual cues in the detection of ironic intent, revealing that the contextual characteristics of the verbal exchange play a key role in its interpretation (e.g. Kreuz & Glucksberg, 1989; Gibbs, 1994; Kumon-Nakamura et al., 1995). Specifically, some experimental results have demonstrated the key role of contextual contrast effects in irony perception (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003). Ivanko & Pexman (2003) performed several experiments investigating the role of context (and specifically the degree of incongruity between the discourse context and the ironic comment) in the interpretation of literal and sarcastic statements in English. A set of 89 listeners rated 12 sentences such as "Brad is a wonderful singer" which were preceded by discourse contexts with different degrees of situational negativity (i.e. bias towards an ironic interpretation of the statement) using strongly negative (i.e. strongly ironic biased), weakly negative (i.e. weakly ironic biased), and ambiguous (i.e. non-biased) discourse contexts. The authors found that in strongly ironic biased situations the ironic statements were perceived to be more mocking than literal statements, whereas in the weakly negative context condition, the same ironic statements were perceived to be only slightly more mocking than literal statements. The authors concluded that the existence of appropriate contextual conditions—specifically, the degree of negativity of the discourse context—facilitated the detection of sarcasm.

Following this line of research, other studies have focused on the interplay between verbal, prosodic and contextual cues (Woodland & Voyer, 2011; Voyer, Thibodeau & DeLong, 2016; Voyer & Vu 2016). Woodland & Voyer (2011) examined how contrasts between discourse context and tone of voice affected the perception of sarcasm. A total of 82 English listeners were presented with a set of short written discourse contexts presented in ironic and literal biased conditions followed by sentences presented in sarcastic and sincere tone of voice and, crucially, presented in congruent and incongruent context/tone of voice pairings. Subjects were asked to rate the perceived degree of ‘sarcastic irony’ by means of a 5-point Likert-type scale (1 = ‘very sincere’ to 5 = ‘very sarcastic’). Results showed that mid-range ratings of perceived degree of irony and longer reaction times were obtained when tone and context were incongruent (i.e., ironic biased context and sincere tone, and vice versa) compared to when they were congruent (i.e., literal biased context and sincere tone, and ironic biased context and sarcastic tone). The authors conclude that the sarcastic tone of voice may serve to exaggerate the contrast between the statement and the discourse context, which leads to a higher perception of sarcasm, which clearly showed the relevance that contrasting contextual and prosodic cues have in the verbal irony recognition process: the more mismatching cues (contextual and prosodic), the more accurate ratings of irony. Recently, Voyer et al. (2016) conducted a follow-up of Woodland & Voyer's (2011) study in which they run a set of perception experiments introducing ambiguous (i.e.

non-biased) discourse contexts to determine whether a milder contrast between context and propositional information would affect the proportion of sarcastic responses and response time. Results were consistent with Woodland & Voyer's (2011) results. These two studies clearly showed that (a) congruent context/tone of voice pairs facilitate the processing of sarcastic remarks; and (b) the incongruence between tone of voice and discourse context creates a failed expectation that leads to increased difficulty in assessing utterance interpretation. All together, these lines of research suggest that congruency between the discourse context and the tone of voice with which an utterance is produced influence how irony is perceived.

However, though the abovementioned studies clearly show that the interaction between contextual and prosodic cues affects verbal irony comprehension, so far no experimental research has assessed how visual cues may affect it. Most past research on irony detection has relied on either written or auditory materials for the detection of irony, and little is known about the role of visual information. To fill this research gap, Chapter 4 of this dissertation presents a study which introduces a novel aspect to the literature of irony perception by investigating the role of audiovisual cues in the communication of the speaker's intentions, and how critical this visual information is when compared to prosodic and contextual cues. In line with other work on the effect of visual cues, I hypothesize that multimodal (i.e. prosodic and gestural) cues of irony will be stronger than contextual cues for the detection of verbal irony.

1.4. General objectives, research questions and hypotheses

In this dissertation we aim at investigating more closely the prosodic and gestural dimensions of verbal irony. Specifically, we are interested on assessing how speakers use prosodic and gestural cues from a production, perception and developmental point of view. The following set of research questions will be addressed, divided in four chapters:

a) What is the rate of appearance of prosodic and gestural characteristics of verbal irony produced by a professional comedian? Are they temporally aligned? (Chapter 2). I hypothesize that (a1) ironic utterances will display a higher density of prosodic and gestural markers than non-ironic utterances, and that (a2) gestural markers will appear both temporally aligned with prosodic prominence but also independently.

b) What is the rate of appearance of prosodic and gestural characteristics of ironic comments produced by speakers in spontaneous speech? What is the contribution of gestures produced during post-utterance codas to the understanding of verbal irony? (Chapter 3). I further hypothesize that (b1) in non-professional spontaneous speech speakers will employ a higher density of prosodic and gestural markers in ironic compared to non-ironic utterances, and (b2) that they will actively use gestural codas to detect ironic intents.

c) What is the relative contribution of prosodic and gestural cues to the detection of irony? How strong are multimodal (i.e. prosodic and gestural) cues compared to contextual cues for the detection of verbal irony? (Chapter 4). I also hypothesize that (c1) listeners will rely more strongly on gestural information than on prosodic information for detecting irony, and (c2) also on multimodal cues than on contextual ones.

d) Do prosodic and gestural cues to emotion facilitate the detection by children of a speaker's ironic intent? (Chapter 5). I hypothesize that (d) prosodic and gestural cues of emotion will facilitate the detection of irony in early stages of irony acquisition (e.g., 5-year old children).

This dissertation is thus organized into four independent studies, which are presented in Chapters 2 to 5. Chapter 2 is a case study of a professional comedian and focuses on describing how prosodic and gestural cues are employed by the comedian in order to mark the presence of an ironic intent. Previous studies have reported that speakers employ contrasts between ironic and the immediately previous non-ironic speech to signal ironic intent, but there are no studies that have investigated (a) gestural contrasts between ironic and non-ironic previous speech and (b) the temporal alignment of gestural and prosodic patterns in ironic speech. In order to correct these issues, we analyzed the gestural and acoustic cues of a corpus of 21 ironic utterances acts produced by a professional comedian on a TV show. The results showed two main findings: (a) that ironic

utterances display a higher density of both prosodic and gestural markers compared to preceding non-ironic utterances and (b) that gestural markers can appear both temporally aligned with prosodic prominence but can also appear independently, as gestural codas.

In order to extend the findings of the first study to non-professional speech and also to experimentally confirm the findings for gestural codas, I conducted a second study (Chapter 3) which included two different experiments, namely, (a) a production experiment eliciting spontaneous ironic speech and (b) a perception experiment on the contribution of gestural codas to the detection of verbal irony. Results reveals (a) that also in non-professional spontaneous speech speakers employ a higher density of prosodic and gestural markers in ironic as compared to non-ironic utterances; and (b) that speakers detect ironic intent significantly better when post-utterance gestural codas were present than when they were not.

The third study (Chapter 4) investigates the relative contribution of contextual vs. prosodic vs. gestural cues to verbal irony comprehension. Previous findings have shown that the ability to detect speakers' ironic intent lies in the ability to detect mismatches between verbal, contextual, and prosodic cues, emphasizing the preeminence of contextual over prosodic cues in verbal irony detection. By means of three perception experiments in which participants were asked to rate the ironic intent of a set of discourse context-utterance pairs produced with sincere or sarcastic multimodal cues, we tested the strength of gestural cues compared

to prosodic and contextual ones. The first experiment showed that (a) listeners detect irony more accurately when they have access to both prosodic and gestural cues than when they just rely on prosodic information, (b) that listeners rely more strongly on gestural information than on prosodic information, and (c) that listeners rely more heavily on gestural cues than on prosodic or contextual ones for detecting irony. Overall, these results highlight the strength of visual cues relative to prosodic and contextual cues in the detection of speakers' intentionality.

Finally, the fourth study (Chapter 5) investigates whether prosodic and gestural cues to emotion facilitate verbal irony detection by children. Previous studies on children's irony appreciation revealed that prosodic cues play a facilitating role in their irony comprehension between the ages of 8 and 9 and that their irony detection skills are correlated with their empathy development, but the contribution of gestural cues to this issue have been not studied yet. We designed an irony detection task in which three groups of 5-, 8- and 11 years-old children were audiovisually presented with six ironic context-utterance pairs produced with prosodic and gestural cues conveying three different type of emotions: one strongly mismatching negative emotion, one slightly mismatching negative emotion and one matching positive emotion. Results showed that strongly mismatching cues to emotion led to significantly higher irony detection rates in the three age groups, suggesting that mismatching gestural cues facilitate irony appreciation in early stages of development.

González-Fuente S. "[La prosodia audiovisual de la ironía verbal: un estudio de caso](#)". Revista de la Sociedad Española de Lingüística. 2015;45(1):77-104.

2. CHAPTER 2: “La prosodia audiovisual de la ironía verbal: un estudio de caso”

2.1. Introducción

El presente artículo se organizará de la siguiente manera: en este primer apartado presentaremos (1.1) el fenómeno lingüístico de la ironía verbal como un subtipo del fenómeno comunicativo del lenguaje indirecto, (1.2) una revisión de la bibliografía científica que se ha ocupado de investigar el papel que desempeñan los componentes prosódico y gestual en la interpretación de los enunciados irónicos y (1.3) un resumen de las perspectivas y los marcos teóricos que constituirán la base de la discusión de los resultados, como son (a) la perspectiva de la «prosodia audiovisual» (Krahmer y Swerts 2009) —y su particular consideración de la función y de la relación existente entre los elementos auditivos y visuales del habla— y (b) la teoría pragmática de orientación cognitiva conocida como Teoría de la Relevancia —Sperber & Wilson, 1986/1995— (concretamente, presentaremos el análisis funcional que desde esta perspectiva se realiza de la contribución de la prosodia y de los elementos no verbales a la interpretación de los enunciados). A este primer apartado, le seguirán un segundo y un tercer apartados dedicados, respectivamente, a exponer la metodología que se ha empleado para la selección, codificación y análisis de los datos, y a dar cumplido detalle de los resultados obtenidos. Finalmente, concluiremos este artículo realizando una breve discusión de los resultados logrados y proponiendo futuras

líneas de investigación en la materia que nos ha ocupado: la contribución de los componentes prosódico y gestual a la interpretación de enunciados irónicos.

2.1.1. Lenguaje indirecto e ironía verbal

Dentro del complejo sistema de la comunicación humana, uno de los recursos más empleados es el del lenguaje indirecto, esto es, aquel acto lingüístico en el que los constituyentes verbales superficiales no son un reflejo del mensaje que el hablante desea transmitir, o, dicho de otro modo, en el que el significado último de la expresión no está contenido tan solo en la forma proposicional del enunciado, sino que se infiere a partir de la interacción entre esta et al. factores, como el conocimiento compartido entre los interlocutores o el modo en el que el enunciado ha sido proferido (Bryant, 2011: 291). Atendamos al Ejemplo 1:

Ejemplo 1.

A: — ¿Sabías que José ha vuelto a suspender el examen de conducir?

B: — ¡No sabes cuánto lo siento!

En el Ejemplo 1, el enunciado emitido por B puede ser interpretado por A de diferentes maneras. Esas posibles interpretaciones vendrán determinadas no solo por el contenido proposicional del enunciado “No sabes cuánto lo siento”, sino por la interacción entre ese

contenido proposicional, factores de carácter contextual (p. ej. la información que comparten los interlocutores sobre ellos mismos, sobre el contexto situacional y sobre el mundo), y factores de carácter formal (p. ej. cómo ha sido pronunciado el enunciado). En una de esas posibles interpretaciones —aquella en la que ambos saben que “B no soporta a José” y, además, en la que B profiere el enunciado “No sabes cuánto lo siento” alargando los sonidos [a] y [n] de la palabra cuánto mientras abre exageradamente los ojos y esboza una pícaro sonrisa a continuación—, B no solo no “lo siente”, sino que se alegra de que José haya vuelto a suspender el examen de conducir, lo cual constituye un claro ejemplo de uso del lenguaje indirecto y, en este caso concreto, de ironía verbal. Se trata de lenguaje indirecto porque existe una incongruencia entre el contenido literal de la proposición y el contenido implícito, y hablamos de “ironía verbal” porque B no pretende esconder esa incongruencia, sino que, por contra, se esfuerza en remarcar esa disociación entre lo que dice y cómo lo dice con tal de guiar al oyente hacia la correcta interpretación del mensaje. Es en esa voluntad del «ironizador» de que su intención sea percibida por el oyente donde encontramos la clave para entender el papel que desempeñan prosodia y gestualidad en la producción e interpretación de enunciados irónicos. Como veremos en el último apartado de esta introducción (1.3), solo aquellas disciplinas que manejan marcos generales que integran en su análisis los aspectos cognitivos y sociales del lenguaje, como la pragmática, la filosofía o la psicología, pueden ofrecer una explicación más satisfactoria a fenómenos como el de la ironía verbal, pues incluyen en su análisis

factores que la lingüística tradicional había considerado «paralingüísticos» o «extralingüísticos». Estos factores, no obstante, a la luz de los resultados obtenidos por los estudios experimentales que se reseñan en el siguiente apartado (1.2), resultan ser absolutamente necesarios para explicar este tipo de fenómenos tan propios y característicos de la comunicación humana (ejemplo de algunos de los marcos filosófico-pragmáticos y psicológicos propuestos son los de Austin, 1962; Grice, 1975; Clark & Gerrig, 1984; Searle, 1979; Sperber & Wilson, 1986/1995; entre otros).

2.1.2. Los componentes prosódico y gestual en el estudio de la ironía verbal

a) Prosodia: el tono irónico

Como acabamos de ver en el Ejemplo 1, además del contenido proposicional del enunciado, los factores que intervienen en la interpretación de los enunciados irónicos son de distinta naturaleza y pueden agruparse en dos macrocategorías: la primera es la del «conocimiento compartido» existente entre emisor y receptor, que se refiere al conocimiento compartido por ambos sobre el contexto situacional, sobre el mundo y sobre las creencias generales de los hablantes, y la segunda la conforman las pistas o marcas comunicativas que señalan la presencia de ironía, las cuales pueden ser «verbales segmentales» (p. ej. el empleo de determinados adjetivos o adverbios o de una particular disposición sintáctica de

los elementos en la frase¹), «verbales no segmentales» (p. ej. modulaciones de la voz) o «no verbales» (p. ej. expresiones faciales y gestos), clasificación esta propuesta por Scharrer et al., 2011. Muchos son los estudios que han descrito las variaciones de carácter prosódico que se observan al comparar el habla irónica con la neutra, razón por la que se asume que el hablante modula su producción prosódica con el fin de facilitar al oyente la interpretación de la ironía (p. ej. Gibbs, 2000; Nakassis & Snedeker, 2002; Anolli et al., 2002; Attardo et al. 2003, 2013; Caucci & Kreuz, 2012; Laval & Bert-Erboul, 2005; Cheang & Pell, 2008, 2009; Bryant & Fox Tree 2005; Bryant, 2010; Scharrer et al., 2011; Padilla, 2011; Rockwell, 2000). La complejidad del fenómeno y la gran diversidad de efectos irónicos que se producen en la comunicación humana complican enormemente la tarea de establecer una caracterización sólida del tono irónico (como se concluye en Bryant, 2010 y 2011), por lo que la mayoría de los estudios se han centrado en la descripción y el análisis de la prosodia de un subtipo de ironía. El subtipo que ha merecido mayor atención ha sido el de la «ironía crítica» o sarcasmo (p. ej. Attardo et al., 2003, 2013, Caucci & Kreuz, 2012, Cheang & Pell, 2008, y Rockwell, 2000, en inglés; Scharrer et al., 2011, en alemán; Laval & Bert-Erboul, 2005, en francés; Cheang & Pell, 2009, en cantonés), aunque también existen estudios específicos sobre la entonación irónica de las preguntas retóricas o de las hipérboles (Becerra et al., 2013, en español), y algunos que incluyen el estudio

¹ Para más información sobre la relación entre disposición sintáctica e ironía, véase Escandell-Vidal y Leonetti 2014.

de lo que se ha llamado «ironía de imagen positiva» (Ruiz Gurillo, 2008: 51) —aquella en la que la intención no es criticar, sino halagar— (Nakassis & Snedeker, 2020, en inglés, Anolli et al., 2002, en italiano). Estas restricciones, bien sean de carácter teórico o metodológico, han permitido obtener unos resultados que apuntan a la existencia de algunas características prosódicas específicas de algunos subtipos de ironía, así como también observar las afinidades y discrepancias existentes entre los diferentes subtipos. En estos estudios se han analizado las variaciones de elementos prosódicos como la altura (picos, contornos y altura global o local —p. ej. focalizaciones— de F0); la intensidad (global o local —p. ej. palabras enfatizadas—) o la duración (global, de palabras enfatizadas, de segmentos concretos, pausas o silabeo). En todos ellos se aprecian variaciones significativas entre el tono de voz irónico y el no irónico en alguno o en varios de los parámetros acústicos analizados. Sin embargo, mientras que la ralentización del habla en la producción de enunciados irónicos —esto es, el incremento de la duración global del enunciado irónico o de algunos de sus segmentos— parece ser un fenómeno característico del habla irónica que aparece reseñado de manera consistente y transversal en todos los estudios realizados (p. ej. Anolli et ál., 2002; Bryant, 2010; Laval & Bert-Erboul, 2005; Padilla, 2011), los resultados parecen diferir en la dirección en la que se producen las modulaciones de altura e intensidad². En resumen, los estudios

² Consúltese Scharrer et al. (2011) para una amplia revisión sobre las discrepancias existentes entre los resultados de los valores de altura e intensidad de los estudios realizados.

realizados hasta la fecha sobre la prosodia de la ironía parecen confirmar que los hablantes modulan el tono de voz cuando emiten un enunciado irónico y que este contrasta con el habla no irónica, pero, como se desprende de las diferencias —e incluso contradicciones— existentes entre los estudios, no de una manera única e inequívoca.

Si bien es cierto que desde el ámbito de la pragmática se considera que el contexto constituye un factor esencial para que la interpretación irónica emerja, e incluso que solo ese factor puede resultar suficiente para propiciar la correcta interpretación de un enunciado irónico (Kreuz & Glucksberg, 1989; Gibbs, 1994; Kumon-Nakamura et al., 1995; Utsumi, 2000; Ruiz Gurillo, 2008), también se ha afirmado que los hablantes se sirven de otros indicadores —como el tono de voz³— para facilitar el complejo proceso cognitivo que implica la comprensión de un enunciado irónico (Ruiz Gurillo, 2008). Recientes estudios experimentales acerca de la interacción entre contexto y entonación señalan que el tono de voz que empleamos cuando ironizamos sirve efectivamente para señalar el contraste existente entre el enunciado y el contexto,

³ El «tono de voz» entre otros indicadores. A este respecto, y en el marco de la Teoría de la Relevancia, en Ruiz Gurillo (2008) se afirma que, al tratarse la ironía de un hecho pragmático de carácter básicamente contextual, el hablante ha de emplear ciertas habilidades «inferiores» (consideran que la ironía es una habilidad metarrepresentacional de carácter superior —p. ej. que requiere de un mayor esfuerzo cognitivo para comprenderse—), para que su enunciado resulte óptimamente relevante. Estas habilidades metarrepresentacionales de orden inferior —p. ej. más fácilmente comprensibles—, serían el tono de voz, la hipérbole o el discurso directo, entre otras), y, como ya se ha dicho, actuarían a modo de índices o indicadores.

y que, aun no siendo estrictamente necesario para la comprensión de un enunciado irónico el empleo de modulaciones prosódicas, los hablantes perciben mayor naturalidad en aquellos enunciados realizados con un tono de voz irónico que en aquellos otros que se realizan con un tono neutro (Woodland & Voyer, 2011; Voyer et al. 2016).

b) Gestualidad

Las investigaciones sobre la utilización de gestos en combinación con el habla sugieren que ambas modalidades discursivas, la verbal y la gestual, surgen de una misma estructura conceptual a través de un proceso integrado de construcción de enunciados (McNeill, 1992, 2005). Así, desde esta perspectiva se sostiene que habla y gestos forman un sistema único y unificado, y que los gestos no solo coocurren con el habla, sino que son coexpresivos semántica y pragmáticamente, poniendo de manifiesto la congruencia de formas y regularidades sistemáticas en cuanto a su posición y su sincronía, y conformando conjuntamente el «producto final» que conciben los hablantes en el diseño o construcción de sus enunciados (Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992, 2005). Según estos investigadores, esa conformación conjunta no implica que el gesto se muestre siempre redundante con el contenido del discurso, sino que en muchas ocasiones completa o complementa — no solo por adición, sino también por restricción— su significado. Desde este prisma, la mayoría de los gestos que se producen conjuntamente con el habla estarían actuando a modo de marcadores o puntualizadores metadiscursivos, reflejando la

función pragmática de un enunciado en el discurso o proporcionando indicios acerca de su estructura. En cuanto a los estudios que han abordado el componente gestual en la producción y percepción de los enunciados irónicos, lo primero que cabe decir es que son escasos y de proceder menos sistemático que los dedicados a la prosodia. Aun con ello, las investigaciones llevadas a cabo muestran que el habla irónica se acompaña frecuentemente de gestos y expresiones faciales como movimientos de cabeza, cejas, boca y brazos, así como de otros elementos no verbales, como la risa o la mirada. Cabe también reseñar que la aproximación al estudio de los componentes gestuales del habla irónica se ha llevado a cabo principalmente desde dos perspectivas diferentes, aunque relacionadas: aquella que aborda su estudio desde el análisis del habla humorística (que incluye el uso de expresiones irónicas) (Attardo et al., 2011; Caucci & Kreuz, 2012; Tabacaru & Lemmens, 2014), y aquella que, de manera inversa, se centra en el análisis de la expresión de la ironía (entre cuyas metas comunicativas se encuentra el humor) (Attardo et al., 2003; Bryant, 2011, 2012; Haiman, 1998; Hancock, 2004; Kreuz, 1996; Williams et al., 2009).

En resumen, y recogiendo lo expuesto en puntos anteriores, es dentro de este «significar» a otros niveles, bien sea junto a la prosodia o de manera independiente, donde debemos buscar la contribución que la gestualidad puede realizar a la correcta interpretación de un enunciado irónico, para lo cual creemos que es necesario caracterizar la naturaleza de esa contribución, relacionarla con la prosodia y encajar ambos componentes en un modelo

pragmático que trate de dar cuenta de la complejidad del fenómeno de la ironía sin relegar la prosodia y la gestualidad al ámbito de lo extra —o de lo para—, o al menos no sin antes haber analizado pormenorizadamente el tipo de informaciones que ambas pueden codificar, tanto de manera conjunta como independiente, y la función específica que ambos elementos puedan desempeñar. En el siguiente punto recogeremos los marcos teóricos que creemos que se ajustan más a este objetivo.

2.1.3. La Teoría de la Relevancia y la perspectiva de la Prosodia Audiovisual

Como manifestación propia del lenguaje indirecto, el fenómeno de la ironía verbal ha sido abordado desde muy diversas perspectivas, dando lugar así a múltiples y variadas tipologías en función del enfoque y del criterio clasificatorio escogido. Desde el ámbito de la psicolingüística, por ejemplo, se ha sugerido que la ironía verbal se emplea para alcanzar metas social y comunicativamente complejas⁴ (Kreuz & Roberts, 1995; Leggit & Gibbs, 2000). En otros estudios, cuyo foco se dirige hacia las funciones sociales de la ironía —como es el caso de la Tinge Hypothesis (Dews & Winner, 1995)—, se ha abordado el fenómeno atendiendo a los matices que el uso de

⁴ Por ejemplo, parece ser que preferiríamos el lenguaje irónico al verbal con el objetivo de dotar de humor a una situación, lo cual estaría en consonancia con lo observado en algunos estudios realizados sobre la percepción de la ironía verbal que concluyen que esta se percibe como más divertida y los hablantes irónicos como más graciosos (Gibbs 2000).

expresiones irónicas —bien sea con intención crítica o halagadora— imprime en la interpretación final del mensaje. Por otro lado, desde teorías pragmáticas de orientación cognitiva (p. ej. la Teoría de la Relevancia —en adelante, TR— (Sperber & Wilson, 1986/1995), o la Pretense Theory (Clark & Gerrig, 1984)), se ha tratado de explicar el fenómeno de la ironía verbal atendiendo a los procesos de producción, comprensión y procesamiento cognitivo de los enunciados irónicos. Los datos experimentales que se presentan en este artículo se discutirán en la sección final desde la perspectiva de la TR⁵, pues varios son los autores que han realizado en este marco propuestas de explicación acerca de la naturaleza y el modo en que prosodia y gestualidad se sitúan y se articulan con otras unidades y niveles de la lengua (p. ej. Wilson y Wharton, 2006, y Escandell-Vidal, 2011a, para los componentes prosódicos; Wharton, 2009, para los elementos no verbales). A nuestro juicio, la potencia explicativa de la TR deriva de considerar que no todos los elementos lingüísticos contribuyen del mismo modo a la interpretación de un enunciado. Así, la TR aboga por la existencia

⁵ Desde la perspectiva de la TR, se afirma que la comunicación humana es posible gracias a la conjunción de tres factores: la ostensión (p. ej. la conducta por la que un ser humano manifiesta la intención de comunicar algo); la inferencia (p. ej. el proceso por el que se produce la interpretación de un enunciado), y el compromiso con la búsqueda de la relevancia. Dicho de otro modo: un acto comunicativo resulta exitoso cuando el emisor produce un enunciado suficientemente relevante como para que el receptor lo interprete de forma satisfactoria. Esa «relevancia suficiente» es el quid de la Teoría de la Relevancia, uno de cuyos pilares consiste en considerar que la cognición humana está claramente orientada a alcanzar los mayores beneficios cognitivos con el menor esfuerzo de procesamiento posible. La «relevancia», por tanto, es un concepto comparativo que relaciona los supuestos que proporciona el emisor (aquellos que él considera suficientes para el éxito del acto comunicativo en curso) con aquellos supuestos que reconstruye el destinatario del mensaje.

de distintos niveles de representación en los que operan unidades cuya aportación es de diferente naturaleza: unidades conceptuales —aquellas que contienen información sobre las representaciones— y unidades procedimentales —aquellas que aportan información sobre cómo operar con esas representaciones— (Wilson & Sperber, 1993: 2). Es en este segundo grupo de unidades —las procedimentales— en el que se inscribiría la contribución de algunas de las características prosódicas y gestuales que acompañan al habla⁶. En este sentido, varios son ya los estudios enmarcados en la perspectiva relevantista que han propuesto que las modulaciones prosódicas codifican instrucciones procedimentales que guían los procesos inferenciales a través de la reducción del rango de posibles interpretaciones de un enunciado (House, 1990, 2006; Clark & Lyndsey, 1990; Fretheim, 2002; Wilson & Wharton, 2006; Escandell-Vidal, 1998, 2011a, 2011b; Prieto et al., 2013). En cuanto a los gestos y expresiones faciales, siendo menor la atención que estos han merecido, los estudios publicados hasta la fecha (p. ej. Wharton, 2009; De Brabanter, 2010; Forceville, 2014) coinciden tanto en la importancia que les otorgan a estos elementos (pues pueden incluso constituir la única marca que manifieste la intención del emisor), como en la necesidad de realizar una distinción clara

⁶ Según la TR, las unidades de procesamiento operan a distintos niveles: el nivel de las «explicaturas inferiores», en el que las unidades —p. ej. determinantes o tiempos verbales— guían al destinatario hacia la identificación del contenido explícito que el emisor quiere comunicar; el nivel de las «explicaturas ilocutivas» (o superiores), en el que las unidades —p. ej. patrones entonativos, marcas léxicas de evidencialidad— dan cuenta de la expresión ilocutiva o actitud del hablante; y el nivel de las «implicaturas», en el que las unidades —p. ej. marcadores del discurso— indican cómo conectar el contenido proposicional con otras informaciones del contexto (véase Escandell-Vidal, 2011a, para una ampliación).

entre los diferentes tipos de gestos que producimos al comunicarnos, pues, de la misma manera que sucede con los elementos prosódicos, su naturaleza puede ir desde lo simbólico (p. ej. universal) hasta lo convencional (p. ej. lingüístico). En este sentido, el refinamiento tipológico de este modelo nos permite no solo caracterizar y asociar con un determinado nivel funcional la contribución de aquellos patrones prosódicos y gestuales que encajan con las funciones de carácter procedimental descritas (véase nota 6), sino también la de aquellos que no necesariamente codifican instrucciones específicas de procesamiento, pero que de manera indudable orientan la interpretación de un enunciado. En estrecha relación con lo expuesto, creemos oportuno señalar que dentro del marco de la TR existen también numerosos estudios dedicados al fenómeno de la comunicación humorística (p. ej. Yus, 1997, 2003; Ruiz Gurillo & Alvarado Ortega, 2013). Dada la estrecha relación existente entre ambos fenómenos —ironía y humor—, se discutirán brevemente los resultados obtenidos en el presente estudio a la luz de estos trabajos.

Por otro lado, desde la perspectiva de la «prosodia audiovisual» se afirma que las características prosódicas del habla son, cuando menos, complementadas por marcas gestuales (Krahmer & Swerts, 2004; Swerts & Krahmer, 2005). Así, existen trabajos recientes en los que incluso se ha observado que los gestos proporcionan indicios más concluyentes que la entonación a la hora de interpretar el contenido pragmático de un enunciado (Borràs-Comes et al., 2011; Prieto et al., 2015). Otros estudios (Cvejic et al., 2010, 2012)

han obtenido resultados que indican que los hablantes son capaces de realizar una suerte de representación prosódica abstracta a partir de las pistas o marcas visuales que obtienen de sus interlocutores, circunstancia que les permite interpretar correctamente un enunciado a pesar de las diferencias inter- e intrasujeto. En esta línea, algunos experimentos más recientes se han centrado en el estudio de la aceptabilidad, del grado de especificidad y de la interpretación resultante de la combinación entre configuraciones gestuales y patrones entonativos (Borràs-Comes & Prieto, 2011). Parece ser que, por un lado —y como se presumía—, las diferentes combinaciones arrojan diferentes interpretaciones y, por otro, que no todas las combinaciones son aceptables. Además, algunas de esas configuraciones gestuales son más generales que otras (y, por lo tanto, más compatibles con las distintas configuraciones entonativas), mientras que otras son mucho más específicas y, por tanto, menos combinables. A la luz de estos resultados, se ha puesto de manifiesto que la combinación entre determinadas configuraciones gestuales y entonativas arroja diferentes interpretaciones, las cuales están relacionadas con categorías prosódicas distintas, dato revelador que supone, creemos, un claro estímulo para abordar el estudio del fenómeno de la ironía desde esta perspectiva. No tenemos constancia de que hasta la fecha se haya realizado un estudio detallado sobre cuándo y cómo interactúan prosodia y gestualidad en los enunciados irónicos. El presente trabajo, de carácter meramente exploratorio —tanto por diseño como por extensión—, pretende tan solo abrir esa puerta a través de la exposición de los resultados de un sencillo estudio de

caso realizado sobre un corpus de veintiún enunciados irónicos, los cuales fueron sometidos a (1) un análisis cuantitativo, que sirvió para caracterizar la producción de marcas prosódico-gestuales globales del corpus, y (2) un análisis cualitativo de dos de los enunciados, en el que nos detuvimos a observar las sincronías existentes entre marcas prosódicas, gestuales y léxicas, así como a determinar la función que, de manera conjunta o independiente, desempeñaban todas ellas.

2.2. Metodología

Para realizar el estudio de caso que presentamos a continuación, elaboramos inicialmente un corpus de treinta enunciados irónicos, cuya selección, filtrado y posterior análisis se realizó según los criterios que describimos a continuación. Los treinta enunciados irónicos⁷ fueron seleccionados y extraídos por el autor de un total de ocho vídeos⁸ que contienen otros tantos monólogos humorísticos. El intérprete de estos monólogos (pertenecientes al

⁷ Hemos delimitado la unidad enunciado siguiendo un criterio de naturaleza discursiva, esto es, atendiendo a razones pragmáticas, y no gramaticales. Siguiendo a Escandell-Vidal, 2006: 28, hemos huido de la identificación enunciado-oración, considerando que «una unidad del discurso no puede tener más límites que los que establece el emisor y su intención comunicativa, independientemente del grado de complejidad de su realización formal». Así, se han considerado «enunciados irónicos» aquellas oraciones —o series de oraciones— cuya intención comunicativa —ironizar— se mantenía constante durante todo el acto comunicativo.

⁸ Los vídeos tienen una duración media de 5'23'', y están disponibles de manera gratuita en la página web <http://www.youtube.com>.

género semiespontáneo⁹ del monólogo televisivo) es el cómico Andreu Buenafuente, y el marco contextual es el de un programa de humor para la televisión. El hecho de haber seleccionado este género se debe a que consideramos que sus características propiciarían, por un lado, la segura aparición de enunciados de carácter irónico —al ser precisamente el humorismo una de las metas de la ironía verbal (véase Attardo et al., 2011, 2013; Ruiz Gurillo, 2013)— y, por otro, la casi segura aparición de marcas prosódicas y gestuales —dado el componente dramático del género—. Creemos que todo ello no va en detrimento del objetivo del estudio —pues no es este el de caracterizar el habla irónica en situaciones espontáneas—, sino que, por contra, precisamente por tratarse de un género que se produce en una situación y un contexto muy determinados, las variables que pudieran afectar a los datos gozan de un mayor control. Además, como apuntan Attardo et al., 2003: 246-247, los datos extraídos de textos literarios o de otros textos no espontáneos pueden llegar a ser «tan reveladores como los datos obtenidos de manera natural». Con tal de garantizar la prototipicidad de los enunciados irónicos seleccionados, nos cercioramos de que todos ellos se ajustaran a la definición de enunciado irónico propuesta por Wilson y Sperber (1992: 59-60)¹⁰.

⁹ Están en su mayoría guionizados, aunque el monologuista puede desviarse del guion, lo cual suele suceder con frecuencia en el caso particular que nos ocupa. De cualquier modo, no creemos que este hecho condicione substancialmente las observaciones realizadas sobre la ironía, puesto que esta no suele estar guionizada, sino que forma parte sustancial del propio género del monólogo.

¹⁰ Desde la propuesta relevantista, se considera que los enunciados irónicos son una variedad de «cita indirecta». Las “citas indirectas” —en oposición a las “citas directas”— son aquellas en las que un enunciado no se reproduce de manera

La variable “subtipo de ironía” no fue contemplada en la selección de los enunciados del corpus, por lo que fueron considerados enunciados irónicos todos aquellos que cumplían con el criterio apuntado arriba, así como con la consideración general que desde el ámbito de la psicología realiza Gibbs, 2000: 13: “Cualquier forma de ironía refleja claramente la idea de un hablante produciendo algún tipo de contraste entre expectativas y realidad”. A continuación, a fin de confirmar la prototipicidad de los enunciados irónicos seleccionados, realizamos un test perceptivo a cuatro informantes —tres hombres y una mujer, con edades comprendidas entre los 27 y los 35 años, y nivel de estudios universitario— que consistió en valorar el grado de ironía (en una escala del 1 al 5 —de ninguna a mucha—) que percibían en cada uno de los enunciados presentados aisladamente. Como resultado de este filtro, fueron finalmente seleccionados para ser objeto de análisis aquellos que obtuvieron una puntuación de 4,5 puntos o superior, lo que redujo el corpus a la cantidad de veintiún enunciados irónicos.

Posteriormente, capturamos el sonido de los archivos de vídeo con el programa de libre distribución Audacity (Audacity Team, 2014), con el que generamos veintiún archivos sonoros (en formato wav y de 16 bits) que contenían el sonido del enunciado considerado irónico y los 10 segundos anteriores al mismo con tal de poder realizar las comparaciones oportunas entre habla irónica y habla no

exacta, sino únicamente su significado. Además, para que la mención indirecta de una “proposición”, de un “significado” o de “un pensamiento” sea considerada irónica, esta debe ser expresada mediante una clara actitud de desaprobación o rechazo hacia el contenido de la misma.

irónica. A continuación, las grabaciones fueron analizadas utilizando el programa de libre distribución PRAAT (Boersma y Weenik 2008), diseñado para el análisis acústico del habla.

2.2.1. Análisis cuantitativo

a) Prosodia

El análisis cuantitativo de los elementos prosódicos de los veintiún enunciados irónicos y de los veintiún enunciados no irónicos consistió en la extracción de cuatro parámetros acústicos relacionados con la frecuencia fundamental (F0), la amplitud y el tiempo (siguiendo la propuesta de Bryant 2010). De los primeros, se extrajeron: (1) la media de F0 del enunciado (en Hz) y (2) la variabilidad de F0 (p. ej. la media de las desviaciones de los valores de F0 respecto a la F0 media en cada punto de la curva melódica de cada uno de los enunciados) (en Hz); en cuanto a la amplitud, se extrajeron los valores de (3) la amplitud media (en dB); y, por último, en cuanto al tiempo, calculamos (4) la duración media de la sílaba (DMS), esto es, el tiempo total que tarda en pronunciarse el enunciado dividido entre el número de sílabas de ese mismo enunciado (en milisegundos), cuyo valor da cuenta conjuntamente tanto de la separación entre palabras, como del silabeo y del alargamiento significativo de segmentos, que son los fenómenos relacionados con la duración que se analizan en el completo y exhaustivo estudio sobre la prosodia irónica del español de Padilla 2011. Los datos de los cuatro parámetros acústicos extraídos fueron sometidos a cuatro tests estadísticos *t-test* con tal de determinar la

independencia de las medias, siendo la variable independiente Tipo de enunciado (irónico frente a no irónico) y los cuatro parámetros acústicos (F0 media, Variabilidad de F0, Amplitud media y DMS —p. ej. duración media de las sílabas—) las variables dependientes.

b) Gestualidad

Las marcas gestuales fueron manualmente anotadas por el autor utilizando el programa informático ELAN (Lausberg y Sloetjes 2009)¹¹ siguiendo el manual de codificación de gestos y expresiones faciales de Allwood, Cerrato, Jokinen, Navarretta & Paggio (2007) y Nonhebel et al. (2004). Los componentes gestuales etiquetados son todos aquellos que se han descrito en la bibliografía como posibles marcas de ironía, como los movimientos de cabeza, cejas y boca, la oclusión/apertura de los ojos y los gestos producidos con las manos, así como las risas y la desviación de la mirada (p. ej. Attardo 2003, 2011; Bryant 2011; Rockwell 2000; Tabacaru & Lemmens, 2014; Williams et al., 2009). Todos los gestos se etiquetaron durante la producción del enunciado, así como en aquellos instantes inmediatamente posteriores a la producción del enunciado en los que consideramos que el gesto producido estaba claramente integrado en el acto de habla en cuestión. Dada la variación observada entre condiciones experimentales en términos de duración de los enunciados, se procedió a dividir el número de marcas gestuales etiquetadas entre las sílabas proferidas en cada

¹¹ ELAN es una herramienta informática de libre acceso que se emplea para el etiquetado y la alineación de transcripciones y contenido audiovisual.

uno de ellos, en aras de que la comparación entre enunciados irónicos y no irónicos no resultara afectada por el mayor contenido segmental de unos respecto a los otros. El ajuste porcentual supuso una reducción del 12% de los datos obtenidos para los enunciados irónicos.

2.2.2. Análisis cualitativo

El segundo de los análisis realizados fue de carácter cualitativo y consistió en la descripción minuciosa de la prosodia y de la gestualidad del emisor durante la producción de los enunciados irónicos y de la relación de ambos con el componente léxico-sintáctico, así como de la función que desempeñaban. Para ello, al análisis de los fenómenos prosódicos expuesto en el anterior apartado añadimos el análisis funcional fonológico de los patrones entonativos del sistema Sp_ToBI (Prieto y Roseano 2010). De manera análoga, a la descripción minuciosa de las marcas gestuales reseñadas en el apartado anterior se añadió en este segundo análisis la clasificación funcional de los gestos que se describe en McNeill 1992, adaptando el significado de los gestos a las necesidades específicas de nuestra tarea¹². Esta clasificación se basa en criterios de forma (configuración manual y trayectorias) y de significado (la relación que se percibe del gesto con el contenido y con la

¹² Realizamos la adaptación siguiendo la sugerencia de Cartmill et al. (2012: 222): “Aunque las directrices para describir la forma del gesto se pueden aplicar de manera útil a cualquier tarea, cuando el objetivo es asignar significado al gesto, necesitamos construir categorías que sean apropiadas para la tarea en cuestión”.

estructura discursiva). Las cuatro categorías que McNeill establece en cuanto a la forma son los gestos deícticos, los cuales dirigen la atención hacia un objeto determinado (bien sea utilizando los brazos o la cabeza); los gestos convencionales, que son símbolos con un significado compartido por una comunidad —p. ej. el gesto de OK—; los gestos representacionales (icónicos y metafóricos), los cuales hacen referencia a objetos, acciones o relaciones por medio de la recreación de la forma o del movimiento y, por último, los gestos rítmicos (beat gestures), que, pese a no tener un significado semántico claro, son prototípicamente un reflejo de la producción del hablante de las estructuras discursivas o narrativas. En cuanto a las fases temporales de la realización de los gestos, esta se divide en tres fases claramente definidas: preparación, stroke (p. ej. ejecución, cuyo punto de mayor extensión e intensidad se denomina ápex ‘cima’) y retracción. En el caso de los gestos rítmicos, los instantes de mayor intensidad —las cimas— suelen aparecer alineados con marcas prosódicas como los picos de F0, lo cual da cuenta de la estrecha relación de los componentes prosódico y gestual. La observación de estas alineaciones constituye uno de los principales objetivos de este segundo análisis.

2.3. Resultados

2.3.1. Análisis cuantitativo

En la Tabla 2 podemos observar los resultados de la media y la desviación estándar de los cuatro parámetros acústicos observados,

tanto para los veintiún enunciados irónicos como para los veintiún enunciados no irónicos producidos en el habla inmediatamente anterior.

Los datos de los 4 parámetros analizados fueron sometidos a 4 tests estadísticos *t-test* con tal de determinar la independencia de las medias, siendo la variable independiente *Tipo de enunciado* (irónico frente a no irónico) y los cuatro parámetros acústicos (*F0 media*, *Variabilidad de F0*, *Amplitud media* y *DMS* —i.e. duración media de las sílabas—), las variables dependientes. El análisis estadístico muestra cómo son únicamente la *Variabilidad de F0* y la *DMS* los parámetros acústicos que globalmente distinguen de manera significativa ambos tipos de enunciados ($p < 0.5$).

En cuanto a los resultados de las marcas gestuales, en la Tabla 3 se presentan los resultados para cada tipo de enunciado de la media de marcas producidas por enunciado. La longitud (medida en sílabas) de los enunciados irónicos seleccionados resultó ser un 12% mayor que la de los no irónicos, por lo que los valores presentados han sufrido una corrección para ajustar el tiempo en ms de ambos y garantizar la viabilidad de la comparación entre los dos tipos de enunciado (consultar el apartado 2.2 — Metodología). Así, observamos cómo los enunciados irónicos se produjeron en general con un mayor número de marcas gestuales en comparación con los no irónicos.

Tabla 2. Media y desviación estándar de los valores de los cuatro parámetros acústicos recogidos de los 21 enunciados irónicos y de sus 21 enunciados no irónicos precedentes. Los valores de *F0_media* y *F0_variabilidad* se muestran en Hz, los valores de *Amplitud_media* en dB, y los valores de DMS en milisegundos.

Parámetro acústico	Enunciados irónicos		Enunciados no irónicos	
	<i>Media</i>	<i>Desv. Estándar</i>	<i>Media</i>	<i>Desv. Estándar</i>
F0_media (Hz.)	159,7	19,40	155,52	15,06
F0_variabil. (Hz.)	42,29*	27,75*	30,76*	13,10*
Ampl._media (dB)	68,13	5,55	65,01	3,82
DMS (ms.)	197*	56*	173*	35*

Nota. La marca ‘*’ señala los valores estadísticamente significativos en los *t-tests* ($p < 0.05$).

Los resultados de los múltiples tests estadísticos chi-cuadrado para variables nominales realizados entre la variable *Tipo de enunciado* (irónico frente a no irónico) y cada una de las marcas gestuales etiquetadas (ausencia frente a presencia (1-3 ocurrencias) frente a abundancia (+ de 4 ocurrencias)) mostraron diferencias significativas entre ambos tipos de enunciado en las siguientes marcas: mueca de la boca (0,9 en enunciados irónicos frente a 0,2 en enunciados no irónicos), fruncido (1,4 frente a 0,3) y arqueado (5,2 frente a 2,1) de cejas, ladeado de cabeza (2,1 frente a 0,4), semioclusión de ojos (1,5 frente a 0,4) y risas/sonrisas (3,9 frente a 1,9). Aunque no queda reflejado en la Tabla 3, cabe destacar que se observó un mayor número de enunciados irónicos con producciones

gestuales posteriores a la pronunciación del contenido verbal segmental del enunciado (en el 66% de los enunciados irónicos y en el 28% de los no irónicos).

Tabla 3. Media de las marcas gestuales que aparecen por enunciado observadas en los 21 enunciados irónicos y en los 21 enunciados no irónicos.

Marca gestual	Enunciados irónicos <i>Media</i>	Enunciados no irónicos <i>Media</i>
Cabeza — Asentimiento	5,2	4,8
Cabeza — Ladeado	2,1*	0,4*
Cabeza — Sacudida	0,3	0,2
Cejas — Arqueado	5,2*	2,1*
Cejas — Fruncido	1,4*	0,3*
Boca — Mueca (labios estirados)	0,9*	0,2*
Ojos — Semioclusión	1,5*	0,4*
Sonrisa/Risa	3,9*	1,9*
Mirada — Desvío	1,1	0,6
Manos — Gesto metafórico	0,6	0,4
Manos — Batido (<i>Beat gesture</i>)	4,5	3,9

Nota. La marca ‘*’ señala los valores estadísticamente significativos en los tests chi-cuadrado ($p < 0.05$).

En resumen, lo que observamos en los resultados del análisis cuantitativo es que los enunciados irónicos se producen con una F0 media y una amplitud media ligeramente superior —aunque no de manera significativa— en los enunciados irónicos respecto a los no irónicos, así como con una variación de F0 y una duración media de la sílaba significativamente diferentes entre ambas condiciones. De manera análoga, también observamos cómo las marcas gestuales aparecen de modo general en mayor porcentaje en los enunciados irónicos que en los no irónicos, y cómo ambas condiciones muestran también diferencias estadísticamente significativas en la frecuencia de aparición de 6 de las 11 marcas.

2.3.2. Análisis cualitativo

A continuación, a través de un análisis más pormenorizado de dos de los 21 enunciados irónicos, señalaremos las conexiones existentes entre las marcas verbales segmentales (i.e. léxico-sintácticas), verbales no segmentales (i.e. prosódicas) y no verbales (i.e. gestuales), así como apuntaremos las diferentes funciones que los componentes prosódicos y gestuales parecen desempeñar.

Enunciado 1

“Es rápido, ¿eh? ¡Qué sagaz! ¡Se tendría que llamar José María Sagaz!”.¹³

¹³ El enunciado no irónico inmediatamente anterior era *“Oye, la noticia del día es que Aznar ha afirmado que ahora- ahora ya sabe que no había armas de*

El análisis de los parámetros acústicos de este enunciado irónico respecto al habla no irónica inmediatamente anterior, muestra una ligera elevación de la F0 media de aquel respecto a este (178,02 Hz. frente a 157,23 Hz), así como una notablemente mayor variabilidad de la F0 (37,16 frente a 26,15). La amplitud media también es ligeramente más elevada (67,55 dB frente a 65,23 dB), y la duración media de la sílaba se muestra notablemente diferente¹⁴ entre ambos tipos de enunciado (201,3 ms. frente a 164,3 ms.). A nivel local, observamos un alargamiento claro de varios sonidos, siendo los dos más destacados el sonido [r] inicial de “rápido” (véase Figura 1—arriba) y el segundo sonido [a] de “sagaz” (véase figura 1—abajo) (la palabra “sagaz” aparece dos veces, y en ambas ocasiones se produce ese alargamiento). En cuanto a los patrones entonativos, cabe destacar que el enunciado consta de tres frases entonativas independientes, que se corresponden con las tres oraciones del enunciado: “Es rápido, eh?”, “Qué sagaz” y “Se tendría que llamar José María Sagaz”, cuyos acentos tonales más destacables recaen en las palabras que ya hemos indicado —“rápido” y, por dos veces, “sagaz”—. En el primer caso, observamos cómo el acento tonal que recae sobre la palabra “rápido” (véase Figura 1—arriba) es el típico

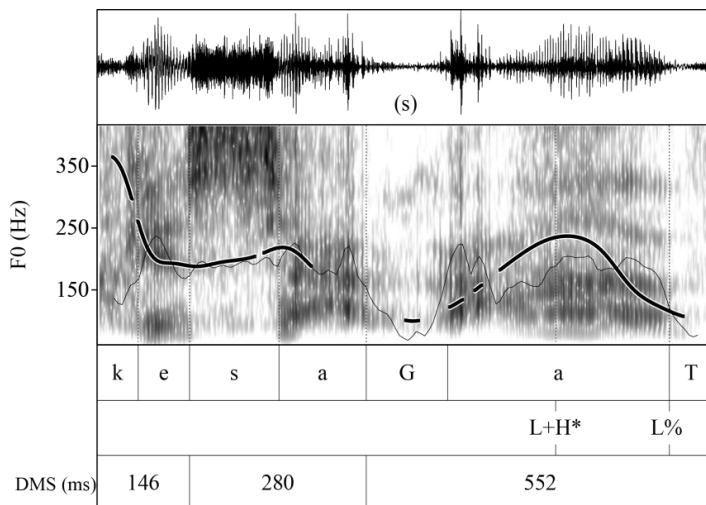
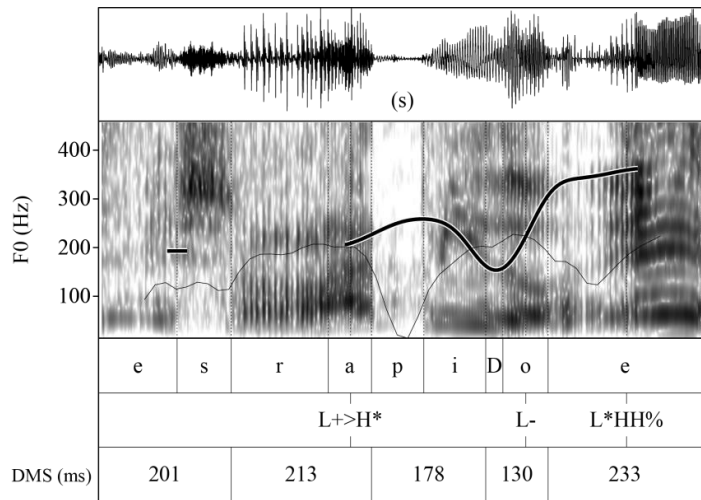
destrucción masiva en Irak”, el cual hace referencia a una noticia aparecida en febrero de 2007.

¹⁴ En todos los enunciados en los que aparecía más de una oración (como es el caso del presente enunciado irónico), el tiempo transcurrido entre las oraciones no fue en ningún caso computado a la hora de calcular la duración media de la sílaba, pues los resultados podrían haberse visto afectados por esta circunstancia. Aún y así, como observamos, la diferencia entre ambas tasas de habla (irónica frente a no irónica) sigue siendo muy notoria.

acento prenuclear de frase declarativa de foco ancho, aunque, al producirse a una frecuencia tan alta, la percepción de ese significado fonológico neutro queda algo desdibujada, y se percibe una clara carga enfática en el desplazamiento de ese pico tonal. En el caso de “sagaz”, en ambas apariciones de la palabra (véase Figura 2—abajo para la segunda de ellas) observamos cómo el acento tonal empleado es el típico acento tonal enfático del foco contrastivo (L+H* L%), entre cuyas funciones pragmáticas se encuentra la de transmitir “obviedad”, lo cual da buena cuenta de la intención irónica del emisor al pronunciar la palabra “sagaz”.

En cuanto a la gestualidad, en este enunciado destacamos la aparición de dos tipos de gestos distintos, unos de carácter metafórico y los otros de carácter rítmico. Los de carácter metafórico aparecen combinados y se producen durante la pronunciación de la palabra “rápido” (véase Figura 2—izquierda y centro). El primero de ellos consiste en un movimiento rápido de batida lateral del brazo hacia ambos lados; el segundo consiste en un chasquido de dedos que marca metafóricamente los puntos espaciales entre los que se realiza ese movimiento, y el tercero es un giro repetido rápido de la cabeza a izquierda y derecha. Todos ellos redundan en el significado del concepto expresado por la palabra “rápidez”, y su función parece ser la de hiperbolizar el significado de “rápido”.

Figura 2. Oscilogramas (franja superior), espectrogramas (franja media), curvas melódicas (líneas negras intensas), curvas de intensidad (líneas negras finas), transcripción fonética (en formato SAMPA (Llisterri et ál. 1993)), patrones acentuales (en formato Sp_ToBI (Prieto & Roseano, 2010) y duración media de la sílaba (*DMS*) de los fragmentos “Es rápido, ¿eh?” (arriba) y “¡Qué sagaz!” (abajo).



En cuanto a los gestos de carácter rítmico (*beat gestures*), estos aparecen alineados con los picos de F0 descritos arriba (situados sobre las palabras “rápido” y “sagaz”). Estos gestos rítmicos adoptan la forma de un ‘fruncido de cejas’ (Figura 3, izda. y centro), de una ‘semioclusión de ojos’ (Figura 3, izda. y centro) y de un ‘asentimiento con la cabeza’ (Figura 3, dcha.).

Figura 3. Instantáneas de la producción de las palabras “rápido” (izquierda y centro) y “sagaz” (derecha).



De todos ellos, es el golpe de cabeza el que se alinea más claramente con el pico de F0 de “sagaz” (en ambas ocasiones), mientras que los otros dos, aunque su momento de mayor intensidad —la cima del gesto— coincide con el pico tonal de “rápido”, permanecen activos en mayor o menor grado durante la mayor parte de la producción del enunciado irónico. Por último, posteriormente a la pronunciación del enunciado, observamos también un leve gesto de mueca con la boca acompañado de una leve sonrisa.

En resumen, los contrastes prosódicos de carácter global entre el enunciado irónico y el no irónico —i.e. una F0 media, una variación

de F0 y una duración media de la sílaba notablemente superiores en el enunciado irónico— parecen ser correlatos del distanciamiento que el hablante muestra sobre el contenido literal del enunciado que está pronunciando: su voz es distinta, se sale de lo habitual, y eso ya actúa como marca que señala el tratamiento especial que el receptor del enunciado va a tener que dispensar al mensaje que está recibiendo. Este distanciamiento general respecto a lo afirmado también lo percibimos en las marcas gestuales que, aun apareciendo en un momento puntual alineadas con un pico de F0, se mantienen a lo largo de la mayor parte del enunciado, como el fruncido de cejas y la semioclusión de ojos. En cuanto a los fenómenos puntuales, observamos cómo los instantes de mayor intensidad —las cimas— de los gestos (fruncido de cejas, semioclusión ocular, batido de cabeza y de manos) aparecen claramente alineados con los picos de F0, y en segmentos en los que se produce una clara ralentización del habla. De manera significativa, esas alineaciones se dan justo sobre aquellas palabras que contienen la mayor carga irónica del enunciado (i.e. “rápido” y “sagaz”), esto es, aquellas que necesitan ser interpretadas en sentido opuesto para alcanzar la interpretación irónica. Además, los acentos tonales que aparecen sobre estas palabras, en especial sobre la última —“sagaz”—, transmiten una información pragmática claramente enfática (p.ej. foco contrastivo u obviedad, en el caso de L+H*), contribuyendo junto al resto de marcas a señalar el especial tratamiento interpretativo por parte del oyente que estas palabras requieren. Finalmente, tras haber señalado prosódica y gestualmente aquellos puntos clave del enunciado que deben interpretarse de manera no literal, el hablante produce un

último gesto —consistente en una mueca con la boca acompañada de una leve sonrisa— con el que sella su principal intención comunicativa al emitir el enunciado irónico: realizar una crítica en clave humorística.

Enunciado 2

“Curiosamente en las ruinas del castillo de Herodes... Unos lince... Unos lince los tíos, ¿eh?”.¹⁵

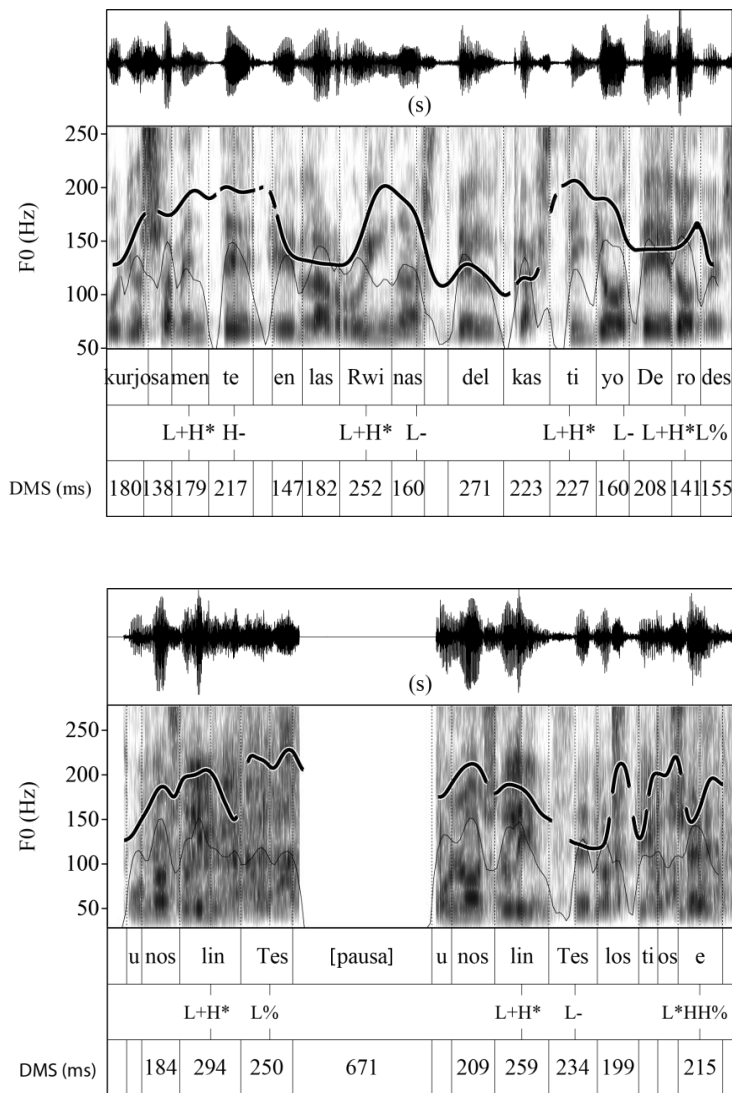
El análisis de los parámetros acústicos de este enunciado irónico respecto al habla no irónica inmediatamente anterior, muestra una elevación de la F0 media de aquel respecto a este (163,02 Hz. frente a 152,02 Hz), así como una mayor variabilidad de la F0 (41,22 frente a 23,73). La amplitud media también es ligeramente más elevada (66,23 dB frente a 67,61 dB), y la duración media de la sílaba se muestra notablemente diferente entre ambos tipos de enunciado (201,3 ms. frente a 164,3 ms.). A nivel local, se observan alargamientos de varios sonidos: el sonido [n] de “curiosamente”, el sonido [r] “ruinas”, y los sonidos [i] y [n] de las dos apariciones de la palabra “lince” (véase Figura 4). En cuanto a los patrones entonativos, el enunciado consta de tres frases entonativas independientes, que, como sucedía en el enunciado anterior, se corresponden con las tres oraciones del enunciado: “Curiosamente, en las ruinas del castillo de Herodes”, “Unos lince” y “Unos lince,

¹⁵ El enunciado no irónico inmediatamente anterior era “*Dicen que unos arqueólogos israelíes han encontrado la tumba del rey Herodes*”, noticia divulgada en mayo de 2007.

los tíos, ¿eh?”, cuyos acentos tonales de carácter enfático (L+H*) recaen sobre la sílaba tónica de las palabras “curiosamente”, “ruinas”, “castillo”, “Herodes” y “lince”. En todas ellas observamos cómo el acento tonal empleado es el típico acento tonal enfático del foco contrastivo (L+H* L%) (véase Figura 4), cuya función pragmática ya hemos reseñado en el análisis del anterior enunciado. Cabe destacar que en la primera oración (i.e. “Curiosamente...”) los tres picos de F0 con acento tonal L+H* (“curiosamente”, “ruinas” y “castillo”) se producen cada vez a una frecuencia más alta, lo cual resulta absolutamente anómalo en una frase enunciativa, a no ser que esta se produzca de manera enfática, como es el caso.

En cuanto a la gestualidad, en primer lugar observamos cómo durante la mayor parte de la pronunciación del enunciado el emisor presenta una ligera inclinación de cabeza hacia su izquierda, así como un leve fruncido de cejas, el cual se intensificará en algunos puntos concretos, pero se mantendrá presente a lo largo de todo el enunciado. En segundo lugar, destacan la aparición de dos tipos de marcas gestuales distintas: unas de carácter rítmico, y una de carácter convencional —que aparece combinada con un gesto rítmico—.

Figura 4. Oscilogramas (franja superior), espectrogramas (franja media), curvas melódicas (líneas negras intensas), curvas de intensidad (líneas negras finas), transcripción fonética (en formato SAMPA (Llisterri et ál. 1993)), patrones acentuales (en formato Sp_ToBI (Prieto y Roseano 2010)) y duración media de la sílaba (DMS) de los fragmentos “Curiosamente, en las ruinas del Castillo de Herodes” (arriba) y “Unos lince... Unos lince los tíos, ¿eh?” (abajo).



De manera análoga a lo que observábamos en el anterior enunciado, las cimas de los gestos rítmicos aparecen sincronizadas con los picos de F0, y algunas de ellas precedidas por un segmento de ralentización del habla. Así, en la primera oración de este segundo enunciado observamos alineaciones entre los picos de F0 situados en las sílabas tónicas de “curiosamente”, “ruinas”, “castillo” y “Herodes” (todos ellos producidos con acentos tonales enfáticos L+H*, véase Figura 4) y las cimas de las siguientes marcas gestuales: batida de ambas manos hacia los lados —en el caso de “curiosamente”— y hacia abajo —en el resto—, fruncido de cejas, y asentimiento con la cabeza (véase Figura 5). En la segunda parte del enunciado (i.e. “Unos lince... Unos lince los tíos, ¿eh?”), los gestos rítmicos que aparecen son: un golpe de cabeza puntual hacia adelante en ambas apariciones de “lince”, un fruncido de cejas y una semioclusión de ojos —solo en la última producción de “lince”—. Además, en la segunda producción de “lince”, el golpe de cabeza se produce con mayor intensidad que en la primera, y este aparece sincronizado con un gesto de batido del brazo hacia adelante, que, además, incorpora en la forma de la mano un signo convencional (el símbolo de “OK”, —véase Figura 6—), al que también acompañan una nueva cima del gesto ‘fruncido de cejas’ unido a una semioclusión de los ojos.

Figura 5. Instantáneas de la producción de gestos rítmicos alineados con los picos de F0 de las palabras con acento léxico de la oración “Curiosamente, en la ruinas del castillo de Herodes” (arriba), junto al espectrograma, curva melódica, curva de intensidad y transcripción fonética (en formato SAMPA (Llisterri et ál. 1993)) de su producción (abajo).

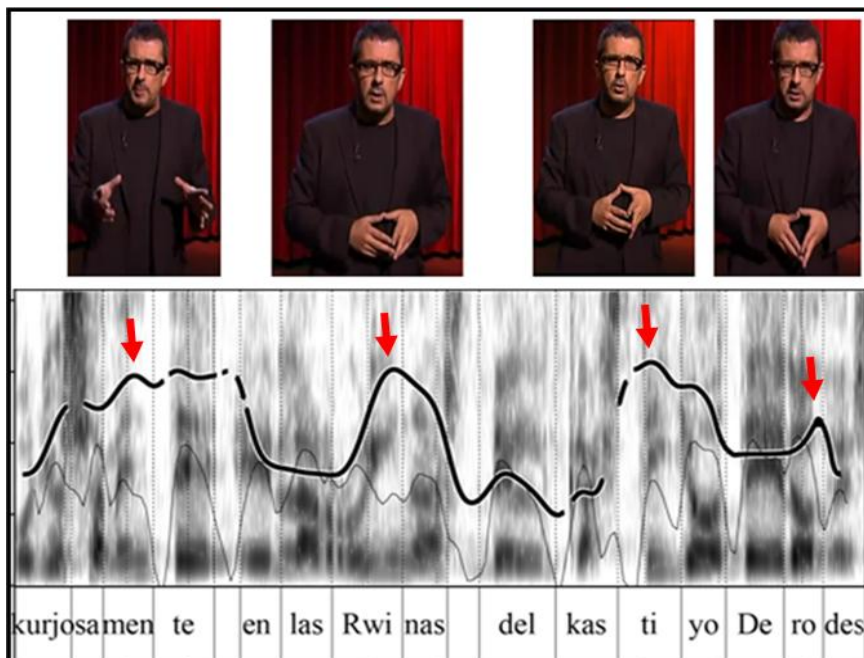


Figura 6. Instantánea de la producción de la marca gestual convencional “OK” producida junto al gesto rítmico del batido del brazo hacia adelante, cuya máxima extensión coincide con el punto de mayor intensidad del fruncido de cejas, de la semioclusión de los ojos y del asentimiento con la cabeza. Todos ellos aparecen alineados con el pico de F0 situado en la sílaba tónica de la palabra “lince” (la instantánea corresponde a la segunda aparición de la palabra).



En resumen, de la misma manera que hemos podido observar en el análisis del enunciado anterior, los contrastes de carácter global existentes entre ambos tipos de enunciado, tanto prosódicos —i.e. una variación de F0 y una duración media de la sílaba notablemente superiores en el enunciado irónico— como gestuales —fruncido de cejas e inclinación lateral de la cabeza activos durante la mayor parte del enunciado—, parecen dar cuenta del distanciamiento general que el hablante muestra hacia el contenido literal del enunciado que está pronunciando: su comportamiento prosódico y gestual no es el mismo que el que se producía en el enunciado anterior. En cuanto a los fenómenos puntuales, también observamos

en este enunciado cómo las cimas de los gestos rítmicos (el fruncido de cejas, la semioclusión ocular y los batidos de cabeza y manos) aparecen claramente alineadas con los picos de F0 y la mayor parte de las veces en segmentos en los que se ralentiza el habla. Si bien en el enunciado anterior esas alineaciones se daban únicamente sobre aquellas palabras que contenían la mayor carga irónica del enunciado (“rápido” y “sagaz”), en este enunciado, además de observar esa misma función enfática de prosodia y gestualidad en las dos apariciones de “lince”, observamos una interacción distinta entre las marcas prosodia/gestualidad y el contenido verbal segmental. Así, apreciamos un incremento paulatino de los tres primeros picos de F0 en la producción de la oración “Curiosamente, en las ruinas del castillo de Herodes” (véase Figura 4) que resulta totalmente anómalo en la producción de una oración enunciativa y que solo se explica si atendemos al contenido semántico del enunciado y a la intención comunicativa del emisor. La intención del emisor es subrayar en tono humorístico lo poco o nada asombroso que resulta el que se haya encontrado la tumba de Herodes en las ruinas del castillo de Herodes. Así, primero presenta la primera parte de la noticia (i.e. que han encontrado la tumba de Herodes) y después, tras el irónico empleo del adverbio “curiosamente” (que aparece convenientemente marcado prosódica y gestualmente, como ya se ha descrito), y a través del empleo anómalo de ese incremento paulatino de los picos de F0 (sincronizando cada uno de ellos con varios y marcados gestos rítmicos), así como de una ralentización general del habla, el hablante genera una tensión discursiva que culmina con la nueva

aparición de la palabra “Herodes”, lo cual deja al descubierto el absurdo que quería señalar el emisor (i.e. que resulte reseñable el que hayan descubierto la tumba de Herodes en el castillo de... Herodes) y tras la que aparece la valoración sarcástica posterior “Unos linces. Unos linces, los tíos, ¿eh?”. Además, en este segundo enunciado observamos un gesto convencional (el gesto de “OK”) con contenido semántico claro (de conformidad o aprobación), cuya función es la de hiperbolizar la valoración positiva que el hablante realiza de la ya de por sí proposición hiperbólica “son unos linces” (en sentido metafórico, que son muy ágiles mentalmente, muy inteligentes) y acentuar así el contraste entre lo manifestado —que “son unos linces”— y lo real —que no lo son— que propicie la interpretación irónica del enunciado por parte del interlocutor y desencadene el efecto humorístico perseguido.

2.4. Discusión y conclusiones

El reciente interés por el estudio de las marcas prosódicas y gestuales que acompañan al habla ha puesto de manifiesto la ineludible necesidad de incluir ambos componentes en cualquier aproximación que se pretenda realizar a los mecanismos que rigen el funcionamiento de la comunicación humana. Así, ciertas investigaciones sobre la utilización de gestos en combinación con el habla sugieren que ambas manifestaciones surgen de una misma estructura conceptual y que conforman un único sistema (McNeill 1992, 2005; Cartmill et ál., 2012). Otras han dejado clara constancia de la contribución decisiva de algunos patrones gestuales y

prosódicos a la detección de diferentes tipos de inferencias pragmáticas (p.ej. Borràs-Comes et ál., 2011; Goldin-Meadow, 2003; Krahmer & Swerts, 2004; Swerts & Krahmer, 2005; Prieto et ál., 2015). En este estudio, hemos presentado una aproximación de carácter meramente exploratorio a la ironía verbal desde la perspectiva de la *prosodia audiovisual*, esto es, a la luz del estudio conjunto de las características prosódicas y gestuales de 21 enunciados irónicos producidos por un humorista profesional en el contexto situacional de un monólogo humorístico.

El análisis cuantitativo ha consistido en la comparación de 4 parámetros acústicos y de 11 marcas gestuales de 21 enunciados irónicos con las de los 21 enunciados no irónicos que los preceden inmediatamente. Del análisis de los resultados hemos extraído dos conclusiones. En primer lugar, estos han mostrado que existen claros contrastes entre el número y la intensidad de las marcas prosódicas y gestuales empleadas en los enunciados irónicos respecto a las empleadas en los no irónicos. Así, observamos que los enunciados irónicos se producen con unos valores de variabilidad de F0 y de duración media de la sílaba (*DMS*) significativamente superiores, así como que 6 de las 11 marcas gestuales estudiadas —mueca con la boca, fruncido y arqueado de cejas, ladeado de cabeza, semioclusión de ojos y risas/sonrisas— aparecían significativamente en mayor número en los enunciados irónicos que en los no irónicos. Estos datos confirman los resultados obtenidos en investigaciones realizadas con anterioridad sobre la prosodia del habla irónica, tanto en producción espontánea, como

controlada (p.ej. Gibbs, 2000; Anolli et ál., 2002; Attardo et ál., 2003, 2013; Caucci & Kreuz, 2012; Laval & Bert-Erboul, 2005; Cheang & Pell, 2008, 2009; Bryant & Fox Tree, 2005; Bryant, 2010; Scharrer et ál., 2011; Padilla, 2011; Rockwell, 2000). El hecho de que, en el presente estudio, los valores de *F0 media* y *Amplitud media* no hayan arrojado significación estadística alguna en la comparación entre enunciado irónicos y no irónicos no invalida nuestro análisis, pues los resultados para los parámetros de amplitud y de F0 se han mostrado discrepantes e incluso contradictorios entre los diferentes estudios realizados hasta la fecha, lo cual puede ser debido tanto a cuestiones relacionadas con la metodología empleada, como con el subtipo de ironía estudiado (o bien, tal y como apuntan Cheang & Pell (2008), a la existencia de diferencias interlingüísticas en la producción del habla irónica). El único fenómeno acústico que documentan todos los estudios como claro correlato de los enunciados irónicos es el de la ralentización del habla (p.ej. Anolli et ál., 2002; Bryant, 2010; Laval & Bert-Erboul, 2005; Padilla, 2011), lo cual puede ser explicado por el esfuerzo que realiza el hablante por acomodarse a las especiales necesidades de procesamiento que requiere la interpretación de los enunciados irónicos (Bryant, 2010, 2011). Si consideramos los resultados de todos estos estudios en su conjunto, y a la luz de los resultados obtenidos en el nuestro, quizá deberíamos decantarnos por afirmar que no existe una manera única e inequívoca de marcar un enunciado irónico, como así concluyen varios estudios recientes (p.ej. Bryant, 2010, 2011, 2012; Padilla, 2011; Attardo et ál., 2011, 2013). Por lo que respecta a las marcas gestuales, los resultados

obtenidos también se encuentran en consonancia con los estudios que se han encargado de estudiar el componente gestual en el habla irónica. Así, todas las expresiones faciales que se han mostrado significativamente relevantes a la hora de marcar enunciados irónicos ya habían sido señaladas como tales anteriormente por, entre otros, Attardo et ál. (2003, 2011), Caucci y Kreuz (2012) y Tabacaru y Lemmens (2014), así como la presencia de risas/sonrisas (Bryant, 2011; Caucci y Kreuz, 2012). Tras confirmar que, efectivamente, enunciados irónicos y no irónicos contrastan tanto en el número como en el modo en el que aparecen las marcas prosódicas y gestuales, la segunda de las conclusiones que extrajimos del análisis cuantitativo de los datos es que ambos tipos de marcas pueden aparecer (1) alineadas (como sucede entre los picos de F0 y el instante de mayor intensidad —la cima— de algunos de los gestos, como el arqueamiento de cejas o el asentimiento con la cabeza) o (2) de manera independiente. Así, se procedió a realizar un segundo y exhaustivo análisis de dos de los enunciados con tal de observar las interacciones que se producían entre los componentes léxico-sintácticos, prosódicos y gestuales durante la producción de dos de los 21 enunciados irónicos.

Los resultados de este segundo análisis han revelado que, efectivamente, se producen contrastes prosódicos y gestuales de carácter global entre enunciados irónicos y no irónicos. En cuanto a los primeros, hemos observado un incremento general de la F0 media, una mayor variabilidad de F0, y un incremento de la duración media de la sílaba; y, respecto a los segundos, aunque

algunas de las marcas gestuales aparecen en un momento puntual alineadas con un pico de F0, muchas de ellas se mantienen presentes a lo largo de la mayor parte del enunciado, como el fruncido de cejas, la semioclusión de ojos o la inclinación lateral de la cabeza. Estas marcas prosódicas y gestuales de carácter global están en consonancia con lo que la bibliografía previa ha señalado acerca de las características prosódicas y gestuales del habla irónica, y cuya función ha sido descrita de manera general como la de facilitar al oyente la interpretación de los enunciados irónicos (p.ej. Gibbs, 2000; Nakassis & Snedeker, 2002; Anolli et ál., 2002; Attardo et ál., 2003, 2013; Caucci & Kreuz, 2012; Laval & Bert-Erboul, 2005; Cheang & Pell, 2008, 2009; Bryant & Fox Tree, 2005; Bryant, 2010; Scharrer et ál., 2011; Padilla, 2011; Rockwell, 2000). Por otro lado, el análisis cualitativo de esos dos enunciados nos ha permitido observar cómo los instantes de mayor intensidad — las cimas— de algunos de los gestos (fruncido de cejas, semioclusión ocular, batido de cabeza y de manos) aparecen claramente alineados con los picos de F0 (producidos, además con acentos tonales de carácter enfático L+H*), y en segmentos en los que se produce una clara ralentización del habla. También hemos podido observar cómo esas marcas prosódico-gestuales interaccionan con el contenido léxico-sintáctico de la oración, bien actuando a modo de índices puntuales —señalando aquellas palabras que contienen la mayor carga irónica del enunciado, como sucede en “rápido”, “sagaz”, “curiosamente” o “lince”—, o bien actuando a modo de señales al servicio de objetivos retóricos o discursivos —como en el caso del incremento paulatino de los

valores de F0 y de la intensidad de las marcas gestuales en la oración “Curiosamente, en las ruinas del castillo... de Herodes”—. Este uso específico y conjunto de prosodia y gestualidad que se muestra tan estrechamente relacionado con la semántica, con la sintaxis, e incluso con la estructura discursiva es precisamente el objeto de estudio de la *prosodia audiovisual*, desde cuyo enfoque ya se ha puesto de manifiesto la importancia que los patrones prosódicos y gestuales revisten en la detección de significados pragmáticos (p.ej. Goldin-Meadow, 2003; Krahmer & Swerts, 2004; Swerts & Krahmer, 2005; Borràs-Comes et ál, 2011; Prieto et ál., 2015), enfoque que, hasta donde alcanzamos a conocer, es en el presente estudio donde se ha empleado por vez primera para examinar el fenómeno de la ironía verbal. Además, las observaciones realizadas en este estudio a la luz de la perspectiva de la *prosodia audiovisual*, encajan con las consideraciones que se han realizado tanto desde el ámbito de estudio de la gestualidad —cuyos principales representantes afirman que los gestos pueden ser entendidos como marcadores o puntualizadores metadiscursivos, reflejando la función pragmática de un enunciado en el discurso o bien proporcionando indicios acerca de la estructura del mismo (Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992, 2005)—, como desde el ámbito de estudio de la prosodia, desde el que se ha propuesto —dentro del marco de la Teoría de la Relevancia— que las modulaciones prosódicas codifican instrucciones procedimentales que guían los procesos inferenciales a través de la reducción del rango de posibles interpretaciones de un enunciado (House, 1990; 2006; Clark & Lindsey, 1990, Fretheim, 2002;

Wilson & Wharton, 2006; Wharton, 2009; Escandell-Vidal, 1998, 2011a, 2011b, y Prieto et ál., 2013). En el caso particular de la ironía verbal, esto significaría que las señales prosódicas y gestuales, al aportar información relevante sobre la intención comunicativa del emisor, actuarían como un índice que guiaría al oyente hacia el reconocimiento de esa actitud distante o irónica que el hablante manifiesta respecto a la proposición expresada, facilitando así el complejo proceso cognitivo que implica la comprensión de los enunciados irónicos (Ruiz Gurillo, 2008). Sin embargo, es importante señalar que, como también se afirma desde la perspectiva de la TR, no todos los recursos procedimentales necesariamente codifican instrucciones específicas de procesamiento. Según hemos podido constatar en el análisis de los resultados, si bien existen patrones prosódicos (frecuentemente alineados con gestos o expresiones faciales) que ostentan un significado procedimental inequívoco (como es el caso del patrón enfático L+H*L%) y que, en el marco de la TR, diríamos que actúan al nivel de las explicaturas de orden superior, otras de las marcas observadas (tanto prosódicas como gestuales) guardan una relación no convencional con el significado que expresan (como el incremento paulatino de la media de F0, el ligero incremento de la intensidad, los batidos laterales con carácter metafórico, los chasquidos de dedos, las marcas gestuales mantenidas,...). A pesar de ello, como se desprende tanto del análisis cuantitativo como del cualitativo de los datos, resulta indudable su contribución a la expresión de los enunciados irónicos, pues, de algún modo, todas esas marcas prosódicas y gestuales de carácter no convencional

también están orientando la interpretación hacia una determinada dirección. A este respecto, Forceville (2014) sugiere que estos últimos elementos, esto es, aquellos elementos que no codifican significados procedimentales, a pesar de que no pueda considerarse, *sensu stricto*, que contribuyan ‘a nivel de las explicaturas’ (pues no son elementos propiamente ‘codificados’), desencadenan procesos cognitivos muy similares a los de los elementos codificados (*‘explicature-like processes’*, los llama), dando así cabida al papel que desempeñan este tipo de marcas en el marco pragmático de la TR. A modo de adenda, y no sin lamentar no poder dedicarle mayor atención, queríamos señalar también que, dada la naturaleza humorística del corpus empleado, los datos recogidos sobre las características prosódicas y gestuales de los enunciados irónicos también encajan con las propuestas que en el marco de la TR se han realizado para explicar el proceso de interpretación de los enunciados emitidos con intención humorística. Según se propone en Yus (2003), con tal de generar una situación humorística, el humorista trata de conducir al oyente hacia una primera interpretación coherente con el *principio de relevancia*, para luego invalidarla conduciéndole hacia una segunda interpretación, si bien menos probable, también correcta. Prosodia y gestualidad tendrían en este tipo de situaciones comunicativas una especial importancia, pues el doble juego de expectativas con el que debe tratar el humorista requiere de un uso más complejo de las marcas con las que pretende orientar —y reorientar— la interpretación que desea obtener del oyente. Este dibujo del acto comunicativo humorístico encuentra un claro reflejo en los datos obtenidos en el presente

estudio, pues explica tanto la proliferación de marcas “orientadoras” hacia una determinada interpretación en los enunciados irónicos, como también —y especialmente— la mayor aparición observada en los datos de elementos gestuales posteriores a la emisión del enunciado (66% en enunciados irónicos frente a 28% en los no irónicos), los cuales claramente desempeñarían la función de reorientar la interpretación del oyente hacia el ámbito de lo irónico/humorístico.

En conclusión, los resultados de este estudio muestran que, efectivamente, los hablantes emplean modulaciones prosódicas y gestuales para señalar la presencia de un enunciado irónico, pero que no lo hacen de un modo único, sino que se valen de diferentes recursos en los que se ven implicados elementos prosódicos, gestuales y léxico-sintácticos. Las observaciones realizadas sobre la interacción entre estos tres componentes abren un camino para que futuras investigaciones aborden la tarea de esclarecer la naturaleza concreta de esa interacción, no solo en el ámbito del habla irónica, sino en el de la comunicación en general. Como han mostrado algunos estudios experimentales recientes, y como explican las teorías pragmáticas de orientación cognitiva como la TR, los elementos prosódicos y gestuales contribuyen fehacientemente al proceso de interpretación de los enunciados, y su aportación es especialmente relevante en aquellas situaciones comunicativas que presentan una mayor complejidad dialéctica entre los diferentes elementos que intervienen en un acto comunicativo, uno de cuyos casos paradigmáticos es el de la ironía verbal. Sería deseable que

futuras investigaciones, de mayor calado y extensión que la presente, encararan la tarea de realizar un inventario exhaustivo de las categorías y distinciones susceptibles de ser expresadas por medios prosódicos y gestuales, tanto de manera conjunta como por separado. La existencia de un catálogo de esta naturaleza permitiría, por un lado, avanzar en el preciso establecimiento del papel que prosodia y gestualidad desempeñan en la producción, interpretación y procesamiento cognitivo de los enunciados irónicos y, por otro, refinar los modelos pragmáticos en pos de una más ajustada explicación sobre los diferentes tipos de significados que estos dos componentes son susceptibles de codificar.

González-Fuente S, Escandell-Vidal V, and Prieto P.
[Gestural codas pave the way to the understanding of verbal irony](#)". Journal of Pragmatics. 2015;90:26-47.

3. CHAPTER 3: “Gestural codas pave the way to verbal irony understanding”

3.1. 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!”) (Curc6, 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 rhetorics¹⁶, 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

¹⁶ See e.g. Quintilian’s *Institutio Oratoria*.

really meant. Similarly, conventional/logical approaches to verbal irony (e.g. 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, 2011). 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 (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 with a set of markers or cues (Attardo et al. 2003, 2013; Bryant & Fox Tree, 2005). 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 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., 2013, 2015; Kraemer & Swerts, 2004; Swerts & Kraemer, 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 aversion is used by speakers when producing sarcastic utterances. Williams et al.’s (2009) experiments found that speakers averted 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.

3.2. Experiment 1

3.2.1. Methods

a) 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.

b) 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 7) 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 2). The video sequences

and sentences were selected in order to prompt incongruent contextual situations that would lead to spontaneous ironic responses (see 2.1.3. 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 7).

Figure 7. Still images of Video A (left panel) and Video B (right panel).



The eight 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 2 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 2. 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’

c) 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 8, 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.

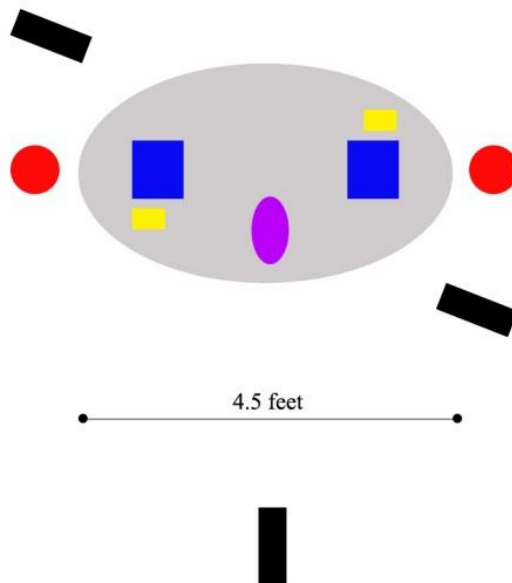
d) Data coding

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)¹⁷ were also

¹⁷ One of our reviewers rightly points out that it is quite possible that speakers about to produce an ironical utterance may start producing ironical cues in the utterances preceding the ironical one. Although we considered the possibility of choosing baseline sentences which were not preceding the target ironic sentences, we finally decided to use the immediately preceding sentences for two main reasons: (1) we follow Bryant’s methodology for comparing ironic and non-ironic sentences, thus making our results directly comparable with previous studies (Bryant, 2010), and (2) our results on the specific correlates of target ironic sentences are stronger if we find clear differences in the ironical cues between the

extracted (as in Bryant, 2010). 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.”

Figure 8. 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.



ironic utterances and the utterances preceding them (which might reflect some less strong cueing).

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).¹⁸ 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 9. Also, the prosodic characteristics of the target utterances were coded using Praat (Boersma & Weenink 2008) and automatically imported into ELAN.

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’).

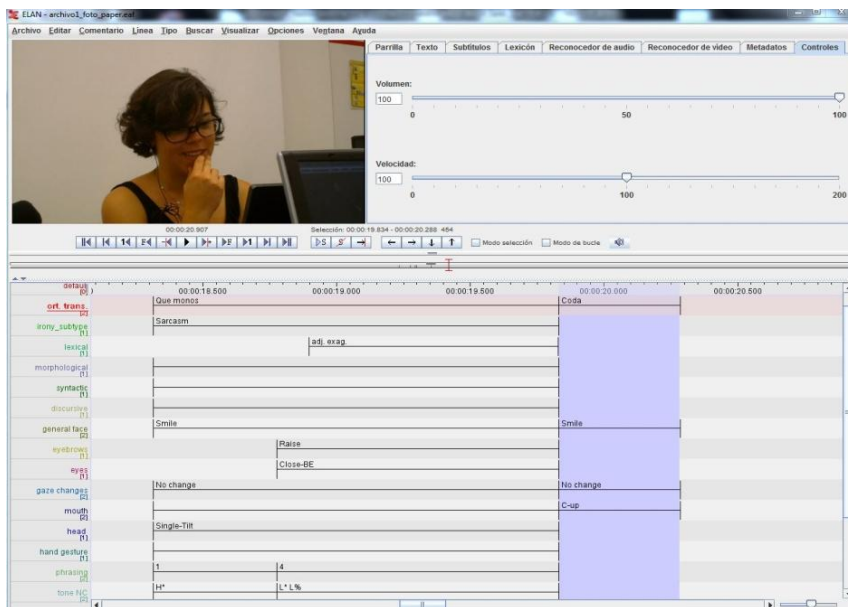
Coding of irony subtypes (tier 2). We followed Gibbs’ (2000)¹⁹ proposal and labelled the following five irony subtypes: ‘sarcasm’,

¹⁸ ELAN is an open source tool used for annotating and aligning transcriptions with video data.

¹⁹ We are aware that Gibbs’ (2000) irony subtype classification has been criticised by Wilson (2013) in the light of experimental work on the development of irony comprehension. She argues that some discursive phenomena as hyperbole, jocularly, understatement and rhetorical questions (which have been generally treated as forms of irony), “display none of the distinctive features of irony in most of their uses” (Wilson 2013: 40). Even though we agree with Wilson that more theoretical accounts and experimental paradigms are needed to clarify what can be considered an ironic remark, we adopted Gibb's classification

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.

Figure 9. Example of labelling with the target ironic sentence “*Que monos!*” (“How cute!”).



as a useful labelling system that makes our results directly comparable with previous studies (e.g., Bryant ,2010).

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 & Leonetti (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).


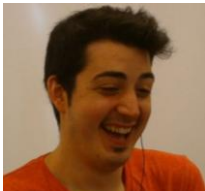






Visual coding (tiers 7-13). Following Allwood et al.’s (2007) gestures coding proposal and McNeill (1992), the following gestural cues produced during and after the utterance of sentences were annotated:

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 4 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 4 show picture stills of the facial and body gestures that were annotated most frequently in the corpus.

Table 4. Examples of facial and body gestures.

<i>Tier</i>	<i>Labelling examples</i>		
<i>General face</i>			
	‘Smile’	‘Laugh’	‘Scowl’
<i>Eyebrow movements</i>			
	‘Raise’	‘Frown’	
<i>Eyes</i>			
	‘Close both’	‘Squint’	‘Exaggerated Opening’

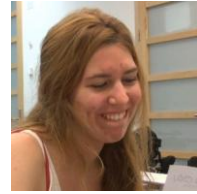
Gaze changes



‘Towards
interlocutor’



‘Gaze
Aversion’²⁰

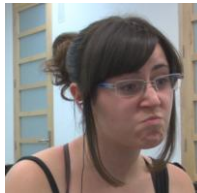


‘Towards
materials’

Mouth



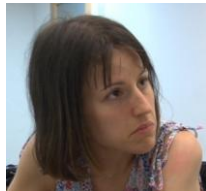
‘Stretched’



‘Protruded’



‘Corners-Up’

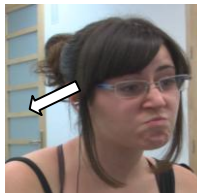


‘Corners-
Down’

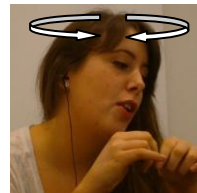
Head



‘Nod’



‘Tilt’



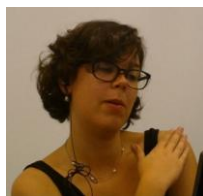
‘Shake’

²⁰ By “gaze aversion” we refer to some kind of brief and transitory shifting of gaze away from the interlocutor.

*Hand
gestures*



‘Beat’



‘Other’²¹

Prosodic coding (tiers 14 - 21).

Phrasing (tier 14). Following the Cat_ToBI proposal (Prieto 2014)²², the following break indices (i.e. the level of boundary strength of prosodic groups) were annotated: prosodic groups 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 with accented syllables) were labelled.

Voice quality (tier 16). In this tier voice quality features (labelled ‘Creaky’, ‘Falsetto’ or ‘Breathy’) were perceptually annotated and

²¹ The ‘Other’ value includes metaphorical, deictic and iconic gestures (McNeill 1992).

²² The Cat_ToBI proposal consists on a description of the prosodic and intonational structure of Catalan within the Autosegmental-Metrical (AM) framework (Pierrehumbert, 1980; Gussenhoven, 2004, among others). This proposal includes an analysis of the phonetic realizations and distributional properties of the phonological intonational patterns found in Catalan, as well as the description of the intonational realization of different pragmatic meanings.

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.

e) 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 2.1.4). 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.

3.2.2. 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.²³

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.

a) Irony subtypes

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

²³ The distribution of ironic productions during the experimental session was as follows: 72% of the ironic 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).

b) 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*”²⁴ [“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.

²⁴ ‘Colocadillos’ is a Spanish word, not Catalan, in this example of code-mixing.

c) 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 and pragmatic meanings (such as L+H*L%, L*HL%, L*!H%, or L+H*L!H%)²⁵. 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

Table 5 shows the mean and standard deviation values of the four acoustic measures (namely, F0 mean and F0 variability, intensity

²⁵ In the Cat_ToBI proposal (Prieto 2014), the nuclear configuration L+H* L% is described to appear in narrow focus statements, exclamatives and imperatives contexts; L* HL% is described to appear in obviousness statements; L* !H% in disapproval statements; and L+H* L!H% in emphatic obviousness statements.

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 5. 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.

Acoustic dimension	Baseline utterances		Ironic utterances	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
F0 (st.)	90,14	3,06	90,76	3,40
F0 variability (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 ($\chi^2(1) = 8.05$, $p < 0.05$).

d) 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

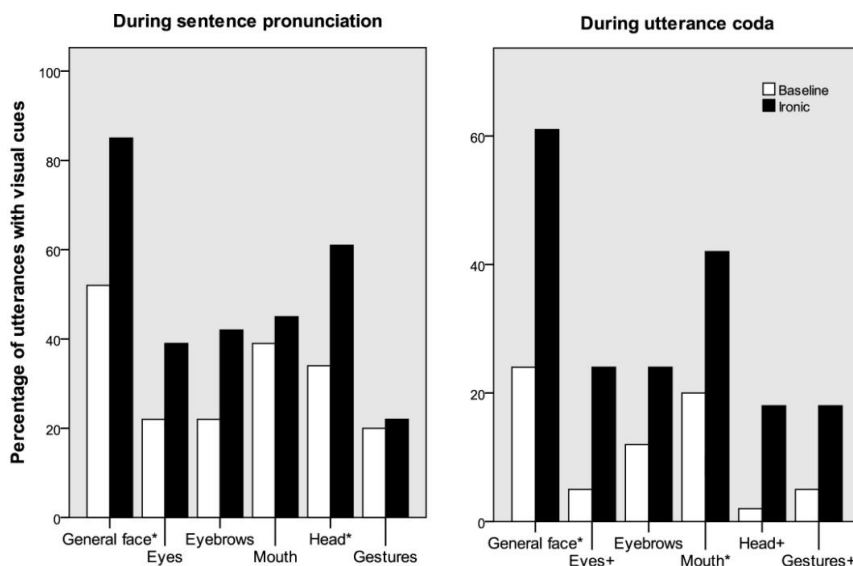
Figure 10 shows the percentage of baseline and ironic target utterances in which ‘General Face’, ‘Eyes’, ‘Eyebrows’, ‘Mouth’, ‘Head’ and ‘Gestures’ values differ from ‘None’ both during the utterance (left panel) and after, i.e. during the coda (right panel). These results show that ironic utterances display higher rates of all gestural cues under analysis than baseline utterances. The results of a set of chi-square tests testing the effect of utterance type (baseline vs. ironic utterances) on all the target visual cues showed that the difference was only significant in the case of ‘General Face’ and ‘Head’ for the non-coda condition and ‘General Face’ and ‘Mouth’ for the coda condition, as can be seen in Figure 10. 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%).

Gaze

Figure 11 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).²⁶

²⁶ This difference between utterances’ type related to gaze patterns has been found to be significant ($\chi^2(1) = 6.08$, $p < 0.01$).

Figure 10. 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).

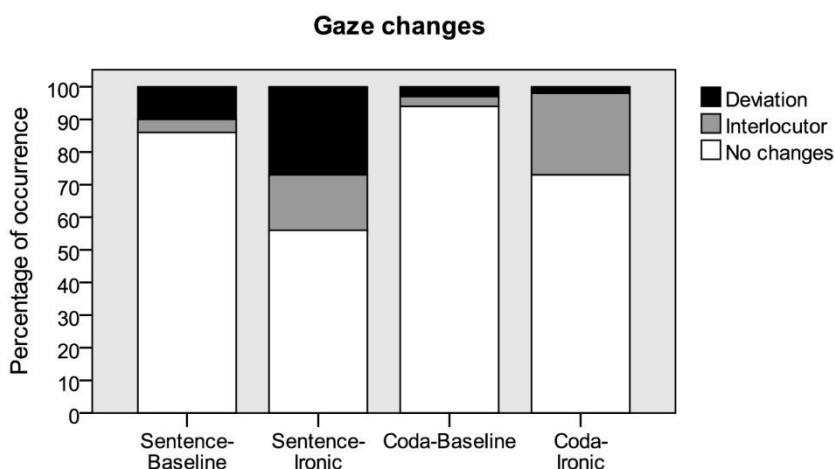


Note. In Figure 10, ‘*’ indicates that $p < 0.05$ and ‘+’ indicates that chi-square test not performed because the expected frequency was less than 5 in more than 20% of the cells.

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 aversion”—black-shaded part of columns). These results are in agreement with those described by

Williams et al. (2009), who concluded that speakers tended to avert their gaze from the interlocutor when being ironic.

Figure 11. Percentages of utterances with ‘No changes’ (white part of columns); gaze change ‘Towards interlocutor’ (grey-shaded part of columns) and ‘Gaze aversion’ (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).



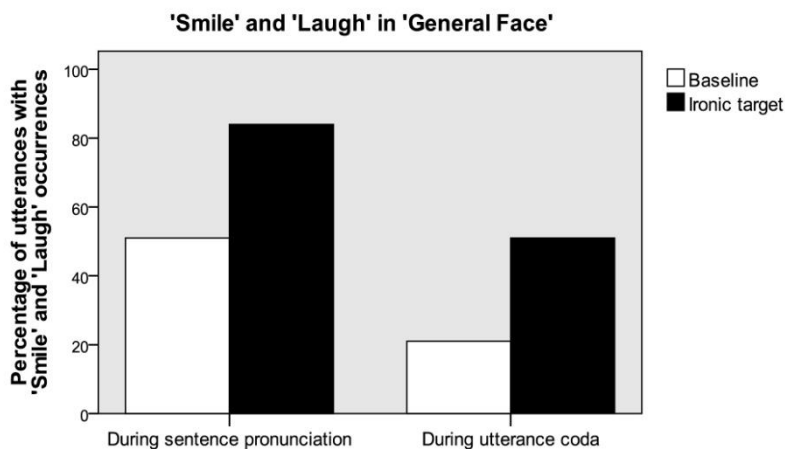
The same pattern can be observed during the production of gestural codas, that is, speakers change their gaze pattern more frequently during ironic utterance codas (27% out of the 70% ironic sentences containing a gestural coda) than during baseline utterance codas (6% out of the 27% of baseline utterances 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, something which is consistent with the ‘information-seeking’ function of gaze proposed by Argyle and Cook (1976), among others.

‘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 12 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 12. 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).



e) Summary of lexico-syntactic, prosodic and visual cues 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.

3.3. 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.

3.3.1. Methods

a) 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).

b) 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 3 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 (2a) and (2b)).

Example 3. 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 720×576 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 × 4 discourse contexts × 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 13 for a set of examples).

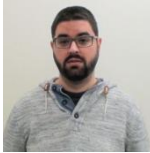



Table 6 shows the percentages of gesture types occurring in ironic and non-ironic utterances broken down by moment of occurrence (i.e. during sentence utterance or during post-utterance codas). First, it will be noted that a different range of gesture types appears in ironic utterances relative to non-ironic utterances. In non-ironic utterances, the most common visual cue was head nodding (83%),

Figure 13. 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!”)

Before utterance	During sentence		During coda (1 sec.)
			
	Raised eyebrows	Head nod Shoulders shrug	Head nod Sustained gaze

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

Before utterance	During sentence		During coda (1 sec.)
			
	Raised eyebrows Shoulders shrug	Head nod	Head tilt Averted gaze Stretched mouth

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.²⁷ 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 6. 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.)
	Smile (corners-up mouth)	16%	16%
Ironic utterances	Stretched mouth	8%	33%
	Raised eyebrows	25%	66%
	Squinted eyes	17%	8%

²⁷ 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.

	Head shake/tilt	17%	50%
	Head nod	8%	0%
	Shoulder shrug	17%	8%
	Smile (corners-up mouth)	33%	41%
	Stretched mouth	0%	0%
Non- ironic utterances	Raised eyebrows	25%	8%
	Squinted eyes	0%	0%
	Head shake/tilt	0%	0%
	Head nod	83%	50%
	Shoulder shrug	33%	0%

Table 7 shows the percentages of occurrence of sustained gaze vs. averted 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 aversions during the sentence utterance (33%) but only in one case during the coda (8%).

Table 7. Percentage of occurrence of sustained gaze vs. averted gaze in ironic and non-ironic utterances, 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%
	Averted gaze	33%	8%
Non-ironic utterances	Sustained gaze	100%	100%
	Averted 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 aversion characterizing ironic productions. In the case of ironic gestural codas, the gaze behaviour that characterizes them is sustained gaze.

c) 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 3.3.1.b) 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 4). 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 4. Ambiguous discourse context.

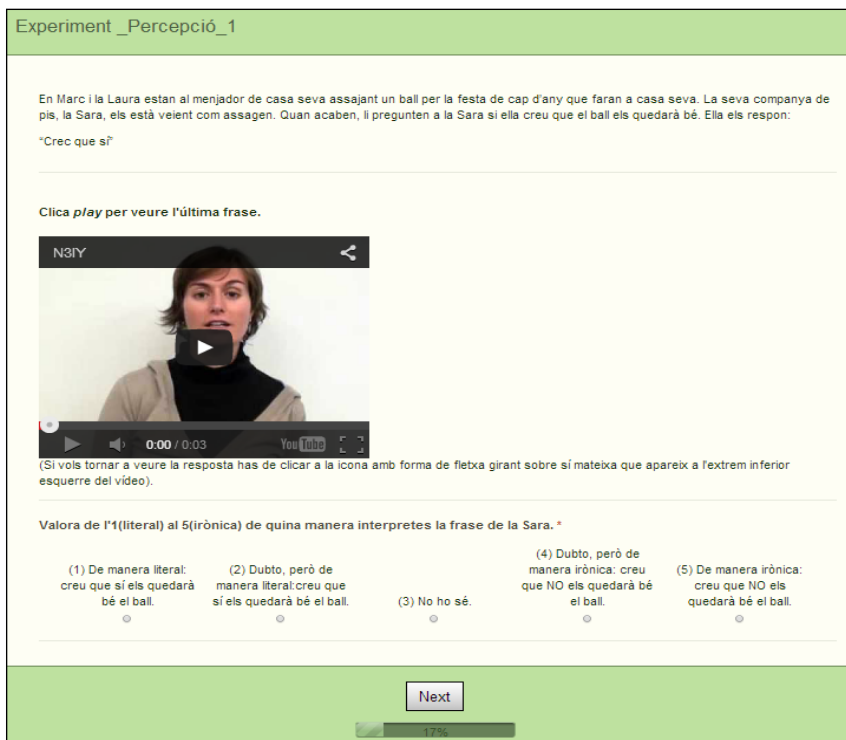
John and Peter are roommates and are having breakfast in the kitchen, which is an interior room of the house with no windows.

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 14).

Figure 14. 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).

d) Procedure

Participants completed one of the two versions of the two online audiovisual questionnaires. 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) \times 24 questions). The mean duration of this experiment per participant was 16 minutes.

e) Measures and statistical analyses

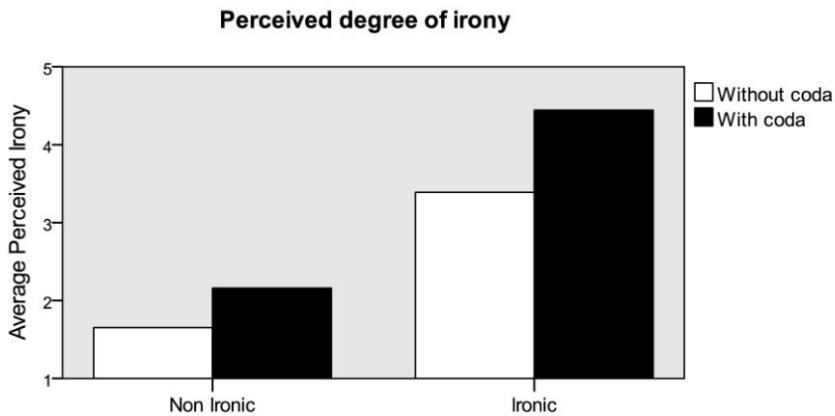
The 576 responses were analyzed with a Generalized Linear Mixed Model (GLMM) using IBM SPSS Statistics 23.0 (IBM Corporation 2015), with 'Perceived degree of irony' as a dependent variable. The fixed factors were 'Utterance performance' (2 levels: 'Non-ironic' intonation/gesture vs. 'Ironic' intonation/gesture), 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.

3.3.2. 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’ \times ‘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 15 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 had the effect of increasing the irony detection.

Figure 15. 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.

3.4. 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., 2015; 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, that 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 averted 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 averted gaze with no specific destination (i.e. fleeting aversions) 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 were 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 et al.’s (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 & Rossano (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, 1991; Stivers & Rossano, 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., 2015). 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.

4. CHAPTER 4: “Communicating irony: when gesture cues are more powerful than prosodic and contextual cues”

4.1. Introduction

In daily conversations, speakers are surrounded by all sorts of information that they have to process in online comprehension. If an utterance is produced with an ironic intent, listeners are expected to recognize that it is not literally true and infer the real intention of the speaker. For that purpose, they need to be able to detect a potential contradiction between the face value semantic meaning of the proposition and all sorts of information coming from a variety of sources, including the discourse context (e.g., the specific situation, shared beliefs between speaker and listener) as well as the manner in which the speaker performs the utterance (i.e., prosodic and gestural modulations).

Consider the situation in Example 5, with two potential follow-up utterances.

Example 5. Laura and Julia live on the same street and are about the same age. They know each other only by sight. Today they have met by chance at the theater. Having greeted each other, they are now waiting for the show to start, seated side by side. Before the play starts, a theater employee announces over the PA system

that unfortunately the performance has to be canceled because the leading actress has lost her voice. Laura turns to Julia and says:

(a) Oh, shit!

(b) Oh, great!

In (1a) Laura is making a negative evaluation of a negative situation (the cancellation announcement) by uttering a negative comment (*Oh, shit!*). By contrast, in (1b) she is uttering a positive comment (*Oh, great!*). Whereas in the first case the interpretation easily arises from detecting the match between the negative proposition and the preceding negative discourse context, in (1b) the ironic interpretation of *Oh, great!* will be achieved when Julia perceives the mismatch between the positive valence of the proposition and the negative valence of the discourse context. However, the utterance may have also been accompanied by multimodal negative cues such as a sad tone of voice and disapproving gestures such as head shaking/tilting, stretched mouth, rolling eyes, etc. Presumably this set of multimodal cues helps a listener to infer information about the attitudinal and emotional states of their interlocutor and thus plays a role in their ability to read ironic intent. But exactly how important is the role of these multimodal modulations? Do gestural cues carry greater weight relative to prosodic cues in the detection of irony, and how do they interact with contextual cues?

This study deals with the role of prosodic and gestural cues in combination with contextual cues in the process of verbal irony perception. By means of three experimental studies, we will

examine the relative contribution of these three elements to the detection of verbal irony, more specifically to the detection of sarcasm. While verbal irony in general is understood as a form of non-literal language in which a speaker produces an utterance whose unstated meaning is at variance with or even opposite to its literal verbal content (Bryant, 2012), many authors consider sarcasm a subtype of verbal irony that is characterized by the utterance of a sentence bearing ostensibly positive content but which implicitly expresses a negative or critical attitude towards an event or person (Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995; Cheang & Pell, 2008). Presumably, to achieve a successful understanding of a speaker's sarcastic intent, listeners can rely on the contrast not only between the positive verbal content and the negative discourse context but also the negative emotion or attitude conveyed by prosodic and gestural cues (e.g., Voyer et al., 2016).

Research on verbal irony understanding has mainly focused on examining the role of contextual cues. Most accounts agree that the contextual characteristics of the verbal exchange such as the place where it is occurring and the set of beliefs shared between speaker and listener play a key role in its interpretation (Kreuz & Glucksberg, 1989; Gibbs, 1994; Kumon-Nakamura et al., 1995; Utsumi, 2000; and others). Specific experimental results have emphasized the role of contextual contrast effects in sarcasm perception (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003). For example, Ivanko and Pexman (2003) performed several experiments investigating the

role of discourse context (and specifically the degree of incongruity between the discourse context and the potential response) in the interpretation of literal and ironic statements in English. A set of 89 listeners rated 12 sentences such as *Brad is a wonderful singer* which were preceded by discourse contexts with different degrees of situational negativity (i.e., bias towards an ironic interpretation of the statement) using strongly negative, weakly negative, and neutral discourse contexts. The results showed that in strongly ironic-biased situations the sarcastic statements were perceived to be more mocking than literal statements, whereas in weakly negative situations, the same sarcastic statements were perceived to be only slightly more mocking than literal statements. The authors concluded that more extreme contrasts between context and propositional content (and specifically the degree of negativity of the discourse context) facilitate sarcasm detection.

Concurrently to with analyzing the role of discourse context in verbal irony interpretation, several studies have investigated the role of prosody in the expression and recognition of verbal irony (e.g., Rockwell, 2000; Attardo, Eisterhold, Hay & Poggi, 2003; Attardo, Pickering & Baker, 2011; Padilla, 2004, 2011; Bryant & Fox Tree, 2005, 2010; Loevenbruck et al., 2013; González-Fuente, Escandell-Vidal & Prieto, 2015; González-Fuente et al., 2016). One of the main underlying questions of this line of research has been to assess whether we can rely on the concept of an ‘ironic tone of voice’. Across studies, acoustic analyses of ironic productions have revealed that ironic utterances are characterized in different languages by longer syllable durations, stronger pitch range

modulations, and the use of specific intonational contours (see e.g. Rockwell, 2000; Attardo et al., 2003; Cheang & Pell, 2009; and Bryant, 2010, for English; Anolli, et al., 2002, for Italian; Cheang & Pell, 2009, for Cantonese; Laval & Bert-Erboul, 2005; Loevenbruck et al., 2013; and González-Fuente et al., 2016, for French; Scharrer et al., 2011, for German; Padilla, 2004, 2011, for Spanish; and González-Fuente et al., 2015, for Catalan). While duration cues seem to be consistent across studies in all languages, since in most cases duration slows down in ironic speech, other prosodic cues like average pitch, pitch variability, and intensity patterns do not show a consistent pattern across studies or languages (see Bryant, 2011, for a discussion). This lack of consistency may be due to methodological issues such as differences in the irony subtype under analysis, the language-specific implementation of irony, and also the intonational phonology of each language, which may privilege either rising or falling pitch accents (Loevenbruck et al., 2013). In this regard, as far as we know, González-Fuente et al. (2016) is the only empirical study investigating the extent to which specific nuclear tonal configurations (together with pitch range expansion and syllable lengthening features) influence irony interpretation. Interestingly, the results showed that duration and intonation contour choice were more powerful than pitch range modulations for irony detection in French. In sum, as contended by Bryant and Fox Tree (2005), it seems clear that the notion of a crosslinguistic “ironic tone of voice” is oversimplified and misguided. In addition to the issues highlighted above, we argue that the notion of pragmatic contrast is essential to explain the great

amount of variation attested to in attempts to characterize ironic prosody. Importantly, prosodic features expressing either positive or negative emotion may be interpreted as ironic given a particular context.

Despite the failure to find a consistent “ironic tone of voice” across production studies, perception studies have shown that ironic intent can be successfully extracted from prosodic cues even in the absence of contextual cues. For example, Loevenbruck et al. (2013) found that French-speaking subjects could accurately distinguish ironic from sincere statements on average 79% of the time despite the lack of a prior discourse context.²⁸ Padilla (2011) found that 92% of the time Spanish listeners could successfully identify the ironic utterances in a set of excerpts from a corpus of spontaneous speech. However, in this case the utterances were presented with the preceding context. Interestingly, when participants were asked whether context or tone of voice had been more helpful in detecting irony 2% responded ‘context’, 48% ‘tone of voice’, and 50% considered both cues to have been equally important for their decision.

From a theoretical point of view, the Relevance-Theory pragmatic account attributes the importance of prosodic modulations for verbal irony understanding to the fact that they act as facilitators of inferential processes (e.g., Sperber & Wilson, 1986/1995; House,

²⁸ Acoustic analysis of results obtained in the production experiment used to create stimulus utterances for the identification task showed that sarcastic comments were consistently produced with significantly higher pitch level, wider range, and longer durations than sincere comments.

1990; 2006, Clark & Lyndsey, 1990; Fretheim, 2002; Wilson & Wharton, 2006; Escandell-Vidal, 1998, 2011a, 2011b). Going a step further, one author within Relevance Theory research has argued that a fully explanatory account of verbal irony must include a listener's ability to interpret the speaker's feelings and emotions, as the recognition of the affective attitude of the speaker "may not only influence the eventual choice of an interpretation, but also the very ascription of irony as utterly offensive, mildly offensive, praising or humorous" (Yus, 2016: 93).

The interplay between context and prosody

To our knowledge, only three studies have investigated the interplay between discourse context and prosodic cues in the perception of irony, two of them investigating sarcasm (Woodland & Voyer, 2011, and Voyer et al., 2016), and one of them investigating ironic compliments (Voyer & Vu, 2016). Woodland and Voyer (2011) examined how contrasts between discourse context and tone of voice affected the perception of sarcasm. A total of 82 English listeners were presented twice with a set of short written discourse contexts in two separate conditions, namely ironic-biased (N = 12) and literal-biased discourse contexts (N = 12). These discourse contexts were followed by a set of follow-up utterances which were auditorily presented in either an intended sarcastic (N = 12) or an intended sincere (N = 12) tone of voice. The 48 stimuli were composed of either congruent combinations of discourse context and utterance prosodic performance (12 ironic-biased discourse contexts with 12 ironic performances of the reaction comment and 12 literal biased contexts with 12 sincere performances of the

reaction comment, for a total of 24 congruent matches) and incongruent combinations of those features (12 ironic biased discourse contexts with 12 sincere performances and 12 non-ironic biased contexts with 12 sincere performances, for a total of 24 incongruent matches). After listening to the auditory stimuli, subjects were asked to rate the perceived degree of ‘sarcastic irony’ (sic) by means of a 5-point Likert-type scale (1 = ‘very sincere’ to 5 = ‘very sarcastic’). Results showed that mid-range ratings of perceived degree of irony and longer reaction times were obtained when tone and context were incongruent (i.e., ironic-biased context and sincere tone or vice versa) compared to when they were congruent (i.e., literal-biased context and sincere tone or ironic-biased context and sarcastic tone). Thus, producing a positive statement (e.g., *Well done!*) with a sarcastic tone of voice may serve to exaggerate the contrast between the verbal message and the discourse context, leading to an increase in the likelihood that the utterance will be rated as sarcastic. These results clearly point to the relevance of contrasting contextual and prosodic cues in the verbal irony recognition process. That is, the greater the mismatch between contextual and prosodic cues, the greater the perception of irony.

In a follow-up of Woodland & Voyer (2011), Voyer et al. (2016) ran a set of perception experiments that introduced two main novelties: (1) both discourse contexts and follow-up utterances were presented as auditory stimuli, and (2) ambiguous (i.e., non-biased) discourse contexts were used to determine whether a milder contrast between context and propositional information would affect the perception of sarcasm on the one hand and reaction times on the

other. As in Woodland & Voyer (2011), in Experiment 1 participants were presented with congruent vs. incongruent context/tone of voice pairs. The results showed that congruent combinations induced faster reaction times and more accurate ratings of the perceived degree of irony compared to incongruent combinations, which produced mid-range accuracy ratings for the degree of irony and slower reaction times. In Experiment 2, when participants were presented with an ambiguous context paired with either a sarcastic or sincere tone of voice, the results were as predicted, with ratings for perceived degree of irony somewhere intermediate between those obtained in the congruent and the incongruent pairings used in Experiment 1. This is, when the negative or positive valence of the discourse context was neutralized, participants' ratings for perceived degree of irony were strongly influenced in the direction of either the sincere or sarcastic tone of voice, but never to the extent induced by the congruent pairings in Experiment 1. Thus the two studies clearly showed that (a) congruent context/tone of voice pairs facilitate the processing of sarcastic remarks; and (b) incongruence between tone of voice and discourse context creates a frustrated expectation which in turn leads to increased difficulty in interpreting the utterance. Furthermore, Voyer and Vu (2016) showed that context and tone of voice interactions also affect the interpretation of ironic compliments, by including in their stimulus materials statements with a negative valence conveying the speaker's praising intent (e.g., *God, you're terrible!* meaning *Well done!*), thus extending their findings to irony subtypes other than sarcasm.

However, though these studies clearly show the interplay between the discourse context and tone of voice, so far no experimental research has included audiovisual materials. Most past research on irony detection has relied on either written or auditory materials for the detection of irony, and little is known about the role of visual information encoded in gesture. Therefore, the present study introduces a novel aspect to the literature of irony perception by investigating the role of audiovisual cues in the communication of a speaker's intentions and the relative importance of this visual information relative to other cues.

The role of gestures

There is growing evidence that gestures play an important role in pragmatic comprehension (e.g., Goldin-Meadow, 2003; Krahmer & Swerts, 2004; Swerts & Krahmer, 2005; Holler & Wilkin, 2009; Borràs-Comes, et al., 2011; Prieto, et al., 2015). Some pragmatic theoretical accounts such as Relevance Theory have signaled the need to include non-verbal features (e.g., facial expressions) in pragmatic comprehension accounts (Wilson & Wharton, 2006; Wharton, 2009). Despite this, very little empirical research to assess the role of gesture in irony perception has been carried out.

Researchers have noted that irony can be communicated by a variety of gestural and facial cues, such as head movements, smiles and laughter, eye gaze aversion, and mouth and eyebrow configurations (e.g., Haiman, 1998; Attardo et al., 2003, 2011; Padilla, 2004; Hancock, 2004; Poggi, 2007; Williams et al., 2009; Bryant, 2011; Caucci & Kreuz, 2012; Tabacaru & Lemmens, 2014;

González-Fuente et al., 2015). For example, Williams et al. (2009) found that speakers averted their gaze when being sarcastic in conversations with an unknown interlocutor. More recently, Caucci and Kreuz (2012) found that one of the largest differences in facial cues produced by English-speakers in a set of 66 sarcastic or sincere utterances was the greater amount of smiling that occurred in sarcastic utterances. Importantly for the purposes of this paper, they suggest that smiling “could be used in addition to or in place of other cues (e.g., changes in tone of voice) to let addressees know an utterance is to be interpreted sarcastically” (Caucci & Kreuz, 2012: 11). The role of gestures in signaling ironic intent has also been recently assessed in González-Fuente et al. (2015), which provided empirical evidence for the facilitating effect of post-utterance gestural cues on verbal irony appreciation.

All in all, to our knowledge no studies have addressed the relative contribution of prosodic and gestural markers in the detection of irony and how they interact between them and with contextual cues. The present study constitutes an attempt to fill this gap by using stimulus materials that are not merely written or auditory, but also audiovisual, which allowed us to add the gestural channel of communication. Three separate experiments were involved. In Experiment 1 we compared the perception of irony in audiovisual vs. audio-only information; in Experiment 2 we tested the effect of congruent and incongruent combinations of gestural and prosodic cues in the same task; and in Experiment 3 we tested the effects of congruent and incongruent combinations of discourse context cues (literal-biased vs. ironic-biased) with multimodal cues (sincere vs.

ironic). The chapter is organised as follows. Sections, 4.2, 4.3, and 4.4 present the methods and results for Experiments 1, 2 and 3, respectively. Finally, Section 4.5 discusses the findings and summarizes the most significant conclusions of the study.

4.2. Experiment 1

The main goal of Experiment 1 was to test the perception of irony in responses to a neutral prompt as conveyed audiovisually and compare this with the perception of irony in the same stimuli conveyed through auditory channels only.

4.2.1. Methods

a) Participants

A total of 52 Catalan-speakers completed the irony detection task. However, the responses of 7 speakers were excluded from subsequent analysis because they reported using Catalan (vs. Spanish) less than 50% of the time in their daily, and since the questionnaire was written in Catalan it was felt that subjects needed to be at least Catalan-dominant bilinguals. Thus the responses submitted to analysis come from the remaining 45 participants (29 women and 16 men; mean age = 32.41, stdev = 12.72), who reported using Catalan on average 84% of the time on a daily basis (stdev = 7.23).

b) Materials

Preliminary study: Discourse Completion Task

In order to obtain the audiovisual materials for all three experiments eight Catalan native speakers participated in a Discourse Completion Task (henceforth DCT; Blum-Kulka, 1989; Billmyer & Farghese, 2000; Félix-Brasdefer, 2010). The DCT methodology consists of a semi-spontaneous elicitation task in which a given situational prompt is presented to the speaker, who is asked to produce a given follow-up sentence in a way that seems to accord with the stipulated context. The DCT was designed to obtain 4 ironic utterances and 4 sincere utterances. A set of 4 discourse contexts were created by the researchers, each one divided into 2 conditions, namely the literal-biased condition (i.e., a positive context), which was intended to trigger a literal interpretation of the positive follow-up sentence, and the ironic-biased condition (i.e., a negative context), intended to trigger a sarcastic interpretation. Example 6 below shows the English translation of one of the discourse contexts used with the two alternative contextual paths (e.g., literal-biased vs. ironic-biased) eliciting the same follow-up utterance.

Example 6. English translation of a discourse context with two alternative contextual paths (a, b) eliciting the same follow-up utterance:

Laura and you live on the same street and you are about the same age. You know each other only by sight. Today you have met by

chance at the theater. Having greeted each other, you are now waiting for the show to start, seated side by side. While waiting, you make small talk.

(a) Literal-biased condition

Before the play starts, a theater employee announces over the PA system that today is International Theater Day and at the end of the show every member of the audience will receive a free ticket for another play. You look at Laura and say to her:

*(That's) fantastic!*²⁹ (follow-up target utterance)

(b) Ironic-biased condition

Before the play starts, a theater employee announces over the PA system that unfortunately the performance has to be canceled because the leading actress has lost her voice. You look at Laura and say to her:

(That's) fantastic! (follow-up target utterance)

Importantly, all the discourse contexts were carefully designed to minimize the role of social variables that might favor an ironic interpretation of the utterance. For example, the two interlocutors in the story are no more than acquaintances, since a closer relationship between them might increase the likelihood of irony in their

²⁹ Though *Oh, great!* might be a more pragmatically accurate English translation of the original Catalan follow-up utterance (*Això és fantàstic!*), for the sake of clarity we have decided to use the more literal translation of the original Catalan expression.

interaction (Spotorno, Koun, Prado, Van Der Henst & Noveck, 2012). In addition, to prevent the participants' biases from affecting their interpretation of the story (and thus their rendering of the sentence), information related with social class, job, and the particular interests of the interlocutors was not presented (Kreuz & Glucksberg, 1989). And most importantly, the discourse context (which would prompt either a sincere or a sarcastic utterance performance) was designed to affect the two interlocutors in the same way.

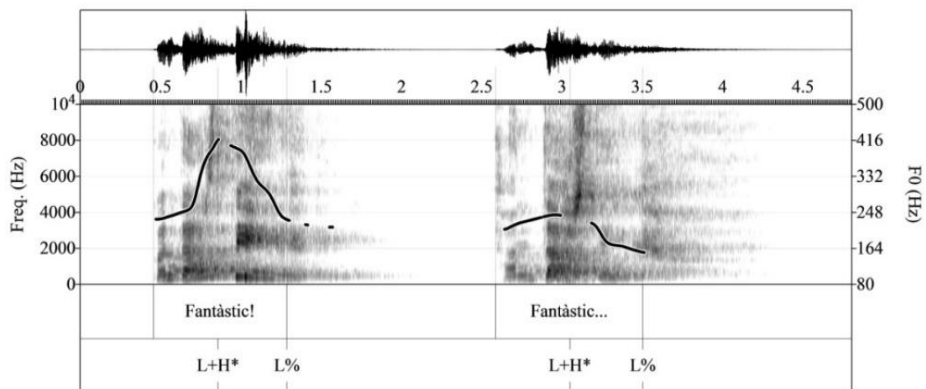
Audiovisual recordings

Audiovisual recordings were performed in a quiet room at the Universitat Pompeu Fabra with a professional digital video camera (Panasonic AG-HMC41). The eight participants were asked to silently read the prompt discourse contexts and were then recorded while producing the target follow-up utterances (e.g., *(That's) fantastic!* in (2)). The speakers were filmed facing the camera and 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 720×576 pixels. The sound was sampled at 44,100 Hz using 16-bit quantization. A total of 64 utterances were obtained, that is, 32 sincere utterances and 32 sarcastic utterances (8 speakers \times 4 discourse contexts \times 2 conditions—literal-biased vs. ironic-biased).

Prosodic and gestural analysis

In order to assess the prosodic cues of the 64 follow-up utterances in the two conditions (32 sincere vs. 32 ironic), they were acoustically analyzed using Praat software (Boersma & Weenink, 2008) and intonational patterns were coded following the Cat_ToBI system (Prieto, 2014). Figure 16 shows an example of one of the follow-up utterances in the two conditions.

Figure 16. Waveforms, spectrograms, and F0 contours of two versions of the Catalan utterance *Fantàstic!* ‘Fantastic!’, namely sincere (left) and sarcastic (right).



As for the acoustical analysis, following Bryant (2010) the recorded follow-up utterances were analyzed in terms of four prosodic features (average F0, F0 variability, MSD [mean syllable duration], and average intensity). 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 follow-up utterance




(in ms.) by the number of syllables. Four Generalized Linear Mixed Model (GLMM) tests were run using IBM SPSS Statistics 23.0 (IBM Corporation, 2015). The variable DISCOURSE CONTEXT (2 levels: literal-biased vs. ironic-biased) was set as the fixed factor, and the dependent variables were the four acoustic dimensions. Subject and Item were set as random factors. The results showed that utterances performed in the ironic-biased context condition were produced with a significantly higher average of F0 ($p < 0.01$), higher F0 variability ($p < 0.01$), and longer durations ($p < 0.01$) than utterances performed in the literal-biased condition.




Regarding gestures, the 64 follow-up utterances were analyzed with ELAN (Lausberg & Sloetjes, 2009) following the guidelines of the MUMIN Multimodal Coding Scheme (Allwood et al., 2007: 278) with the addition of two more gestures, ‘shoulder shrug’ and ‘averted gaze’ (see also González-Fuente et al., 2015). The analysis revealed that although several gestures or facial expressions (raised eyebrows, for example) were common to both sincere and sarcastic productions, confirmation head nods and smiles (i.e., mouth with a corners-up movement) were exclusively used during sincere utterance performances, while head movements such as shaking, tilting, and turning, as well as stretching of the mouth and averted gaze were used only during sarcastic performances (see Table 8).

Out of the 64 recordings (4 follow-up utterances \times 2 discourse context conditions \times 8 speakers), a subset of 8 video files (4 speakers \times 2 video files) were selected by the authors to prepare the stimuli to be used in all three subsequent experiments. Note that in

the end recordings of only four of the eight speakers were needed, with each speaker producing sincere and sarcastic comments for only one discourse context.

Table 8. Examples of the typical gestures produced in the literal-biased condition (i.e., sincere utterances) vs. the ironic-biased condition (i.e., sarcastic utterances).

<p>Literal- biased condition (sincere utterances)</p>			
	<p>Smile (corners-up mouth)</p>	<p>Head nod Raised eyebrows</p>	<p>Smile (corners-up mouth)</p>

<p>Ironic- biased condition (ironic sarcastic utterances)</p>			
	<p>Averted Gaze</p>	<p>Shoulder shrug Head tilt Raised eyebrows</p>	<p>Stretched mouth</p>

Audiovisual stimuli for Experiment 1 and Experiment 3 consisted of all 8 of the selected video image files (4 sincere and 4 ironic) and their corresponding 8 audio files.

For Experiment 2, the audiovisual stimuli consisted of 8 mismatched audiovisual files, which were prepared by digitally remixing the audio and visual tracks of the original 8 video recordings.

Experimental materials

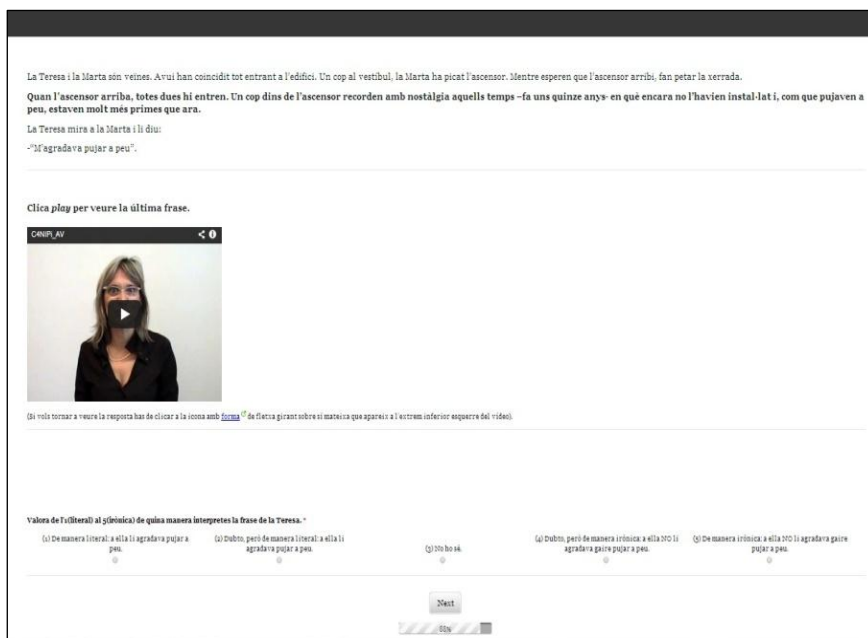
The experimental materials obtained from the DCT task as described above were put together as an audiovisual questionnaire using SurveyGizmo (Vanek & McDaniel, 2006), open source software for generating and administering online questionnaires (see Figure 17).

The questionnaire contained a total of 8 trials. Each trial consisted of the written description of a discourse context followed by an embedded audiovisual or audio clip of a speaker uttering the follow-up statement (e.g. *Fantastic!*).

In order to avoid the potential influence of contextual cues (following Voyer et al., 2016), we decided to use only ambiguous discourse contexts which would not give any clues about the literal vs. ironic interpretation of the follow-up utterance. For a complete set of the four ambiguous discourse contexts, see the Appendix A. Crucially, the follow-up sentences were presented in two multimodal³⁰ realizations, sincere or sarcastic, and in two modality conditions, audio-only (AO) or audiovisual (AV).

³⁰ We will use the term ‘multimodal’ henceforth to refer to the combination of prosodic and gestural cues. On the other hand, we will use ‘modality’ with regard

Figure 17. Screenshot of the online questionnaire used in Experiment 1. The written description of a discourse appears above an embedded audiovisual recording of a speaker making a comment in reaction to the context.



Two different questionnaire forms were designed so that each questionnaire contained 4 of the 8 AO stimulus files (block 1) and 4 of the 8 AV stimulus files (block 2), with different sets in each form. The files were distributed in such a way that in neither questionnaire would participants be exposed to the same sentence produced in the same multimodal realization in the two blocks.

to the communicative mode or channel, there being two possibilities in the present study, audio-only or audio + visual.

c) Procedure

The 45 participants in the irony detection task individually completed one of the two online versions of the questionnaire, each of which consisted of 8 trials. Participants were presented first with the AO condition (4 trials, block 1) and then with the AV condition (4 trials, block 2), with the order of trials automatically randomized by Survey Gizmo within each block. For each trial, after reading each ambiguous discourse context, participants were asked to listen to the follow-up utterances presented in either sincere or sarcastic multimodal conditions and in either AO or AV modality conditions. They could read the contexts and listen to/watch the recordings as many times as they wished. They were then asked to rate the degree of irony they perceived in the speaker's performance on a 5-point Likert scale included in the questionnaire, with 1 indicating 'very literal' and 5 indicating 'very ironic'. It took participants 11 minutes on average to complete the questionnaire. A total of 360 responses were obtained (45 participants \times 8 experimental trials).

d) Measures and statistical analyses

The 360 Likert scale responses were analyzed with a GLMM using IBM SPSS Statistics 20.0 (IBM Corporation, 2011), with Perceived Degree of Irony as a dependent variable. The fixed factors were multimodal cues (2 levels: Sincere and Sarcastic), and modality (2 levels: AO and AV). Subject and Item were set as random factors.

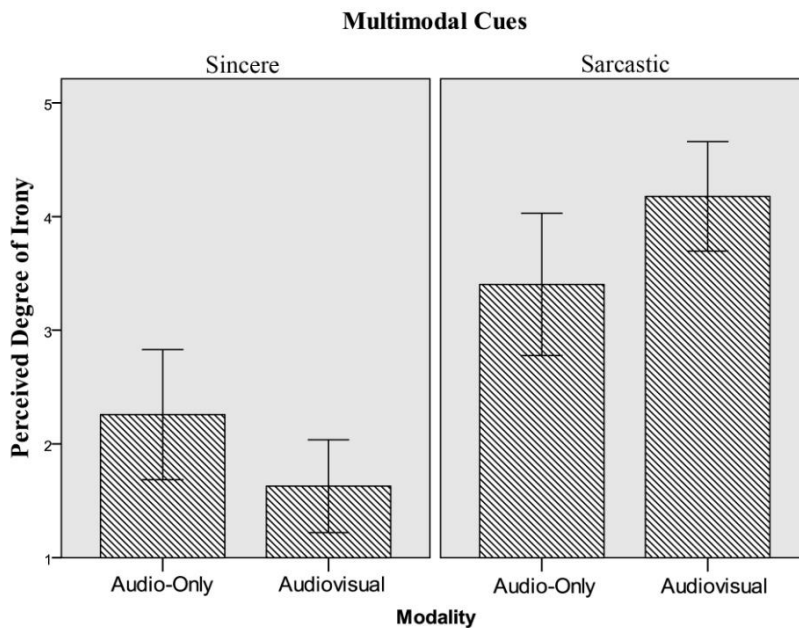
4.2.2. Results

Results of the GLMM showed a main effect of MULTIMODAL CUES ($F(1, 352) = 506.64, p < .001$), as well as a main effect of MODALITY ($F(1, 352) = 23.82, p < .001$). Post-hoc analyses revealed a statistically significant interaction between MULTIMODAL CUES and MODALITY ($F(1, 352) = 9.45, p < .005$), indicating that the effect of AO or AV presentation on the Perceived Degree of Irony variable was significantly different depending on whether the follow-up utterance had been produced with sincere or sarcastic multimodal cues. Interestingly, Perceived Degree of Irony scores were significantly higher when utterances were presented in the AV modality condition than when they were presented in the AO modality ($p < 0.01$). The bar graphs in Figure 18 show the average Perceived Degree of Irony (from 1 = very sincere to 5 = very ironic, y-axes) as a function of multimodal realization (sincere (left panel) vs. sarcastic (right panel) prosodic-gestural cues), and modality (AO (left bars) vs. AV (right bars)).

The results of Experiment 1 show clearly that the AV presentation modality had a stronger effect than the AO modality, thus suggesting that visual cues make a strong contribution to irony detection. In order to test the relative contribution of the visual and prosodic cues, a new irony detection task was designed (Experiment 2) with a set of materials containing mismatched prosodic and gestural cues. In other words, in this set of trial ironic-looking gestures were juxtaposed with sincere-sounding prosody and

ironic-sounding prosody was juxtaposed with sincere-looking gestures.

Figure 18. Average Perceived Degree of Irony as a function of multimodal realization (sincere, left panel, vs. sarcastic, right panel), and modality (AO, left bars, vs. AV, right bars).



4.3. Experiment 2

The main goal of Experiment 2 was to test the relative contribution of prosodic vs. gestural cues to the perception of irony. To do this, the same task set-up described in Experiment 1 was used, but now with a set of stimuli in which the visual (gestural) and audio (prosodic) messages were artificially mismatched.

4.3.1. Methods

a) Participants

A total of 18 Catalan speakers participated in Experiment 2, none of them having participated in Experiment 1. Results from one participant were excluded from subsequent analysis because he reported using Catalan less than 50% of the time on a daily basis. The responses of the remaining 17 participants (8 women and 9 men; mean age = 25.41, stdev = 4.33), who reported using Catalan on average 77% of the time (stdev = 3.23), were submitted to analysis.

b) Materials

As in Experiment 1, an audiovisual questionnaire form consisting of 8 trials was prepared using SurveyGizmo and the discourse contexts and recordings described in section 4.2.1.b, though in this case just one version of the form required. Each trial consisted of an ambiguous discourse context followed by a follow-up positive statement (e.g. *Fantastic!*). Discourse contexts for Experiment 2 were the same ambiguous discourse contexts as those employed in Experiment 1. In this case, however, a new set of mismatched audiovisual stimuli was prepared using Adobe Premiere Pro CS5 by digitally crossing the audio and visual tracks of the audiovisual recordings (i.e., sarcastic audio tracks were juxtaposed with sincere audiovisual tracks, and vice versa). An informal inspection of the stimuli did not reveal cases of undesired lip-sync problems and the

resulting cross-modal mappings were judged by the authors to appear natural. To confirm these impressions, we asked three independent judges to check the stimuli in terms of whether they felt that the A + V mappings were temporally congruent or not. They reported no problematic cases of temporal incongruence.

c) Procedure

As in Experiment 1, participants in Experiment 2 responded to the online questionnaire individually. They were instructed to read each discourse context and then watch the video clip that followed as many times as they wished. They were then asked to rate, using a 5-point Likert scale, the perceived degree of irony conveyed in the recording. The order of the 8 trials was randomly changed for each participant by the SurveyGizmo application. It took participants an average of 9 minutes to complete the questionnaire.

We obtained a total of 136 responses (17 subjects \times 8 trials).

d) Measures and statistical analyses

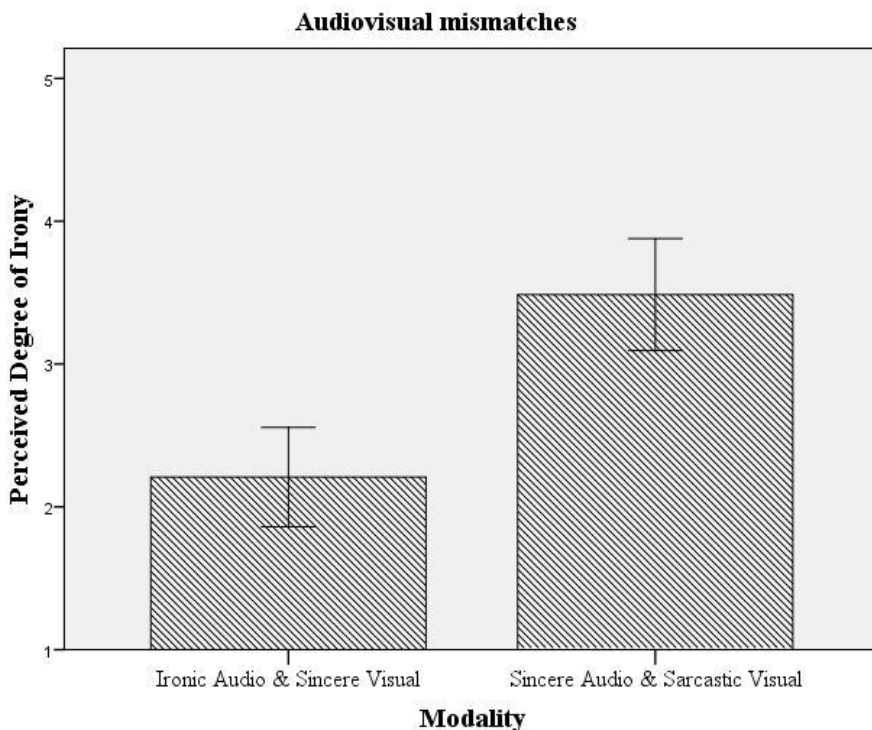
The Likert scale responses were analyzed with a GLMM in which the fixed factor was audiovisual mismatch (2 levels: sarcastic audio track + sincere visual track vs. sincere audio track + sarcastic visual track). Subject and Item were set as random factors.

4.3.2. Results

The results of the GLMM analysis revealed a main effect of audiovisual mismatch ($F(1, 134) = 23.59, p < .001$), which indicates

that when the prosodic (audio) and gestural (visual) components of follow-up utterance performances were incongruent, the type of mismatch had a clear effect on the listeners' detection of irony. The bar graphs in Figure 19 show that listeners rated as more ironic utterances presented in the sarcastic audio track + sincere visual track condition (left bar) than in the sincere audio track + sarcastic visual track (right bar).

Figure 19. Average Perceived Degree of Irony as a function of two types of audiovisual mismatches. The left-hand bar shows the results for stimuli in which sarcastic audio tracks were combined with sincere visual tracks. The right-hand bar shows results for stimuli in which sincere audio tracks were combined with sarcastic visual tracks.



The results of Experiment 2 clearly show that listeners exposed to mismatched A + V combinations relied more strongly on visual information than on audio information, suggesting that the gestural component was stronger than the prosodic component in the detection of ironic intent. However, it was not clear how these features might be interacting with the contrastive valences introduced by discourse context. Therefore a third experiment was designed to try to control for this factor.

4.4. Experiment 3

Experiment 3 aimed to assess the relative role played by three types of cues—contextual, prosodic, and gestural—in the detection of verbal irony. To do this, listeners were presented with a set of discourse contexts of two types, literal-biased or ironic-biased, each one of which was followed by a positive sentence performed in one of two multimodal conditions (with prosody and gesture intended to convey either sincerity or sarcasm) and presented in one of two modality conditions (audio-only or audiovisual). This design would allow us to test the effect of congruent and incongruent pairings of contextual and multimodal cues in verbal irony perception. Our hypothesis was that when exposed to incongruent pairings, listeners would find multimodal cues stronger than contextual ones.

4.4.1. Methods

a) Participants

Participants in Experiment 3 were a group of 34 Catalan speakers none of whom had participated in either Experiment 1 or 2. However, results from four of them were excluded from analysis because they reported that they used Catalan less than 50% of the time in their daily life. The responses of the remaining 30 participants (17 women and 13 men; mean age = 28.13, stdev = 12.23), who reported an average daily use of Catalan of 76% (stdev = 4.5), were submitted to analysis.

b) Materials

The materials used in Experiment 3 were similar to those used in Experiment 1 (see section 4.2.1.b). However, instead of ambiguous discourse contexts, two types of discourse contextual paths were used. Literal-biased discourse contexts were intended to trigger an interpretation of the follow-up utterance as being sincere while ironic-biased were intended to trigger an interpretation of the follow-up utterance as being sarcastic (see the Appendix A for the full set of 8 literal-biased and ironic-biased discourse contexts). As in Experiment 1, embedded recordings of the follow-up utterances were presented in one of two multimodal conditions with gestural and prosodic cues conveying either sincerity or sarcasm. Thus, in this experiment, the information coming from discourse contexts and multimodal cues could be congruent (i.e., positive comments produced with sincere multimodal cues after a literal-biased

context, or positive comments accompanied by sarcastic multimodal cues after an ironic-biased context) or incongruent (i.e., positive comments accompanied by sarcastic multimodal cues after a literal-biased context, and vice versa). The follow-up utterances were presented in one of two modalities, namely AO or AV.

As in Experiment 1, two different questionnaire forms were prepared, each one including 4 audio and 4 visual stimuli (see section 4.2.1.b). The visual and audio stimuli were strategically distributed in the two questionnaires so that participants were not exposed in the same questionnaire to the same sentence produced in the same multimodal condition (sincere or sarcastic) in the two blocks (AO and AV). For each questionnaire, they were presented first in the AO condition (4 trials, block 1) and then in the AV condition (4 trials, block 2) to avoid carryover effects of expected matches between modalities.

c) Procedure

The experimental procedure was the same as to the one used in Experiment 1 (see section 4.2.1.b). Participants were first presented with 4 recordings in the AO condition and then with 4 recordings in the AV condition with the order of trials within each condition automatically randomized by Survey Gizmo. Again, participants were then asked to rate the Degree of Perceived Irony on a 5-point Likert scale. The experiment lasted approximately 14 minutes. A total of 240 responses were obtained (30 participants \times 8 trials).

d) Measures and statistical analyses

The Likert scale responses for Perceived Degree of Irony were submitted to a GLMM in which the fixed factors were DISCOURSE CONTEXT (2 levels: literal-biased vs. ironic-biased), MULTIMODAL CUES (2 levels: sincere vs. sarcastic), and MODALITY (2 levels: AO and AV). Subject and Item were set as random factors.

4.4.2. Results

The GLMM analysis run with Perceived Degree of Irony as a dependent variable revealed a main effect of multimodal cues ($F(1,232) = 245.58, p < .001$), as well as a main effect of discourse context ($F(1,232) = 71.93, p < .001$), but no main effect of modality ($F(1,232) = 0.82, p = .37$). Post-hoc analyses revealed a statistically significant interaction between modality and multimodal cues ($F(1, 232) = 33.41, p < .001$); and modality and discourse context ($F(1, 232) = 25.85, p < .001$), indicating that the effect of ‘Audio-Only’ or ‘Audiovisual’ presentation on the Perceived Degree of Irony scores was significantly different depending on whether multimodal cues or discourse context variables were presented in ‘sincere’ vs. ‘sarcastic’ or ‘literal-biased’ vs. ‘ironic-biased’ conditions respectively. Moreover, a significant interaction between discourse context and multimodal cues ($F(1, 232) = 12.61, p < .001$) was also found, indicating that the effects of multimodal cues (sincere prosody/gestures vs. sarcastic prosody/gestures) on the Perceived Degree of Irony scores were different depending on the previous discourse context (literal-biased vs. ironic-biased). As we can see in

Figure 21, these significant interactions between variables emerge when discourse context and multimodal cues pairings were incongruent (i.e., when a literal-biased discourse context was followed by an utterance performed with sarcastic multimodal cues, or vice versa).

Figure 20 shows the mean Perceived Degree of Irony as a function of modality (AO vs. AV) in congruent Discourse Context/Multimodal Cues pairs. As expected, listeners interpreted the follow-up utterance to be literal when a literal discourse context was followed by a sincere utterance, both in the AO and the AV conditions (see the left panel). The mean values for the AO condition were 1.24 and for the AV condition 1.19. By contrast, listeners judged the follow-up utterance to be sarcastic when the discourse context and multimodal cues of the follow-up utterance were ironic (see the right panel). The mean values for the AO condition were 4.92 and for the AV condition 5.

Figure 21 plots the mean Perceived Degree of Irony as a function of modality (AO vs. AV) in incongruent Discourse Context/Multimodal Cues pairs. These results show that in the AO condition (left bars of each panel), participants tended to rely more on discourse context than on prosodic cues. Thus, an average score of 1.8 was obtained when the discourse context was biased toward a literal interpretation and an average of 3.8 when it was biased toward an ironic interpretation. By contrast, in the AV condition (right bars of each panel), the patterns reverses, with listeners tending to rely more on the AV contrasting information cues than

on contextual information. An average score of 3.6 was obtained when participants had access to mismatching (i.e., sarcastic) multimodal cues (right bar of left panel), and an average score of 2.1 when multimodal cues conveyed sincerity (right bar of right panel).

Figure 20. Average Perceived Degree of Irony (from 1 ‘very sincere’ to 5 ‘very ironic’, y-axes) in congruent Discourse Context/Multimodal Cues pairs, broken down by modality condition (AO left bars, AV right bars). Left-hand graph shows results for utterances reacting to literal-biased contexts and conveying sincere prosodic-gestural cues, while right-hand graph shows results for utterances reacting to ironic-biased contexts and conveying sarcastic prosodic-gestural cues.

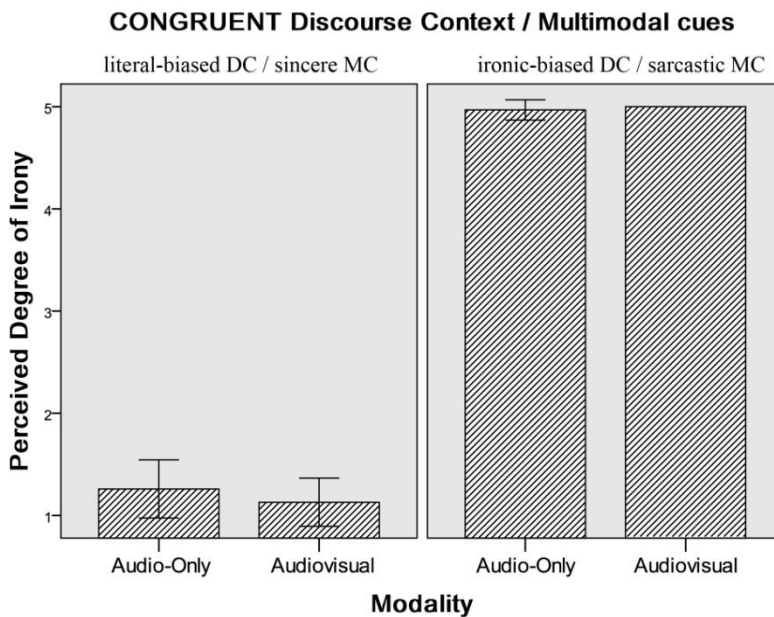
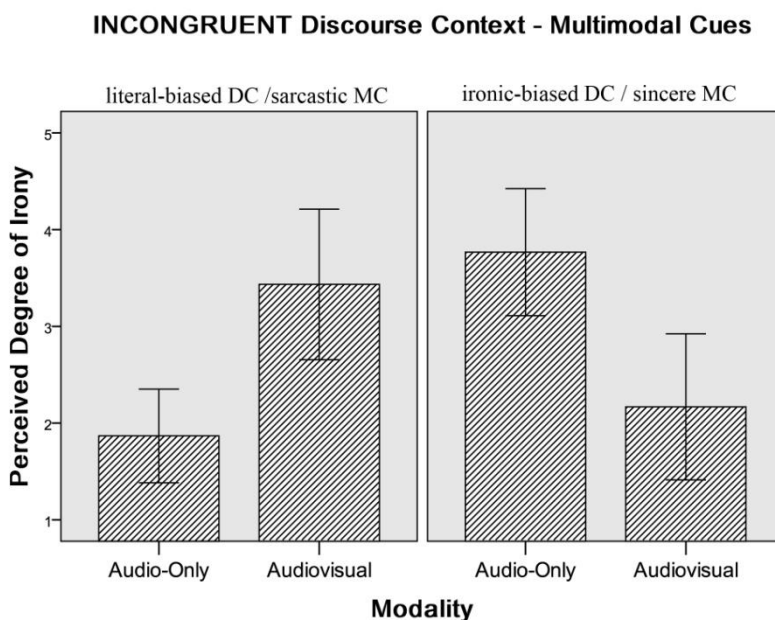


Figure 21. Average Perceived Degree of Irony (from 1 ‘very sincere’ to 5 ‘very ironic’, y-axes) in incongruent Discourse Context/Multimodal Cues pairs, broken down by modality condition (AO left bars, AV right bars). Left-hand graph shows results for utterances reacting to literal-biased contexts and conveying sarcastic prosodic-gestural cues, while right-hand graph shows results for utterances reacting to ironic-biased contexts and conveying sincere prosodic-gestural cues.



In general, the results show that when discourse context and multimodal cues are incongruent, participants rate the perceived irony of the follow-up utterances differently depending on the availability of AV information: crucially, while in the AO condition they rely more on the discourse context, in the AV condition they rely more on the prosodic and gestural realization of the follow-up utterance.

4.5. Discussion and conclusions

Despite the growing evidence that gestures play an important role in pragmatic comprehension (e.g., Goldin-Meadow, 2003; Krahmer & Swerts, 2004; Swerts & Krahmer, 2005; Holler & Wilkin, 2009; Borràs-Comes et al., 2011; Prieto et al., 2015), the relevance of gestures for irony detection has received little attention in the literature. Nor do we know much about the interplay between gesture, prosody, and context in the perception of irony. The purpose of the current study was to address this gap in our knowledge.

The results of Experiment 1 showed that, when the discourse context is ambiguous (i.e. when the listener cannot infer the intention of the speaker from situational cues), listeners rely on the prosodic and gestural characteristics of the utterance to infer ironic intent. Interestingly, the results of the experiment reveal an important asymmetry between the effects of auditory and visual cues. Specifically, the degree of irony perceived was significantly higher when utterances were presented in the AV condition than in the AO condition (4.3 over 5 vs. 3.2) and post-hoc analyses revealed an interaction between MULTIMODAL CUES and MODALITY ($F(1,352) = 9.45, p < .005$). Thus, our results are consistent with González et al. (2015), which empirically demonstrated that post-utterance gestural cues facilitate the detection of ironic intent. They also strongly support Caucci and Kreutz's (2012) suggestion

that visual cues can be used by listeners in addition to—or even in the absence of—to prosodic cues to detect an ironic intent.

Interestingly, the prosodic cues of ironically performed sentences only obtained mid-range scores in triggering ironic interpretations (3.2 over 5 in perceived degree of irony). These results seem to be partially in disagreement with previous research reporting that ironic intent can be totally recovered from the prosody of an utterance (Bryant & Fox Tree, 2005; Padilla, 2011; Loevenbruck et al., 2013). Indeed, though our experiment showed that listeners can use sarcastic (3.2 over 5) and even sincere (1.3 over 5) prosodic realizations to detect ironic intent, these cues are less successful than visual cues at leading them to an ironic interpretation of the sentence, indicating again a facilitatory effect for multimodal signals.

The results of Experiment 2 showed that, when participants were presented with an ambiguous discourse context followed by a set of mismatching audiovisual presentations of the follow-up utterance (i.e., a sincere audio track combined with ironic visual track, or vice versa), they crucially relied more strongly on the visual track. While the mean rate for perceived degree of irony was 2.5 (in the ironic audio track + sincere visual track condition), a 3.8 was obtained for the sincere audio track + ironic visual track condition. This clearly suggests that the ironic cues encoded by gesture can have stronger perceptive effects than those encoded by prosody.

The results of Experiments 1 and 2 are consistent with the results of some studies on audiovisual speech processing showing that when

listeners are exposed to information coming from auditory and visual sources, speakers tend to rely more on visual than on auditory information (e.g., McGurk & MacDonald, 1976; Gentilucci & Cattaneo, 2005; O'Shea, 2005; Bristow, Dehaene-Lambertz, Mattout, Soares, Gliga & Baillet, 2009). In relation to this, studies investigating certain types of attitudinal or emotional correlates clearly show that visual information is far more important for communicative purposes than acoustic information (Mehrabian & Ferris, 1967; Swerts & Krahmer, 2005; Dijkstra, Kramer & Swertz, 2006). Moreover, some studies have shown that the visual component is strongly related to prosody in the communication of a set of pragmatic meanings, such as prominence and focus marking (Hadar et al., 1983; Cavé et al., 1996; Krahmer & Swerts, 2007; Swerts & Krahmer, 2008; Dohen & Loevenbruck, 2009; Prieto et al., 2015), face-to-face grounding (Nakano et al., 2003), and question intonation (Srinivasan & Massaro, 2003). In strong agreement with these studies, the results of Experiment 1 demonstrate that the visual component is crucial in understanding irony.

In addition, if we compare the results of Experiment 2 to those of Experiment 1 (in which multimodal cues were congruent), we see that rating values for irony perception were more ambiguous in Experiment 2. These results support the idea that both visual and auditory components are important in the detection, perception, and processing of ironic speech, and that bimodal integration of visual and acoustic cues is necessary for accurate and fast irony detection processing (see, e.g., Goldin-Meadow, 2003; Swerts & Krahmer,

2005; Holler & Wilkin, 2009; Borràs-Comes & Prieto, 2011; Prieto et al., 2013, 2015).

The results of Experiment 3 showed that, as expected, when discourse context and multimodal cues were congruent (both literal/sincere or both ironic), listeners' ratings of perceived irony strongly agreed with the ironic or literal interpretations of the utterances (1.2 over 5 for the literal-biased context/sincere multimodal cues pairs and 4.9 over 5 for the ironic-biased context/ironic multimodal cues pairs). By contrast, when discourse context and multimodal cues were incongruent, participants rated the perceived irony of the follow-up utterance differently depending on the availability of the audiovisual information: crucially, while in the AO condition listeners relied more on the discourse context for irony detection, in the AV condition they strongly relied on the prosodic and gestural realization of the follow-up utterance, something which is consistent with the results of Experiment 2. This is one of the critical findings of this study. These results regarding incongruent pairs demonstrate that the role played by visual and prosodic cues in irony detection is by no means secondary, and that this strengthens the importance of this role is particularly clear when the prosodic and gestural cues accompanying a sentence do not agree with the expectations triggered by the discourse context. These results are consistent with the results reported in Woodland & Voyer (2011) and Voyer et al. (2016), as we also find that incongruent discourse contexts/tone of voice pairings tend to reflect a response from the participant that is closer to "neutral" on the scale rather than either sincere or ironic.

In addition, our study also finds that when visual information is added to auditory information in incongruent discourse context/multimodal cues pairings, more accurate responses are obtained in the direction of the multimodal cues.

In sum, though previous results on irony detection have claimed that irony detection processes seem to rest on the existence of pragmatic contrasts between discourse context and the propositional content of the utterance (e.g., Colston, 2002; Gerrig & Goldvarg, 2000; Ivanko & Pexman, 2003), as well as on the interaction between context and tone of voice (Woodland & Voyer, 2011; Voyer et al., 2016; Voyer & Vu, 2016), the results of the present experiments have shown that the multimodal cues that accompany utterances are key to irony understanding. Essentially, gestural information has been found to be even more critical than prosody and contextual information in its contribution to the detection of verbal irony. While some studies have pointed to the importance of gestural cues in the production and detection of verbal irony (Rockwell, 2000; Attardo et al., 2003, 2011; Caucci & Kreuz, 2012; González-Fuente et al., 2015), no previous investigations have analyzed the interplay between visual features and other types of information for irony detection. In general, our results are consistent with research on multimodal communication claiming that gestures play an important role in pragmatic meaning comprehension (e.g., Goldin-Meadow, 2003; Swerts & Kraemer, 2005; Holler & Wilkin, 2009; Borràs-Comes & Prieto, 2011; Prieto et al., 2013, 2015).

From a pragmatic point of view, we claim that in order to detect irony listeners need to attend to different levels of contrasting information. Thus it is important to assess not only the discrepancies between propositional content and contextual information (e.g., uttering *That's great* after a negative event, as emphasized by authors like Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995), but also the emotionally positive or negative attitude encoded by prosody and gesture and how this information interacts with propositional content (e.g., uttering *That's great* with a sad or sarcastic voice). Related to this, our results strongly support recent claims within Relevance Theory which argue that the speaker's feelings and emotions are a key factor for verbal irony expression and comprehension, since recognizing the affective attitude of an ironic speaker may be crucial to understand an ironic intent (Yus, 2016). It is precisely this affective stance expressed through prosody and gestures that is crucial for the detection of verbal irony. Our results thus expand on results reported elsewhere that highlight the role of contrast effects between context and propositional content in sarcasm perception (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Ivanko & Pexman, 2003). Our findings complement in particular those studies that have explored the interplay between prosodic and contextual cues in the perception of irony (Woodland & Voyer, 2011; Voyer et al., 2016).

All in all, the findings presented in this study show that not only verbal but also non-verbal components of communication are crucial for verbal irony detection. As Bryant (2011), Padilla (2004),

Poggi (2007) and Attardo et al. (2013) pointed out, verbal irony is a multimodal affair in which all the factors involved in the communication of an ironic utterance affect its successful interpretation. In other words, the verbal component of an ironic comment is only a single piece of the complex mechanism behind the communication of irony. Actually, ironic communication exists even in the absence of explicit linguistic expressions, as irony can be found in all kinds of human non-verbal expressions, including painting, music, and other art forms, whenever the implied meaning of the non-verbal expression is in contradiction with its external. As one form of ironic communication, verbal irony entails the presence of contradictory information coming from different sources simultaneously. One of these sources is the verbal component, but contrastin information is crucial to complete the ironic inferential process. This study presented novel empirical evidence of the stronger effects of multimodal—and especially gestural—cues in comparison with contextual cues in verbal irony detection. This crucial finding allows us to claim that the study of prosodic and gestural cues to verbal irony should be at the core of any pragmatic or psycholinguistic account of verbal irony production and comprehension.

5. CHAPTER 5: “Mismatching prosodic and gestural cues of emotion facilitate the detection of verbal irony in children”

5.1. Introduction

In everyday social interaction, speakers need to assess and integrate multiple sources of information in order to successfully understand others. In this comprehension process, listeners evaluate not just the verbal component of the message (i.e., *what* is being said by the speaker) but also the prosodic and gestural components with which it has been uttered (i.e., *how* it is being said by the speaker). These cues help listeners to infer information about the attitudinal and emotional states of their interlocutors such as uncertainty, incredulity, anger, sadness, etc., an interpretational process which is crucial to successful communication. As Van Lancker (2008: 206) points out, “one of the greatest challenges in psycho- and neurolinguistics research lies in understanding how these two components (i.e., verbal and emotional) interact in human communication”.

Verbal irony is a form of non-literal language in which there is an incongruity between what is said (i.e., the propositional content of an utterance) and what is meant. Although there are other forms of non-literal language such as metaphors or idiomatic expressions, verbal irony is probably the most complex, as comprehension of irony requires the listener to integrate information of various sorts,

including contextual cues (e.g., the specific situation, shared beliefs between speaker and listener) and emotional-attitudinal cues, typically conveyed through prosody and gesture.

Previous research on the perception of verbal irony has clearly shown that detecting ironic intent heavily relies on the perceptual contrast between the pragmatic context of an utterance and its propositional content (Colston, 2002). A few experimental studies have shown that the degree of mismatch between the propositional content of an utterance and its situational context (e.g., the comment *Well done!* uttered in a negative context) is correlated with their irony detection rates (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003). Moreover, recent studies investigating how contrasts between contextual, prosodic, and gestural cues affect verbal irony comprehension have shown that adults are more likely to detect irony in a statement when they have access to mismatching contextual cues (a) together with mismatching prosodic cues (Woodland & Voyer, 2011; Voyer et al., 2016) or (b) together with mismatching prosodic and gestural cues (González-Fuente, Zabalbeascoa & Prieto, submitted). Interestingly, González-Fuente et al. (submitted) study showed that listeners rely more heavily on prosodic/gestural cues than on contextual ones, and also more strongly on gestural information than on prosodic information for detecting irony. These results thus suggest that detecting speakers' ironic intent strongly relies on the ability to detect the mismatches between the valence of the propositional content of the sentence and the valences of contextual and, especially, prosodic and gestural realizations of that sentence. These

findings are consistent with relevance-theory accounts of irony which state that in order to understand an ironic remark it is necessary to identify not only the propositional attitude but also the affective attitude of the speaker towards the utterance (Yus, 2016).

With respect to acquisition, studies on the development of irony comprehension suggest that appreciation of the speaker's intent (understanding whether the speaker is trying to be pleasant or unpleasant) requires the assessment and integration of multiple cognitive and emotional cues, which entails a sophisticated inference process that becomes more accurate as children grow older (Ackerman, 1983; de Groot et al., 1995; Creusere, 2000; Nakassis & Snedeker 2002; Harris & Pexman 2003; Filippova & Astington, 2008). Albeit with some divergences among studies, it has been shown that children begin to detect certain aspects of ironic intent between 5 to 11 years of age (e.g., Milosky & Ford, 2009) and that they do so by means of contextual and prosodic cues. However, to our knowledge no previous studies have specifically tested the effect of facial gestures in combination with prosodic cues on verbal irony detection in children. The main goals of this study will therefore be (a) to test whether prosodic-gestural cues to emotion, in this case prosody and gesture, can facilitate the detection of a speaker's ironic intent by young children, and, if this is the case, (b) to determine at what age this effect first appears.

One of the most common forms of verbal irony is sarcasm, which is generally defined as a figure of speech that occurs when an utterance has an intended meaning that is precisely the opposite of its literal meaning and conveys an explicitly critical attitude towards

a particular event or person (Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995; Cheang & Pell, 2008). Researchers studying adult and child comprehension of verbal irony have typically focused on sarcastic remarks (Ackerman, 1983; Demorest, Mey, Phelps, Gardner & Winner, 1984; Capelli, Nakagawa & Madden, 1990; Nicholson, Whalen & Pexman, 2013). In order to make our results comparable with most of the literature, this form of irony will therefore be the focus of the present study.

The majority of developmental studies related to the acquisition of verbal irony agree that both context (Ackerman, 1983; Capelli et al., 1990; Winner & Leekman, 1991) and prosody (Ackerman, 1982, 1983; Capelli et al., 1990; Winner & Leekman, 1991; de Groot et al., 1995; Keenan & Quigley, 1999; Nakassis & Snedeker, 2002; Harris & Pexman, 2003; Climie & Pexman, 2008) are useful cues for children to understand sarcastic remarks. However, they diverge on the specific age at which children start to be able to use such cues for this purpose. Regarding contextual cues, whereas some studies have found that they do not play a role in children's detection of sarcastic remarks until they are 11 years of age (Capelli et al., 1990), other studies have found that they do so in children as young as 6 (Ackerman, 1983; Winner & Leekman, 1991). Similarly, whereas some studies have shown that children can use prosody as a cue to detect sarcastic remarks already at age 6 (Keenan & Quigley, 1999), others detected no such evidence until children were 8 (Ackerman, 1983; Capelli et al., 1990) or even older (Winner, Windmueller, Rosenblatt, Bosco, Best & Gardner, 1987). However, as Nakassis and Snedeker (2002) and Laval and

208

Bert-Eboul (2005) have pointed out, some of these divergences across experimental results may be attributable to differences in the materials and procedures used, especially those related to the operationalization of the ‘ironic tone of voice’. Though the abovementioned studies typically employed a distinction between ‘sincere’ and ‘ironic’ tones of voice, there was no consensus across studies about precisely what constituted an ironic tone of voice, with descriptions ranging from a “mocking intonation” (Capelli et al., 1990) to “stressed intonation patterns” (Ackerman, 1983), or even simply an “ironic tone of voice” (Nicholson et al., 2013).

It is important to stress at this point that there is in fact no single way to verbally express irony—i.e., there is no such thing as an ‘ironic tone of voice’ (Bryant, 2011; González-Fuente et al., 2015)—since the attitudes and emotions that can be expressed through an ironic remark range from the very positive to the very negative (Laval & Bert-Eboul, 2005; Wilson, 2013; Yus, 2016). Interestingly, to our knowledge the only study that has explored this issue in depth is Nakassis & Snedeker (2002), which tested the role of intonational cues expressing positive and negative emotions on adults’ and 6-year-old children’s comprehension of irony. They found that intonation acted as a relational cue, that is, that intonation facilitated children’s comprehension of an ironic remark when the valence of the intonational cue agreed with the ironic interpretation of the utterance (for example, negative-sounding intonation increased the probability that the child would understand that the speaker had a critical attitude, which in turn led them to understand that the speaker was expressing irony). In the light of

these results, the authors suggested that the research question “Does intonation affect irony comprehension?” should be reformulated to “What kinds of intonations in what kinds of contextual relationships affect irony comprehension?” All in all, we hypothesize that at least part of the reason for the discrepant results on the effects of prosody in irony detection across studies might lie on the lack of proper control of the emotional valences conveyed by prosody in contrast with those of the literal interpretation of the sentence. In order to avoid this shortcoming, in the present study we specifically controlled for the emotional valence—ranging from positive to negative—conveyed not only by prosody but also by facial gestures.

It is well known that facial gestures are a central cue to emotion detection in children. For example, Hübscher, Esteve-Gibert, Igualada and Prieto (2016) performed an uncertainty detection task with 4- to 6-year-old children using a series of materials designed to control for the presence of lexical, intonational, and gestural cues of uncertainty. Their results showed that children performed better overall in detecting uncertainty when gestural cues were present. Moreover, they found that the younger children were more sensitive to gestural and intonational marking of speaker uncertainty than to lexical marking (e.g., the use of adverbial forms such as *perhaps*), which suggests that the intonational and gestural features of communicative interactions may act as bootstrapping mechanisms in early pragmatic development. These findings are comparable to those of Armstrong et al. (2014), where facial gestures also seemed to scaffold children’s performance in detecting another type of belief state, incredulity.

Despite this evidence that facial gestural cues facilitate the comprehension of belief state in child development, however, as far as we know no studies have been conducted to investigate their role in the ability of children to detect irony. Interestingly, independent evidence has shown that there is a strong relationship between the perception of irony by children and their ability to detect emotions in others (i.e., their empathy skills). Nicholson et al. (2013) ran an irony perception and processing experiment with 6- to 7- and 8- to 9-year-olds. Whereas the 6- to 7-year-olds did not detect the ironic intention of the speaker (a near-zero accuracy for ironic statements was reported), in the 8- to 9-year-old group the authors found a strong correlation between the children's empathy skills as measured through the Empathy Quotient for Children and their accuracy in detecting irony (48% of correct responses, measured through an object selection task).

As noted above, the main goal of the present study is to test whether incongruity between multimodal cues (in this case verbal content on the one hand and prosodic and gestural cues on the other) can facilitate the detection by children of a speaker's ironic intent. Following Nicholson et al.'s (2013) experimental design, we had children undertake an audiovisual irony detection task with six congruent prompts (three short narratives with positive outcomes followed by a video of a speaker expressing a positive reaction and three short narratives with negative outcomes followed by a negative reaction) and six incongruent prompts (six short narratives with negative outcomes followed by a positive reaction). The

juxtaposition of a negative context with a positive reaction in the latter case was intended to simulate irony. Crucially, these latter six ‘ironic’ comments were presented in three conditions which manipulated the degree of congruence between the literal valence of the verbal utterance and the emotional valence of the prosodic and gestural cues which accompanied it. Specifically, while in the matching condition positive comments were produced with prosodic and gestural markers overtly conveying positive emotion, in the strongly mismatching condition the positive comments were produced with prosodic-gestural markers conveying a negative emotional valence. A weakly mismatching third condition was added that combined the ironic comments with prosodic and gestural cues in which negative emotional content was restrained to the extent possible. Our main hypothesis was that the stronger the degree of incongruity between the prosodic and gestural cues to emotion and the literal valence of a verbal comment, the higher the irony detection scores would be at all ages. Moreover, we predicted that a facilitating effect of the prosodic and gestural cues would be especially clear at the younger ages. In other words, while the contrast between a negative event and a positive verbal message in itself signals irony, it may be that this contrast will be intensified and therefore more recognizable by children if it is reinforced by emotionally negative prosodic and gestural signals. Our secondary research question concerned the age at which children are able to detect irony. Here we hypothesized that it would be earlier than what has been found in previous studies like Nicholson et al. (2013). All in all, the main novelty of the study was the fact that we

controlled for and manipulated the information conveyed through prosodic and gestural cues to test whether it would give children an advantage in the detection of irony in a verbal message.

5.2. Methods

5.2.1. Preliminary study: Discourse Completion Task

In order to obtain the audiovisual materials to be used in the irony detection task with children, we first asked 15 adult native Catalan-speakers (mean age = 24.7, stdev = 5.3) to participate in a Discourse Completion Task (henceforth DCT; Blum-Kulka, House & Kasper, 1989; Billmyer & Varghese, 2000; Félix-Brasdefer, 2010). The DCT methodology consists of a semi-spontaneous elicitation task in which participants are presented with a discourse context containing a situational prompt followed by a final target sentence which the participant is asked to produce out loud. The DCT was divided into two blocks. The first block was designed to obtain ironic utterances in three conditions (accompanied by matching, weakly mismatching, and strongly mismatching prosodic-gestural cues) and the second to obtain literal (i.e. non-ironic) utterances in two conditions (positive or negative, i.e., literal compliments or literal criticisms), both of which would serve as control stimuli.

a) Ironic utterances

To obtain the sarcastic reactions to be used in the irony detection task, a set of three discourse contexts were created by the

researchers (see Table 9 for an example). In each discourse context, a prompt describing a negative situation (e.g., *Un amic teu està fent volar un estel. De sobte, l'estel cau i es trenca.* ‘A friend of yours is flying a kite. Suddenly, the kite falls to the ground and is smashed’) was followed by the same positive comment (e.g., *Que ben fet!* ‘Well done!’). Since the aim of this study is to investigate the potential facilitation role played by prosodic and gestural cues in the detection of verbal irony, speakers were asked to produce the target sentences in three different conditions. We labeled these three prosodic-gestural conditions according to the degree of match between the valence of the sentence (which was always positive, i.e., ‘Well done!’) and the valence of the conveyed emotion (positive, weakly negative, or strongly negative). Thus, our three prosodic-gestural conditions were (1) matching (in which participants were told to pronounce the positive comment in an exaggeratedly positive or congratulatory manner in terms of both prosody and facial gesture), (2) weakly mismatching (in which they were asked to restrict the negative emotional content of their prosody and facial expression as much as possible), and (3) strongly mismatching (in which participants were told to accompany the comment with prosodic and gestural cues that would make it seem negative or critical). Thus, each speaker produced a total of nine utterances (3 discourse contexts \times 3 prosodic-gestural conditions). For a complete list of discourse prompts, see Appendix B. Conditions were presented to speakers in random orders.

b) Literal utterances

In order to obtain two sets of literal control stimuli, one positive (literal compliments) and the other negative (literal criticisms), participants were again presented with the same three prompt situations used to obtain the ironic utterances. In this case, however, each event had two outcomes, one positive and the other negative, followed by appropriate reactions. For example, “A friend of yours is flying a kite. He/She makes the kite do a loop” is reacted to with “Well done!” whereas “A friend of yours is flying a kite. Suddenly, the kite falls to the ground and is smashed” is reacted to with “What a terrible job!”. Thus, each speaker produced six control utterances (3 discourse contexts \times 2 conditions). Conditions were presented to speakers in random orders.

c) Recording procedure

With regard to the DCT procedure, participants were asked to read the situational prompt contexts and were then video-recorded producing the stipulated follow-up comments. Recordings were made using a Panasonic AG-HMC41 professional digital video camera in a quiet room at the Universitat Pompeu Fabra. Speakers were asked to face the camera and were filmed against a white backdrop, with their head and upper torso included within the video frame. The recordings were digitized at 25 frames per second, with a resolution of 720×576 pixels. The sound was sampled at 44,100 Hz using 16-bit quantization.

Table 9. Example of one of the discourse contexts used in the first part of the DCT to obtain ironic utterances. The original Catalan of the script is shown in italics with the English translation below. The discourse context contains a negative situational prompt (left column) followed by a positive target comment *Que ben fet!* ‘Well done!’(right column) in three prosodic-gestural conditions (middle column): matching (produced with prosodic and gestural cues conveying a positive emotion); weakly mismatching (cues with restrained emotion); and strongly mismatching (cues conveying a negative emotion).

Situational prompt (a negative event)	Prosodic-gestural conditions	Target sentence (a positive comment)
<p><i>Un amic teu està fent volar un estel. De sobte, l'estel cau i es trenca.</i></p> <p>A friend of yours is flying a kite. Suddenly, the kite falls to the ground and is smashed.</p>	<p><u>Matching</u></p> <p><i>Llavors, li dius al teu amic/ga amb entusiasme exagerat:</i></p> <p>You say to your friend with exaggerated enthusiasm:</p>	<p><i>Que ben fet!</i></p> <p>‘Well done!’</p>
	<p><u>Weakly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga amb emoció continguda:</i></p> <p>You say to your friend with restrained emotion:</p>	
	<p><u>Strongly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga ofensivament:</i></p> <p>You say to your friend in a critical manner:</p>	

d) Analysis of the video recordings

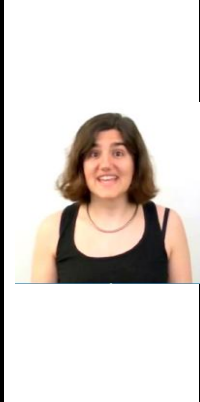
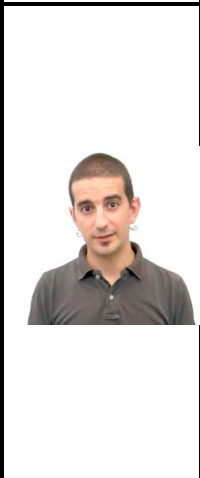
A total of 135 ironic utterances (3 discourse contexts × 3 ironic conditions × 15 participants) and 90 literal utterances (3 discourse contexts × 2 literal conditions × 15 participants) were obtained. The utterances were then prosodically analyzed with Praat (Boersma & Weenink, 2008) following the Cat_ToBI system (Prieto, 2014). Facial gestural cues accompanying the utterances (including any gestures appearing at the ends of utterances, i.e., gestural codas) were annotated with the help of the ELAN system (Lausberg & Sloetjes, 2009) following the guidelines of the MUMIN Multimodal Coding Scheme (Allwood et al., 2007: 278) with the addition of two more gestures, ‘Wrinkled nose’ and ‘Averted gaze’ (see also González-Fuente et al., 2015).


Table 10 shows the distribution of intonational and gestural cues produced by the participants in the DCT as they uttered critical comments in the three stipulated conditions (the table only reports cues that appeared in more than 30% of cases.) As can be seen, in the matching condition, participants mainly used a L+H* L% intonational pattern (73%) together with prominent or repeated head nods and raising of the eyebrows, and also smiles (53%); the gestural codas in this condition consisted of head nods, stretched mouth, and raising of eyebrows. In the weakly mismatching prosodic-gestural condition, the prosody used most often was the marked pattern L*!H% (66%) (which is used in Catalan to express skepticism or disagreement). The most common gestural cues produced in this condition were head tilts, raising of eyebrows, and

averted gazes, which took place as the comment was uttered, and head tilts, stretched mouth, produced at the gestural coda. Finally, in the prosodic-gestural strongly mismatching condition, participants used a L* L% intonational pattern in 87% of cases. Disapproval gestures such as head shakes and tilts, furrowed eyebrows, nose wrinkles, and squinted eyes overlapped with speech. During the gestural coda, speakers produced head shakes and furrowed eyebrows.

The prosodic and gestural cues used by speakers when they produced the literal control utterances were likewise analyzed (see Table 11). The results were as expected. On the one hand, literal compliments were generally (84%) produced with an emphatic L+H* L% intonational pattern and with gestures signaling approval, usually nods and raised eyebrows during sentence pronunciation and again nods and smiles during gestural codas. On the other hand, literal criticisms were almost always (95%) produced with a L*L% intonational pattern accompanied with disapproval gestures such as head shakes and tilts, furrowed eyebrows, wrinkled noses, and squinted eyes, both while the comment was being uttered and during the gestural coda.


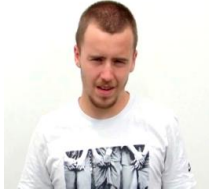
Table 10. Frequency of occurrence of the various intonational and gestural cues that characterized production by speakers of the 135 ironic utterances in the DCT, broken down by condition. The rightmost column shows video stills illustrating the most representative facial gesture for each condition.

Condition	Intonation	Gestures	Video still
<p>Matching</p> <p>(positive verbal content matches exaggeratedly enthusiastic gestural-prosodic cues) (N = 15)</p>	<p>L+H* L% (73%)</p>	<p>During speech</p> <p>Head nod (86%) Raised eyebrows (80%) Head tilt (50%) Smile (53%)</p> <p>Codas</p> <p>Head nod (73%) Raised eyebrows (36%) Stretched mouth (33%)</p>	
<p>Weakly mismatching</p> <p>(positive verbal content is accompanied by gestural-prosodic cues with emotion restrained) (N = 15)</p>	<p>L*!H% (66%)</p>	<p>During speech</p> <p>Raised eyebrows (66%) Head tilt (40%) Averted gaze (31%)</p> <p>Codas</p> <p>Head tilt (33%) Stretched mouth (40%) Raised eyebrows (33%)</p>	

<p>Strongly mismatching</p> <p>(positive verbal content is contradicted by gestural-prosodic cues signaling criticism) (N = 15)</p>	<p>L* L% (87%)</p>	<p>During speech Head shake (87%) Furrowed eyebrows (66%) Head tilt (33%), Wrinkled nose (40%), Squinted eyes (40%)</p> <p>Codas Head shake (53%) Furrowed eyebrows (33%)</p>	
--	------------------------	---	--

In sum, the results of the DCT confirmed previous findings about the audiovisual markers that accompany verbal irony. Verbal irony itself tends to be signaled by the use of specific pitch contours (e.g., the use of marked contrasting tonal-nuclear configurations; see González-Fuente et al., 2015) together with speech-accompanying gestures produced both during and after ironic speech (e.g., facial expressions involving specific eye and eyebrow configurations, laughter and smiles, etc.; see Attardo et al., 2003, 2011; Caucci & Kreuz, 2012; González-Fuente et al., 2015). In the following subsection, we will describe how the recordings obtained were then used as stimuli in an experiment designed to measure children's ability to detect irony.

Table 11. Summary and distribution of prosodic and gestural cues detected in the 90 literal utterances elicited by the DCT in the two literal control conditions. The rightmost column shows video stills illustrating the most representative facial gesture for each condition.

Literal control condition	Prosody	Gestures	Video still
literal compliment (N = 45)	L+H* L% (84%)	<p>During speech Head nod (93%) Raised eyebrows (77%)</p> <p>Gestural codas Head nod (73%) Smile (49%)</p>	
literal criticism (N = 45)	L* L% (95%)	<p>During speech Head shake (88%) Furrowed eyebrows (71%) Squinted eyes (48%) Wrinkled nose (33%) Head tilt (31%)</p> <p>Gestural codas Furrowed eyebrows (35%) Head shake (33%)</p>	

5.2.2. Experimental materials

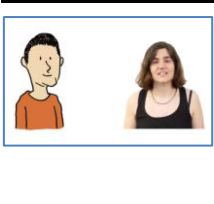
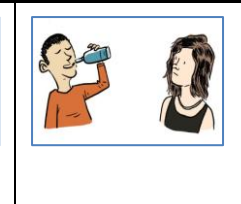
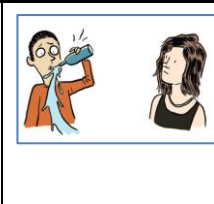
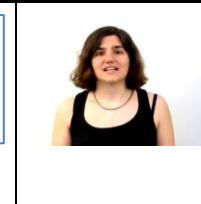
The video recordings that resulted from this preliminary DCT furnished us with the material we needed to create the stimuli for

the irony detection experiment involving children (the design for which, as noted above, is based in part on Nicholson et al., 2013). In this experiment, young children were presented with a set of twelve PowerPoint presentations each involving a short narrative followed by an embedded video of an adult reacting to the narrative (videos in fact recorded during the DCT described above). In six of the trials, the adult's reaction would be congruent with the outcome of the narrative, with a positive outcome to the narrative inducing a complimentary or congratulatory response and a negative outcome inducing a critical or hostile response. In the other six, the adult reacted incongruently, in that a negative outcome was greeted with a compliment, thus simulating irony. These six ironic reactions were audiovisually presented in the three prosodic-gestural conditions described above, with two reactions in the matching condition (the positive verbal content matching the enthusiastic prosodic-gestural cues), two reactions in the weakly mismatching condition (prosodic-gestural cues exhibiting restrained emotion), and the last two reactions in the strongly mismatching condition (the positive verbal content inconsistent with the hostility signaled by the prosodic-gestural cues).

Each discourse context depicted scenarios that were likely to be familiar to children and involved two characters, namely a cartoon character and a real human, who were different for every discourse context. Each discourse context was presented through a sequence of four slides. The first slide introduced the characters, the second and the third slides presented a short narrative with either a positive

or a negative outcome, and the fourth slide displayed an embedded video of the human character making a comment in reaction to the outcome of the event. This is exemplified in Figure 22 (see Appendix C for the full set of 12 sequences).

Figure 22. Sample slides from one of the PowerPoint presentations used in the irony detection task. Slide 1 introduces the two characters, slides 2 and 3 narrate an event (in this case with a narrative outcome), and slide 4 contains an embedded video file showing the human character reacting to the event.

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

The 12 embedded videos depicting humans reacting to the event described were selected by the authors from the set of 225 videos (90 literal + 135 ironic reactions) obtained in the DCT task described in section 5.2.1. Specifically, the authors selected 3 ‘literal compliments’, 3 ‘literal criticisms’, and 6 ‘ironic comments’ (2 for each prosodic-gestural condition: matching, weakly mismatching, and strongly mismatching). The selection was made based on the prosodic and gestural marking described in section 5.2.1.d, with each production chosen because it depicted the set of cues that seemed to be most prototypical for each condition. In

addition, each one of the selected utterances was performed by a different speaker.

Once the 12 stimulus presentations had been created, they were subjected to a validation process using the Survey Gizmo online survey platform (Vanek & McDaniel, 2006). Thirty-six Catalan-speaking adults viewed the 12 presentations online. After each presentation they were asked to indicate whether they interpreted the human character's reaction as being literal or ironic, and then in the latter case to rate the degree of criticism they perceived on a 7-point Likert scale. The results showed near total agreement (98.6%) among the 36 survey-takers in distinguishing the intended literal reactions from the intended ironic reactions. As for their ratings for degree of criticism on the 7-point scale, a Generalized Linear Mixed Model statistical test was conducted to test for significant differences between the three prosodic-gestural conditions using SPSS Statistics 23 software (IBM Corp., 2015). The dependent variable was PERCEIVED CRITICISM (with values from '0' 'not critical' to '7' 'highly critical'). The fixed factor was PROSODIC-GESTURAL CUES (3 levels: matching vs. weakly mismatching vs. strongly mismatching). SUBJECT and ITEM were set as random factors. The results showed that participants significantly distinguished between the three prosodic-gestural conditions ($F(2,213) = 23.671, p < .01$). Matching condition reactions obtained a mean criticism rating of 2.5 (SD 0.4), weakly mismatching condition reactions obtained a mean criticism rating of 4.1 (SD 1.0),

and strongly mismatching condition reactions obtained a mean criticism rating of 6.3 (SD 0.3).

5.2.3. Participants

A total of 92 children participated in the experiment. The children were separated into three groups according to age: a 5-year-old group (N = 31, mean age 5 years 3 months, stdev = 5.13), an 8-year-old group (N = 30, mean age 8 years 2 months, stdev = 4.45), and an 11-year-old group (N = 31, mean age 11 years 3 months, stdev = 5.02). All the children were from middle-class families and were enrolled as preschoolers or pupils at three Catalan public schools located in the Barcelona area. A language exposure questionnaire (based on Bosch & Sebastián-Gallés, 2001) was administered to the parents in order to ensure that the children were predominantly exposed to Catalan (as opposed to Spanish) on a daily basis (mean percentage of overall exposure to Catalan: 83%, stdev = 11.2). Parents were also informed about the experiment's goal and signed a participation consent form permitting their child to participate and the experimental procedure to be video-recorded. All children were tested individually at their respective schools and did not receive any sort of compensation for participating.

5.2.4. Procedure

The experiment took place in a quiet room at each of the three participating schools. The child was seated facing a computer

screen with one researcher, a male Catalan-speaking adult (the first author of this paper), seated next to him/her. A second researcher was seated behind the child to manually take note of the child's actions, and a video camera was positioned facing the child so that the full procedure could be recorded. Four training PowerPoint presentations were shown on the computer, followed by the twelve stimulus presentations. The four training presentations were used to train the child to show whether they judged the human character's reaction to a story as "*amable*" ("nice") or "*dolent*" ("mean") and followed the same structure as stimulus videos (see section 5.2.2), namely they depicted a narrative with either a positive or negative outcome followed by an embedded video of a person reacting. Of the four stories depicted, two had positive outcomes and two had negative outcomes. However, none of the reactions in the training presentations was ironic. After watching the person's reaction to each story, the children were told to signal their "nice"/"mean" judgment manually by placing one of two plastic toys into a plastic "answer bin" placed between them and the computer screen (see Figure 23 below). It was explained that one toy was the "nice duck" while the other was the "mean shark". If the child felt that the human character's reaction showed that he/she was behaving "nicely like the duck", they were to place the duck in the bin. On the other hand, if their reaction was "mean like the shark", they were to place the shark in the bin. Once the four training presentations were finished and the experiment proper began, the child was given no further prompting about placing the toy. The

respective location of duck and shark to the right or left of the child was counterbalanced across participants.

Figure 23. Experimental set-up of the irony detection task showing the plastic “answer bin” between the child subject and the computer screen, with (in this case) the “nice duck” to the child’s right and “mean shark” to the child’s left.



After the training trials, children performed a total of 12 experimental trials, consisting of six stories with congruent, literal reactions (three positive reactions to positive outcomes and three negative reactions to negative outcomes) and six stories with incongruent reactions, of which two had matching verbal and prosodic-gestural messages, two were accompanied by weakly mismatching prosodic-gestural cues, and two had mismatched verbal and prosodic-gestural. The 12 trials were presented to the children in different presentation orders. A total of nine PowerPoint presentations were created by ordering the trials in nine different ways, and children were randomly assigned to one of the nine

presentation orders. In total, the full procedure lasted around 15 minutes per child.

5.3. Results

An initial total of 101 children participated in the irony detection task but results from 9 of these children were excluded from the final analysis because inappropriate responses during the training trials suggested that these children had not properly understood the procedure. Thus data was obtained from 92 children each performing 12 experimental trials, yielding a total of 1104 responses (92 children \times 12 responses). Data consisted of child responses as noted by the second research during the experiment and cross-checked during subsequent viewing of the video-recordings. These responses were coded as ‘correct’ or ‘incorrect’. A ‘correct’ score was awarded under three conditions: a) the child selected the “nice duck” after seeing a complimentary reaction to a positive event outcome; b) the child selected the “mean shark” after seeing a critical reaction to a negative event outcome; or c) the child selected the “mean shark” after seeing a complimentary reaction to a negative event outcome in any of the three prosodic-gestural conditions. All other responses were coded as ‘incorrect’.

The results showed that participants were at ceiling in the accuracy of their responses for all literal control conditions, whether positive or negative. The degree of accuracy in the positive non-ironic condition (i.e., compliments following positive outcomes) was

100% and the degree of accuracy in the negative literal condition (i.e., criticisms following negative outcomes) was 96%. Thus these two conditions were not included in further analyses. A Generalized Linear Mixed Model (henceforth GLMM) test was used to analyze responses to the ironic reactions in the three prosodic-gestural conditions by means of SPSS Statistics 23 software (IBM Corp., 2015). The dependent variable was RESPONSE, a numerical measure obtained by calculating the mean proportion of correct to incorrect responses. The fixed factors were PROSODIC-GESTURAL CUES (3 levels: matching vs. weakly mismatching vs. strongly mismatching), AGE (3 levels: 5 years-old vs. 8 years-old vs. 11-years old), and their interactions. SUBJECT and ITEM were set as random factors.

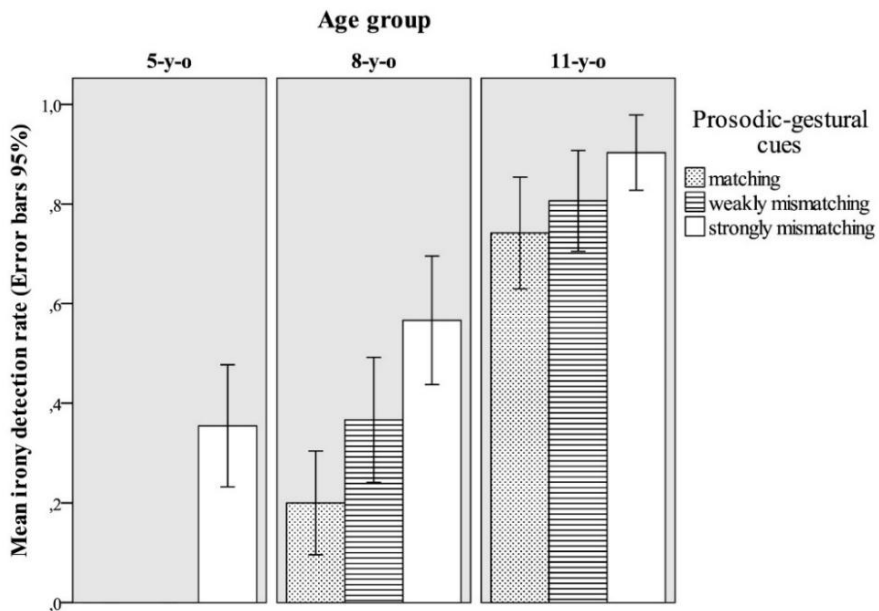
Results of the GLMM showed a main effect of AGE ($F(2,1089) = 56.82, p < .001$), indicating that 11-year-olds performed significantly better than 8-year-olds, who in turn performed significantly better than 5-year-olds. This points to a clear developmental pattern in the irony detection skills of children as they grow older. Also, a main effect of PROSODIC-GESTURAL CUES ($F(2, 1089) = 502.24, p < .001$) was found, correct response scores being significantly different between all three levels of the PROSODIC-GESTURAL CUES condition. Thus, the ‘strongly mismatching’ prosodic-gestural condition triggered significantly more correct responses than the ‘weakly mismatching’ prosodic-gestural condition, which in turn triggered significantly more correct responses than the ‘matching’ prosodic-gestural

condition. Crucially, the GLMM results also showed a main effect of the interaction between the two variables AGE and PROSODIC-GESTURAL CUES ($F(4, 1089) = 62.74, p < .001$), indicating that the effect of the prosodic-gestural condition on the responses differed depending on the age group. Figure 24 shows the mean proportion of correct responses broken down by prosodic-gestural condition (i.e., matching, weakly mismatching, and strongly mismatching) and age (5-year-olds, 8-year-olds, and 11-year-olds).

Post-hoc pairwise comparisons (with Bonferroni correction) showed that age groups significantly differed from each other depending on the prosodic and gestural cues which accompanied an ironic reaction comment. All age groups detected verbal irony significantly better in the strongly mismatching condition than in the other two conditions, and also significantly better in the weakly mismatching condition than in the matching condition. This indicates that the children were significantly more likely to perceive irony when the emotion conveyed by prosodic and gestural cues clearly contrasted with the propositional content of the utterance, in other words, when the reaction to a negative outcome was *Que ben fet* ‘Well done’ but the accompanying prosody and facial gestures signaled hostility or criticism. Moreover, in the case of 5-year-olds, the strongly mismatching condition was the only one of the three conditions that was ever interpreted as ironic (in 38% of cases). Taken as a whole, these results would seem to confirm our hypotheses that prosodic-gestural markers can facilitate the

interpretation by children of a speaker's ironic intent at early stages of their development, and that such markers are especially effective in this regard when they convey pragmatic information that strongly conflicts with the semantic content of the utterance.

Figure 24. Mean proportion of correct irony detection responses broken down by age and prosodic-gestural condition (dotted columns: prosodic-gestural cues matched verbal content; striped columns: prosodic-gestural cues weakly mismatched verbal content; white columns: prosodic-gestural cues strongly mismatched verbal content).



5.4. Discussion and Conclusions

This study examined whether prosodic and gestural cues to emotion can facilitate the detection of a speaker's ironic intent by children. By means of an audiovisual irony detection task we were able to show that strongly mismatching prosodic and gestural cues led to significantly higher irony detection rates than weakly mismatching and matching prosodic-gestural cues, not only in 8- and 11-year-olds but also in 5-year-olds. These results are in line with previous research that found that prosodic cues facilitated 6-year-old children's comprehension of a sarcastic remark when the valence of the prosodic cue contrasted with the literal interpretation of the utterance, that is, when the intonation conveying a negative emotional valence did not match the positive verbal content of a sentence (Nakassis & Snedeker, 2002). Going a step further of Nakassis and Snedeker's (2002) study, our results showed that the stronger the degree of incongruity between the valences of the prosodic-gestural cues and the verbal comment, the higher were the irony detection scores by children, as strongly mismatching prosodic and gestural cues led to significantly higher irony detection rates than weakly mismatching prosodic-gestural cues, which in turn triggered significantly more correct responses than matching prosodic-gestural cues. These findings thus expand on the results of previous studies which emphasize the role of contrast effects in sarcasm perception (Colston & O'Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003; Woodland & Voyer, 2011; Voyer et al., 2016; González-Fuente et

al., submitted). Specifically, our results agree with those studies which showed that the optimal combination of mismatched prosodic and/or gestural cues (i.e. the use of an ‘ironic’ tone of voice together with negative emotional facial expressions) together with mismatched discourse contexts (i.e. a negative situation) obtained the highest irony detection rates in adults (Woodland & Voyer, 2011; Voyer et al. 2016, González-Fuente et al., submitted).

On the other hand, our results show a very clear developmental pattern in irony detection skills by children, as their average irony perception scores increased with age in all prosodic-gestural conditions. In this regard our results are also consistent with previous literature (e.g., de Groot et al., 1995; Creusere, 2000; Filippova & Astington, 2008; Harris & Pexman, 2003; Nakassis & Snedeker, 2002). Yet one of the main questions addressed by this study was whether the access to visual cues would help younger children to detect irony. As noted above, our experimental design was based on that used in Nicholson et al. (2013). This was because, unlike all previous studies on children’s irony perception, Nicholson et al. (2013) included visual information in the experimental set-up. In their study, however, the visual cues were provided by puppets. In the present study, we went a step forward by using video-recordings of real humans, who were able to utter verbal messages while producing mismatched prosodic and facial gestural cues. Nicholson et al.’s (2013) study tested 6- to 7-year-olds and 8- to 9-year-olds and found that the younger age group had near-zero accuracy in detecting irony. By contrast, our

results for 5-year-old children showed that, while these young children likewise failed to detect irony when the prosodic and gestural signals accompanying an incongruently positive verbal message were either positive or weakly critical/negative, they did detect irony 38% of the time when positive verbal content was accompanied by clearly hostile prosodic and gestural cues. These results suggest that the perception of irony can appear at very early stages of development provided that the children have access to strongly mismatching prosodic and gestural cues to emotion, that is, when prosodic and gestural cues show a clear contrast with the propositional content of a sentence. In this regard, these findings are compatible with those of Armstrong et al. (2014) and Hübscher et al. (2016), where visual cues facilitated 4- to 6-year-old children's performance in detecting pragmatic meanings such as incredulity or uncertainty. In general, our results are in line with the growing consensus that pragmatic gestures act as bootstrapping devices in language (and specifically pragmatic) development (e.g., McNeill, 1998; McNeill, Cassell & McCullough, 1994; Kelly, 2001; Butcher & Goldin-Meadow, 2000).

In sum, our results clearly show that children are especially sensitive to emotional expressions conveyed by prosodic and gestural cues, and that they can actively use them to make judgments about a speaker's intent. This result is in line with previous studies which experimentally showing a strong relationship between irony appreciation in children and their empathy skills, that is, their ability to detect emotions in others

(Nicholson et al., 2013). As we signaled above, and as recent claims from pragmatic cognitive accounts such as the Relevance Theory point out, detecting the ‘affective attitude’ of the speaker is crucial to understanding an ironic remark (Wilson, 2013; Yus, 2016). Taking into account this emotional valence perspective, our claim is that research on the development of irony detection needs to incorporate a fine-grained control of the emotional-laden visual and prosodic information which accompanies ironic utterances. As Bryant (2012) claims, understanding verbal irony is a multimodal affair in which all the factors involved in the communication of an ironic utterance affect its successful interpretation. The study of irony detection not only needs to investigate the contrasts between the literal meaning of words and their pragmatic context but also needs to incorporate the study of the interplay of between contextual, propositional, and prosodic and gestural cues in verbal irony comprehension.

6. GENERAL DISCUSSION AND CONCLUSIONS

6.1. Summary of findings

The goal of this dissertation was to investigate the role of prosody and gestures in verbal irony production, perception, and development. Four independent studies were presented, each one in a separate chapter.

The first study investigated the role of prosody and gestures in the production of ironic utterances in the speech of a professional comedian, as well as their temporal interplay (Chapter 2). Two main results were obtained. First, we found that the professional comedian produced ironic utterances with a significantly higher density of prosodic and gestural markers compared to non-ironic immediately preceding utterances, and, as a novelty with respect to the literature on visual cues in ironic speech, we documented a strong presence of gestural markers at the end of ironic utterances (e.g., the so-called gestural codas, in 66% of cases) in comparison with non-ironic utterances (in 28% of cases). Second, we also found that the gestural markers associated with speech typically co-occurred with prosodic features and temporally aligned with intonational pitch accents. This fine-grained analysis of the temporal interplay between gestures and prosody in ironic speech constitutes a novel contribution to the literature in this dissertation.

The second study investigated the role of prosody and gestures in the production of ironic utterances in non-professional spontaneous speech, with special attention paid to the perceptual role of gestural codas in the detection of ironic intent (Chapter 3). Two main results were reported. First, as in the first study, speakers of non-professional spontaneous speech marked ironic utterances with a higher density of prosodic and gestural cues as compared to the immediately preceding non-ironic utterances. Moreover, the results revealed a more prevalent presence of gestural codas in ironic utterances (as 70% of the ironic utterances were followed by gestural codas compared to 27% of non-ironic utterances). Second, crucially, the results from a perception experiment confirmed the relevance of gestural codas in the detection of ironic intent, as listeners detected ironic intent significantly better when post-utterance gestural codas were present than when they were not.

The third study (Chapter 4) investigated the interplay between multimodal (i.e. prosodic and gestural) and contextual cues in a set of verbal irony detection tasks. Three main findings were obtained. First, we found that listeners detected irony more accurately when they had access to both prosodic and visual cues than when they just relied on prosodic information; second, that listeners relied more strongly on gestural information than on prosodic information; and third, that listeners relied more heavily on gestural cues than on prosodic or contextual ones for detecting irony. Overall, the findings of this study contribute to clarifying the role of the gestural component to verbal irony detection in terms of strength by

providing empirical evidence of the key role that gestural cues play in perception as compared to prosodic and contextual information in the detection of verbal irony.

Finally, the fourth study (Chapter 5) investigated how prosodic and gestural cues influence children's detection of verbal irony, with two main findings. First, we found that strongly mismatching multimodal cues (i.e. positive utterances produced with prosodic and gestural cues conveying a negative intent/emotion) led to significantly higher irony detection rates in 5-, 8- and 11-year-old children as compared to slightly mismatching and matching multimodal cues. Second, we also found that strongly mismatching multimodal cues facilitate irony appreciation at early ages. These findings constitute a novel contribution to the literature on verbal irony comprehension by children from both a processing perspective (as mismatching multimodal cues facilitate irony detection in children) and a developmental point of view (as mismatching multimodal cues facilitate irony detection at early stages).

In the next sections I will discuss these findings in relation to the previous literature and show how they contribute to the current existing body of research in the field.

6.2. Is there an ironic tone of voice or an ironic gestural pattern?

Previous research on the prosodic and gestural features of verbal irony has shown that speakers convey prosodic and gestural modulations in their ironic speech. One of the most widely discussed issues within this literature has been whether there is a consistent tone of voice (or an ironic gestural pattern) that we can identify in ironic speech.

With respect to prosody, ironic utterances have been reported to be produced with acoustic modulations in pitch (with a higher or lower F0 mean and variability), intensity variations (with higher intensity values) and a diverse set of duration features (e.g., slower speech rate, more and longer pauses) (e.g. Gibbs, 2000; Nakassis & Snedeker, 2002; Loevenbruck et al., 2013; Anolli et al., 2002; Attardo et al. 2003, 2011; Laval & Bert-Erboul, 2005; Cheang & Pell, 2009; Bryant & Fox Tree, 2002, 2005; Bryant, 2010; Scharrer et al., 2011; Padilla, 2004, 2011; Loevenbruck et al., 2013; González-Fuente et al., 2016). Pitch and intensity cues have yielded mixed results across studies and languages, and the only prosodic feature that has been found to be consistent across languages is the presence of slower speech rates. Moreover, some intonational features have been reported to be relevant for signaling ironic intent, such as rising final inflectional patterns in Spanish (e.g., Padilla, 2004, 2009), or specific nuclear tonal configurations in French (e.g., González-Fuente et al., 2016). As for visual cues,

previous literature has shown that speakers use a wide range of visual cues when communicating irony, as, for example, eyebrow raising, head movements, stretched lips, as well as smiles, laughter, and averted gaze (Attardo et al. 2003, 2011, Bryant 2011, 2012, Haiman, 1998, Hancock, 2004, Kreuz, 1996, Caucci & Kreuz, 2012, Gibbs, 2000; Williams et al, 2011, Padilla, 2004). In general, these gestural cues have been reported to convey information about the emotions and attitudes of the speaker, and they have been related to a wide range of different socio-communicative functions such as, for example, reinforcing a shared positive affective experience (Smoski & Bachorowski, 2003), strengthening friendship ties (Alvarado & Padilla, 2010) or being critical of something or someone (Sperber & Wilson, 1986/1995).

The studies included in Chapters 2 and 3 were designed to analyze the prosodic and gestural features in the speech produced by a professional comedian (Chapter 2) and by non-professional speakers involved in an informal conversation (Chapter 3). The two studies acoustically and gesturally analyzed the target ironic utterances with the immediately preceding non-ironic utterances. As an empirical novelty, in these studies we present a fine-grained analysis of the temporal alignment between prosodic and gestural features in ironic speech. In both studies, results showed that (a) spontaneously produced ironic utterances significantly contrast with the immediately preceding non-ironic utterances, in terms of both prosody and gesture and, (b) crucially, that some gestural cues can

appear aligned with prominent pitch accents and also independently, as gestural codas in a post-utterance position.

With respect to prosody, the results of the first study showed that relative to non-ironic utterances, ironic speech is characterized by a significantly higher F0 variability and slower speech tempo. Similarly, the results of the second study showed that ironic utterances were produced with higher rates of specific 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), with more pauses (45% in ironic target utterances vs. 18% in baseline utterances), and also with slower speech tempo than the preceding non-ironic utterances. These findings are consistent with previous studies reporting the use of intonational patterns to mark irony in different languages (e.g., Attardo, 2001, for English; Padilla, 2004, 2009, 2011, for Spanish; González-Fuente et al., 2016, for French). However, the use of specific tonal-nuclear configurations in verbal irony still remains quite unexplored. In this direction, recent research on the pragmatic meanings of intonation has shown that specific tonal nuclear configurations are closely related to the discursive functions of insubordinate clauses (Elvira-García, 2016; Elvira-García, Roseano & Fernández-Planas, 2017), which strongly suggests that much more research is needed to achieve a complete understanding of the role of specific tonal-nuclear configurations in signaling all kinds of linguistic meanings. As for gestural cues, both studies revealed that speakers produce ironic utterances with higher rates of visual cues

(e.g., eyebrow raising, head movements, smiles, laughter, and gaze changes), compared to non-ironic utterances.

All in all, the results of both experiments indicated that (a) speakers can signal an ironic intent by combining a variety of prosodic and gestural modulations, and (b) that there is no unique way to signal ironic intent through prosody and gesture, which leads us to conclude that we cannot identify a particular “ironic tone of voice” or an “ironic gestural pattern” that is specific to the marking of irony. Previous research suggested that different—and also contradictory—results across studies investigating the “ironic tone of voice” could be explained by differences in the methodological design, in the irony subtype under analysis, in the language-specific implementation of irony, and also in the specific intonational phonology of each language (Bryant, 2011; Loevenbruck et al., 2013). However, results of the fine-grained analysis of the prosodic and gestural markers carried out in Chapter 3 showed a non-significant relation between “irony subtype” and multimodal cues conveyed by speakers, this is, we did not find a characteristic or unique tone of voice for ‘sarcastic’ or for ‘hyperbolic’ ironic utterances.

Our results thus suggest that even with an exhaustive control of the ‘ironic subtypes’ under analysis, prosodic and gestural characteristics of spontaneous ironic speech are varied and lead to conflicting results when searching for a consistent ironic tone of voice or an ironic gestural pattern. Crucially, we claim that this is because intention and emotions that can be expressed through an

ironic remark range from the very positive to the very negative (Wilson, 2013; Yus, 2016), and that prosodic and gestural signals expressing these specific emotions and attitudes are extremely varied and overlap with a wide range of communicative goals. For example, ironic intent can express a variety of valences ranging from a positive valence (such as to bring humor to a situation, e.g. Attardo, 2013), or expressing surprise (Colston & Keller, 1996), to expressing a more negative or critical attitude (e.g. Sperber & Wilson, 1986/1995). We claim that prosodic patterns (and also gestural patterns) across languages are specially suited for signaling intentionality, and thus they can be extremely varied. In the case of gestures, speakers typically employ a variety of smiles and laughter to convey positive intent, while they use head shakes or eyebrow frowning to convey negative intent (see Wharton, 2009). Moreover, it is important to point out that the wide range of visual cues that can appear during communication can be polysemic (Poggi, 2007) and that the same eyebrow configuration combined with a different head movement could lead to different or even opposite interpretations. It is thus important to carry out more studies that take into account the complexity of these gestural and prosodic patterns in online communication.

In sum, the results of the studies presented in Chapters 2 and 3 show that speakers actively use prosodic and gestural cues to signal ironic intent, but also, and crucially, that there is no unique way to prosodically and gesturally mark ironic speech, as both prosodic and gestural markers can be correlates of a wide range of different

communicative functions. In general, the results of both experiments agree with previous claims that there is no particular “ironic tone of voice” or an “ironic gestural pattern” that is specific to signaling ironic intent, and that speakers can indicate the presence of verbal irony by combining and contrasting a variety of prosodic modulations that are not specific to verbal irony (Attardo et al., 2003; Bryant, 2010, 2011; Padilla, 2009, 2011). Together with Bryant (2012), we think that a better understanding of the role of prosodic and gestural markers in verbal irony production and comprehension should depart from the traditional conceptions of verbal irony (which constrained the notion of what verbal irony is and consequently the empirical research on this topic) and approach this issue from pragmatic and cognitive psychological perspectives which integrate the psychological processes and the communicative functions underlying this complex phenomenon of inferential communication. In the perception studies presented in this dissertation (Chapters 4 and 5), we adopted this cognitive approach and our results will be discussed in the light of Contrast and Assimilation Theory (e.g., Colston, 2002) and Relevance Theory (e.g., Sperber & Wilson, 1986/1995).

6.3. Gestural cues can appear both aligned and misaligned with prosodic cues: the particularity of gestural codas.

Another important finding of the first two studies in this dissertation is related to the temporal alignment between prosodic and visual

cues of irony. The results of both studies showed that prosodic and gestural cues can appear both temporally aligned with speech, but also misaligned. Related to this issue, the main findings were: (a) that pitch accents associated with emphatic tonal nuclear configurations (e.g. L+H* L%, L+H* L!H% and L* !H%) are often temporally associated with gestures and with some facial gestures, specifically eyebrow and head movements and (b) that some visual cues are used independently from prosodic ones, especially in what we have called post-utterance gestural codas, or gestural patterns which appear in post-utterance position, e.g. when prosodic and lexical information are not present.

In relation with the appearance of gestural codas, the results presented in Chapters 2 and 3 revealed a more frequent presence of post-utterance gestural codas in ironic utterances compared to preceding non-ironic utterances (66% vs. 28% in Study 1, and 70% vs. 27%, respectively in Study 2). A reasonable explanation for the higher incidence of gestural codas in ironic than in non-ironic speech could be the fact that, as suggested by Poggi (2007), in contrast with deception, ironic communication expects the recipient to understand the implied meaning and intentionality of the speaker, and gestural cues (and specifically gestural codas) are used in order to facilitate the success of the comprehension process. Crucially, the relevance of gestural codas for irony detection was confirmed by the results of the perception experiment presented in Chapter 3, which showed that the presence of these explicit codas helped listeners to be significantly more successful at irony detection.

These gestural codas consisted of gestures and facial expressions such as smiles, laughter, averted gazes, eyebrow raising, head movements, and mouth stretching. It is important to highlight the fact that these visual cues are not an exclusive feature of gestural codas. Our results also reveal that similar types of visual cues can appear during and after (and even before) the ironic utterance being pronounced.

In general, the strong presence of prosodic and gestural cues (and especially gestural codas) in ironic speech found in our studies highlights the relevance of multimodal cues in ironic communication. These findings are consistent with the proposals put forth within Relevance Theory, namely, that prosodic and gestural markers are used by speakers in order to reduce the processing effort of the interlocutor, supposedly until the speaker ensures that the ironic understanding process has been completed (e.g., House, 1990, 2006; Clark and Lyndsey, 1990; Fretheim, 2002; Wilson & Wharton, 2006; Escandell-Vidal, 1998, 2011a, 2011b; and Wharton, 2009). Moreover, the presence of gestural codas contributes to clarifying and illustrating the notion of “ironic utterance”, emphasizing the claim made from pragmatic accounts that an ironic speech act is not concluded until all the relevant information has been expressed, this is, that “[...] a discourse unit has no more limits than those established by the speaker and his

communicative intention, regardless of the degree of complexity of its formal realization” (Escandell-Vidal, 2006: 28)³¹.

6.4. Prosodic and gestural contrasts signal ironic intent

Previous research on the processing of verbal irony has provided empirical evidence of the important role that contrasting information between situational context and propositional content has in the understanding of verbal irony. Specifically, it has been shown that irony detection is facilitated by manipulating the degree of negativity of the discourse context: the more mismatching contextual cues are, the more accurate ratings of irony (Colston & O’Brien, 2000; Gerrig & Goldvarg, 2000; Colston, 2002; Ivanko & Pexman, 2003). Within the so-called Contrast and Assimilation theory (which argues that contrasts may serve as a guide to a more accurate detection of ironic intent) a series of studies have shown that contrasts between ironic tone of voice and the positive or negative valence of the utterance also facilitate the perception of irony (Woodland & Voyer, 2011; Voyer et al. 2016). This dissertation provides novel empirical evidence which is consistent with this line of research, providing new data both for adult and child populations. Thus, the results of the first experiment in Chapter 4 demonstrated that listeners detected more accurately the

³¹ Author's translation from Spanish: “[...] una unidad del discurso no puede tener más límites que los que establece el emisor y su intención comunicativa, independientemente del grado de complejidad de su realización formal”.

ironic intent of the speaker when they had access to audiovisual information than when they relied only on speech information (e.g., the audio-only condition). Similarly, the results of the third experiment showed that when both contextual and multimodal cues contrasted with the positive valence of the target sentence, speakers detected ironic intent more accurately. Crucially, results for incongruent context/multimodal cue pairs showed that having conflicting information between contextual and multimodal cues (i.e. when only contextual or multimodal cues were contrasting with the positive valence of the utterance) led to lower ratings of perception of irony. Thus, the results of these two experiments demonstrated that the greater the amount of contrasting information, the more easily the speaker's intent was detected. Furthermore, the results of the irony detection task with children presented in the fourth study (Chapter 5) showed that stronger mismatches between the positive utterance and the multimodal cues crucially benefited irony detection by children in the three age groups. Again, the important role of contrasting multimodal information in verbal irony detection was confirmed, in this case for a child population, as suggested by Nakassis & Snedeker's (2002) criteria for prosodic cues.

In sum, the studies presented in Chapters 4 and 5 clearly showed that multimodal cues provide important contrasting information which, together with contextual cues, facilitate the detection of irony by speakers. From a pragmatic point of view, we claim that in order to detect irony listeners need to attend to different levels of

contrasting information. Thus it is not only important to assess the discrepancies between propositional content and contextual information (e.g., uttering “That's great” after a negative event, like many authors have emphasized, e.g., Kreuz & Glucksberg, 1989; Kumon-Nakamura et al., 1995), but also the emotionally positive or negative attitude encoded by prosody and gesture and how this information interacts with propositional content (e.g., uttering “That's great” with a sad and sarcastic voice). Our results thus clearly fit the Contrast and Assimilation perspective. Even though previous results on irony detection had claimed that irony comprehension processes seem to rest on the existence of pragmatic contrasts between discourse context and the propositional content of the utterance (e.g., Colston 2002, Gerrig & Goldvarg, 2000; Ivanko & Pexman, 2003), as well as on the interaction between context and tone of voice (Woodland & Voyer. 2011; Voyer et al., 2016, Voyer & Vu, 2016)), the results of the experiments presented in Chapters 4 and 5 have shown that the way the utterances are multimodally produced are key to irony comprehension. Finally, our results also agree with the claims made within relevance theoretic accounts 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 in order to help the interlocutor to arrive at the ironic interpretation.

6.5. Visual cues are stronger than prosodic and contextual cues for verbal irony detection in both adult and child populations

Previous research on the audiovisual perspective has shown that prosodic characteristics of speech are complemented by gestural markers and that they can jointly convey a set of pragmatic meanings, such as prominence and focus marking (Hadar et al., 1983; Cavé et al., 1996; Krahmer & Swerts, 2007; Swerts & Krahmer, 2008; Dohen & Loevenbruck, 2009; Prieto et al., 2015), face-to-face grounding (Nakano et al., 2003), and question intonation (Srinivasan & Massaro 2003). Some of these studies have also shown that gestural information provides more conclusive evidence than intonation when interpreting the pragmatic content of a statement (Borràs-Comes et al., 2011; Goldin-Meadow, 2003; Holler & Wilkin, 2009; Krahmer & Swerts, 2004; Prieto et al., 2015; Swerts & Krahmer, 2005).

This dissertation has contributed to this line of research by providing new empirical evidence on the relative strength of the visual cues as compared to prosodic and contextual cues in verbal irony detection.

The results of the second experiment in Chapter 4 have shown that, when presented with mismatching audiovisual presentations of a set of ironic and sincere utterances (i.e. sincere audio stream combined with ironic video stream, and vice versa), participants relied more

strongly on visual information than on auditory information for detecting ironic intent (3.8 over 5 vs. 2.5 over 5). These results are consistent with previous findings on audiovisual speech processing which demonstrate that when exposed to information coming from auditory and visual sources, listeners tend to rely more on visual than on auditory information (e.g., McGurk et al., 1976; Gentilucci et al., 2005; Bristow et al., 2009; O'Shea, 2005). Our results also confirm results from studies investigating attitudinal or emotional understanding which clearly show that visual information is more important for communicative purposes than acoustic information (Mehrabian & Ferris, 1967; Swerts & Krahmer, 2005; Dijkstra et al., 2006).

Moreover, one of the critical findings of this dissertation is that gestural and prosodic cues could be stronger than contextual cues in signalling ironic intent. Results of the third experiment in Chapter 4 showed that, when participants were presented with incongruent discourse context and multimodal cues, they rated the perceived degree of irony of the target utterance differently depending on the availability of the audiovisual information. Specifically, while they relied more strongly on the discourse context when they only had access to auditory information, they strongly relied on the prosodic and gestural realization of the target sentence when they had access to audiovisual information. These results demonstrate that multimodal cues play a fundamental role in irony detection, especially when the realization of a sentence does not agree with the expectations triggered by the discourse context.

In sum, the results presented in this dissertation have contributed to clarifying the role of the gestural component to verbal irony detection in terms of strength. Even though some studies had pointed out the relevance of gestural cues in the production and detection of verbal irony (Attardo et al. 2003, 2011; Caucci et al. 2012; Rockwell, 2000, Padilla, 2004; Poggi, 2007), our results clearly show that visual information can be even more critical than prosodic and contextual information in the detection of verbal irony, This constitutes a novel contribution to the literature on verbal irony comprehension.

6.6. Multimodal cues facilitate early detection of irony in children

Previous literature on the development of verbal irony detection has shown that the ability to detect a speaker's ironic intent develops between the ages of 5 and 11 (e.g., de Groot et al., 1995; Creusere, 2000; Filippova & Astington, 2008; Harris & Pexman, 2003; Nakassis & Snedeker, 2002; among many others). So far, most studies agree in considering that by 8-9 year of age, children are successful in identifying ironic intent through prosodic cues (Capelli et al., 1990; Milosky & Ford, 2009), and recent studies showed that development of irony appreciation and development of emotion appreciation (i.e., empathy skills) are closely linked (Nicholson et al., 2013). However, to date no experimental studies had taken into account the visual component in the perception of irony by children. Nicholson et al.'s (2013) experimental set-up

included visual information, but it was provided by puppets. In the fourth study presented in this dissertation (Chapter 5), we presented a follow-up of Nicholson et al.'s (2013) irony perception experiment to assess the developmental benefits of visual cues by using video recordings of real humans and by testing children's detection skills in three different age groups, namely 5-, 8- and 11-year-olds. Our hypothesis was that visual cues of verbal irony would facilitate irony detection in 5-year-old children, an age where previous studies reported they had difficulties in such tasks with no gestural information.

The results presented in Chapter 5 first revealed a clear developmental pattern in children's irony detection skills, as the average irony perception scores increased with age, which is consistent with the previous literature (e.g., de Groot et al., 1995; Creusere, 2000; Filippova & Astington, 2008; Harris & Pexman, 2003; Nakassis & Snedeker, 2002). Second, crucially, results showed that gestural cues facilitated the detection of irony. In contrast with the previous literature, 5 year-old children started to detect irony in the condition where participants had access to the emotional cues (or affective attitude, in Yus'(2016) terms) expressed in gesture which strongly contradicted the positive valence of the utterance. These results crucially contrast with Nicholson et al.'s (2013) results, as they found that 6- to 7-year-old children displayed near-zero accuracy in detecting irony using a similar task. These results also suggest that research on the development of irony detection needs to address irony as a

multimodal affair and incorporate a more fine-grained control of the visual and prosodic information which accompanies ironic utterances.

Thus our results expand on previous results on the detection of irony from a developmental point of view. They suggest that the perception of irony can appear at very early stages of development provided that children have access to strongly mismatching prosodic and gestural cues conveying ironic intent. These findings are compatible with those of Armstrong et al. (2014) and Hübscher et al. (2016), where visual cues facilitated 4- to 6-year-old children's performance in detecting belief state meanings such as incredulity or uncertainty.

In sum, our results clearly show that children are especially sensitive to emotional expressions conveyed by prosodic and gestural cues, and that they actively use them to make judgments about a speaker's intent. These results are in line with previous studies which experimentally showed a strong relationship between irony appreciation in children and their empathy skills, this is, their ability to detect emotions in others (Nicholson et al., 2013). Moreover, our results are in line with the growing consensus that pragmatic gestures act as bootstrapping devices in language (and specifically pragmatic) development (e.g., McNeill, 1998; McNeill, Cassell & McCullough, 1994; Kelly, 2001; Butcher & Goldin-Meadow, 2000).

In conclusion, this dissertation has contributed to clarifying some aspects of the role of prosodic and visual cues in verbal irony production, perception and development. In a nutshell, this dissertation has provided more evidence and expanded previous research showing how a wide set of prosodic and gestural features appear in the production of ironic remarks, and that these multimodal cues could appear temporally aligned or manifest themselves independently from each other. From a perception point of view, it has also been shown that prosodic and gestural mismatches with the propositional content are a key factor in signaling ironic intent. Moreover, this dissertation has provided novel empirical evidence of the privileged effects of multimodal — and especially gestural—cues in comparison with contextual cues, both in adult and child populations. We believe that the reason why multimodal cues are so strong in the communication of irony is that the two are strong indicators of speaker intent. In general, the findings presented in this dissertation lead us to suggest that the study of multimodal cues of verbal irony and a more fine-grained analysis of their interplay with verbal and contextual cues should be at the core of any pragmatic or psycholinguistic account of this complex pragmatic phenomenon.

7. References

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

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

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

Alvarado Ortega, M.B., & Padilla, X. (2010): Being polite through irony. In Dale April & Lidia Rodríguez (Eds.): *Dialogue in Spanish. Studies in functions and contexts. Dialogue Studies*, (pp. 55-68). Amsterdam: John Benjamins.

Levinson, S.C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society*, 369: 20130302.

Anolli, L., Ciceri, R., & Infantino, M.G. (2002). "From 'blame by praise' to 'praise by blame': Analysis of vocal patterns in ironic communication". *International Journal of Psychology*, 37, 266-276.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Attardo, S., Wagner, M. M., & Urios-Aparisi, E. (2013). "Prosody and Humour". In S. Attardo, M. M. Wagner, & E. Urios-Aparisi (Eds.), *Prosody and Humor* (pp. 189-201). Amsterdam: John Benjamins.

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

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

Attardo, S. (2001) *Humorous Texts: A Semantic and Pragmatic Analysis*. Berlin: Mouton de Gruyter.

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32: 793-826.

Audacity Team (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0 retrieved April 20th, 2014, from <http://audacity.sourceforge.net/>

Austin, J.L. (1962). *How to Do Things with Words*. Oxford: Clarendon.

Beattie, G. , & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123, 1-30.

Becerra Valderrama, M.I., & Igoa González, J.M. (2014). “La prosodia en la ironía verbal”. In M. A. Penas Ibáñez (Ed.), *Panorama de la fonética española actual* (pp. 453-486). Madrid: Arco Libros.

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

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

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

Borràs-Comes, J., & Prieto, P. (2011). 'Seeing tunes'. The role of visual gestures in tune interpretation. *Journal of Laboratory Phonology*, 2(2), 335-380.

Borràs-Comes, J., Roseano, P., Vanrell, M.d.M., & Prieto, P. (2011). “Perceiving uncertainty: facial gestures, intonation, and

lexical choice”. In *Proceedings of GESPIN 2011*. Bielefeld, Germany.

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

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., & Baillet, S., (2009). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, 21, 905-921.

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

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

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

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

Bryant, G.A., & Fox Tree, J.E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, 17(2), 99-117.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to twoword speech: When hand and mouth

come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235-257). New York: Cambridge University Press.

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

Carston, R. (2002). Linguistics meaning, communicated meaning and cognitive pragmatics. *Mind and Language, 17*, 127-148.

Cartmill, E.A., Ece Demir, Ö., & Goldin-Meadow, S. (2012). *Studying gesture, Research Methods in Child Language: A Practical Guide*. Blackwell Publishing Ltd.

Caucci, G. M., & Kreuz, R. J. (2012). Social and paralinguistic cues to sarcasm. *Humor: International Journal of Humor Research, 25*, 1-22.

Cavé, C., Guaitella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. In *Proceedings of the 4th International Conference on Spoken Language Processing* (pp. 2175-2179). Philadelphia, USA.

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *Journal of the Acoustical Society of America, 126*(3), 1394-1405.

- Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication, 50*, 366-381.
- Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General, 113*(1), 121-126.
- Clark, B., & Lyndsey, G., (1990). Intonation, grammar and utterance interpretation. In *UCL Working Papers in Linguistics 2* (pp. 32-51).
- Climie, E. A., & Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *Journal of Cognitive Development, 9*, 257-285.
- Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics, 34*, 111-142.
- Colston, H. L. (1999). "Not good" is "bad", but "not bad" is not "good": An analysis of three accounts of negation asymmetry. *Discourse Processes, 28*, 237-256.
- Colston, H. L., & Keller, S. B. (1998). You'll never believe this: Irony and hyperbole in expressing surprise. *Journal of Psycholinguistic Research, 27*, 499-513.
- Colston, H. L., & O'Brien, J. (2000). Contrast of kind vs. contrast of magnitude: The pragmatic accomplishments of irony and hyperbole. *Discourse Processes, 30*, 179-199.

Cosnier, J., (1991). Les gestes de la question. In C. Kerbrat-Orecchioni (Ed.), *La question, Presses Universitaires de Lyon* (pp. 163-171). Lyon, France.

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

Curcó, C. (2000). Irony: Negation, echo and metarepresentation. *Lingua, 110*, 257-280.

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

Cvejic E., Kim J., & Davis, C. (2012). Recognizing prosody across modalities, face areas and speakers: Examining perceivers' sensitivity to variable realizations of visual prosody. *Cognition, 122*(3), 442-453.

Cvejic, E., Kim, J., & Davis, C. (2010). Abstracting visual prosody across speakers and face areas, In *Proceedings of AVSP 2010*, Hakone, Kanagawa, Japan.

De Brabanter, P. (2010). "Uttering sentences made up of words and gestures". In Soria, B. & E. Romero (Eds.), *Explicit Communication: Robyn Carston's Pragmatics* (pp. 199-216). Basingstoke: Palgrave MacMillan.

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

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

Dews, S. & Winner, E. (1995). "Muting the meaning: A social function of irony". *Metaphor and Symbolic Activity*, *10*, 3-19.

Dijkstra, C., Krahmer E., & Swerts, M. (2006). Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. In *Proceedings of the 3rd International Conference on Speech Prosody*.

Dohen, M., & Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, *52*(2-3), 177-206.

Eisterhold, J., Attardo, S., & Boxer, D. (2006). Reactions to irony in discourse: Evidence for the Least Disruption Principle. *Journal of Pragmatics*, *38*(8), 1239-1256.

Elvira-García, W., Roseano, P., & Fernández Planas, A. M. (2017). Prosody as a cue for syntactic dependency. Evidence from dependent and independent clauses with subordination marks in Spanish. *Journal of Pragmatics*, *109*, 29-46.

Elvira-García, W. (2016). *La prosodia de las construcciones insubordinadas conectivo-argumentativas en español* (PhD dissertation). Universitat de Barcelona, Barcelona, Spain.

Escandell-Vidal, M.V., & Leonetti, M. (2014). Fronting and Irony in Spanish. In A. Dufter & A. Octavio de Toledo (Eds.), *Left Sentence Peripheries in Spanish: Diachronic, Variationist and Typological Perspectives 3* (pp. 309-342). Amsterdam: John Benjamins.

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

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

Escandell-Vidal, M.V. (2006). *Introducción a la pragmática*. Barcelona: Ariel.

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

Escudero, D., Aguilar, L., Vanrell, M del M., & Prieto, P. (2012). Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labelling system. *Speech Communication 54*(4), 566-582.

Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A.

Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 41-56). Amsterdam/Philadelphia: John Benjamins.

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

Fleiss, J. L., (1981). *Statistical methods for rates and proportions*. London: John Wiley.

Forceville, C. (2014). Relevance Theory as model for analysing visual and multimodal communication. In D. Machin (Ed.), *Visual Communication* (pp. 51-70). Berlin: Mouton de Gruyter.

Fretheim, T. (2002). Intonation as a constraint on inferential processing. In *Proceedings of Speech Prosody International Conference*. Aix-en-Provence, France.

Gale, C., & Monk, A. (2000). Where am I looking? The accuracy of video-mediated gaze awareness. *Perception & Psychophysics, 62*, 586-595.

Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*, 66-75.

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

Gibbs, R. W. (2012). Are ironic acts deliberate? *Journal of Pragmatics, 44*, 104-115.

Gibbs, R. W. (2000). Metarepresentations in staged communicative acts. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective [Vancouver Studies in Cognitive Science 10]* (pp. 389-410). New York: Oxford University Press.

Gibbs, R. W. (1994). *The poetics of mind: figurative thought, language, and understanding*. New York: Cambridge University Press.

Giora, R. (1995). On irony and negation. *Discourse Processes, 19*(2), 239-264.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting gaze disengages the environment and facilitates remembering. *Memory & Cognition, 26*, 651-658.

Goldin-Meadow, S. (2003). *Hearing Gesture: How our hands help us think*. Cambridge: Harvard University Press.

González-Fuente, S., Prieto, P., & Noveck, I. (2016). A fine-grained analysis of the acoustic cues involved in verbal irony recognition in

French. In *Proceedings of the Speech Prosody 2016* (pp. 902-906). Boston, MA (USA).

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

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

Grassmann, S., & Tomasello, M. (2009). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*, 252-263.

Grice, P. (1975). Logic and conversation. In P. Cole, J.L. Morgan (Eds.), *Speech Acts* (pp.41-58). New York: Academic Press.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, 1-14.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Haiman, J. (1998). *Talk is cheap: Sarcasm, Alienation, and the Evolution of Language*. New York: Oxford University Press.

Hadar, U., Steiner, T. J., Grant, E. C., & Clifford, R. F. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech*, *26*, 117-129.

Hancock, J. T. (2004). Ironic use in face-to-face and computer-mediated conversation. *Journal of Language and Social Psychology, 23*, 447-463.

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

Holler, J., & Wilkin, K. (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, J. (2006). Constructing a context with intonation. *Journal of Pragmatics, 38*, 1542-1558.

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

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

Hutcheon, L. (1996) *Irony's Edge: The Theory and Politics of Irony*. London, England: Routledge.

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

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

Iverson, J.M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science* 16 (5), 367-371.

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

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. UK: Cambridge University Press.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, 28(2), 325-349.

Krahmer, S., & Swerts, M. (2009). Audiovisual Prosody—Introduction to the Special Issue. *Language and Speech*, 52, 129-133.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396-414.

Krahmer, E., & Swerts, M. (2004). More about brows. In Z. Ruttkay & C. Pelachaud (Eds.), *From brows to trust: Evaluating*

embodied conversational agents (pp. 191-216). Dordrecht: Kluwer Academic Press.

Kreuz, R. J. (1996). The use of verbal irony: Cues and constraints. In J. S. Mio and A. N. Katz (Eds.), *Metaphor: Implications and Applications* (pp. 23-38). Mahwah, NJ: Lawrence Erlbaum.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Journal of Experimental Psychology: General* 118, 372-386.

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

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

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

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

Leggit, J. S., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse Processes*, 29, 1-24.

Llisterri, J., & Mariño, J. (1993). Spanish Adaptation of SAMPA and automatic phonetic transcription. In *Informe SAMPA/UPC/001/VI*. Madrid.

Loevenbruck, H., BenJannet, M., D'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcasm speech in French: slower, higher, wider. In *Proceedings of Interspeech 2013*. Lyon, France.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, D. (1998). Speech and gesture integration. *New Directions for Child Development*, 79, 11-27.

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

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*, 248-252.

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

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

Muñoa Barredo, I. (1997). Pragmatic Functions of Code-Switching among Basque-Spanish Bilinguals. University of Urbana-Champaign, Chicago. Retrieved April 28th, 2012 from <http://webs.uvigo.es/ssl/actas1997/04/Munhoa.pdf>.

Nakano, Y. I., Reinstein, G., Stocky, T., & Cassell, J. (2003). Towards a model of face-toface grounding. *Annual Meeting of the Association for Computational Linguistics (ACL)*, 553-561.

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

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

Nonhebel, A., Crasborn, O., & van der Kooij, E. (2004). Sign language transcription conventions for the ECHO project. Retrieved October 25th, 2012 from http://www.let.kun.nl/sign-lang/echo/docs/transcr_conv.pdf.

O'Shea, M. (2005). *The Brain: A Very Short Introduction*. Oxford: Oxford University Press.

Padilla, X. (2011). ¿Existen rasgos prosódicos objetivos en los enunciados irónicos? *Oralia*, 14, 203-224.

Padilla, X. (2009): Marcas acústico-melódicas: el tono irónico. In L. Ruiz Gurillo & X. Padilla, X. (Eds.) *Dime cómo ironizas y te diré quién eres* (pp. 371-390). Frankfurt: Peter Lang.

Padilla, X. (2004). El tono irónico. Estudio fonopragmático. *Español Actual*, 81: 85-98.

Pelachaud C., & Poggi I. (Eds.) (2001): Multimodal Communication and Context in Embodied Agents. In *Proceedings of the Workshop W7 at the 5th International Conference on Autonomous Agents*, Montreal, Canada, May 29, 2001.

Poggi, I. (2007). *Mind, hands, face and body. A goal and belief view of multimodal communication*. Berlin: Weidler.

Poggi, I (2006). *Le Parole Del Corpo. Introduzione Alla Comunicazione Multimodale*. Roma: Carocci Editore.

Poggi, I., & Pelachaud, C. (2000). Emotional meaning and expression in animated faces. In A. Paiva (Ed.), *Affect in interactions*. Berlin: Springer Verlag.

Poggi, I., Pezzato, N., & Pelachaud, C. (1999). *Gaze and its meaning in animated faces*. In *The Eighth International Workshop on the Cognitive Science of Natural Language Processing (CSNLP-8) "Language, Vision and Music"*. Galway, Ireland.

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, 17, 286-290.

Pierrehumbert, J. B. (1980). *The Phonetics and Phonology of English Intonation*. Ph.D. Dissertation, Massachusetts Institute of Technology.

Potts, C. (2005). Lexicalized intonational meaning. In S. Kawahara (Ed.), *University of Massachusetts Occasional Papers 30* (pp. 129-146). Amherst, MA: GLSA.

Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics*, 49(1): 41-54.

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

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

Prieto, P., & Roseano, P. (Eds.) (2010). *Transcription of Intonation of the Spanish Language*. München: Lincom Europa.

Randolph, J. J. (2008). Online Kappa Calculator. Online computer program. Retrieved September 26th, 2012. From <http://justus.randolph.name/kappa>.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29, 483-495.

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

Ruiz Gurillo, L. (2013): Narrative strategies in Buenafuente's humorous monologues. In , L. Ruiz Gurillo & M. B. Alvarado Ortega (Eds.), *Irony and humor: From Pragmatics to Discourse* (pp. 107-140). Amsterdam: John Benjamins.

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

Ruiz Gurillo, L., & Alvarado Ortega, M. B. (Eds.) (2013). *Irony and Humor. From pragmatics to discourse*. Amsterdam: John Benjamins.

Searle, J. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Scharrer, L., Christmann, U., & Knoll, M. (2011). Voice Modulations in German Ironic Speech. *Language and Speech*, 54(4), 435-465.

Smoski, M. J., & Bachorowski, J. (2003). Antiphonal laughter between friends and strangers. *Cognition & Emotion*, 17, 327-340.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford : Blackwell.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.B., & Noveck, I. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, 63(1), 25-39.

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

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

Swerts, M., & Krahmer, E. (2008). Facial expressions and prosodic prominence: Comparing modalities and facial areas. *Journal of Phonetics*, 36(2), 219-238.

Swerts, M., & Kraemer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53, 81-94.

Tabacaru, S., & Lemmens, M. (2014). Raised eyebrows as gestural triggers in humour: the case of sarcasm and hyper-understanding. *European Journal of Humour Research*, 2, 11-31.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from non-ironic. *Journal of Pragmatics*, 32, 1777-1806.

Vanek, C., & McDaniel, S. (2006). SurveyGizmo Online Survey Software. Computer program. Retrieved 14 December 2012 from <http://www.surveygizmo.com/>

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

Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24, 64-69.

Vengaliene, D. (2011). *Irony within the scope of conceptual blending in lithuanian and american on-line news headlines: a comparative analysis* (PhD dissertation). Vilnius University, Vilnius, Lithuania.

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

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

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

Wharton, T. (2009). *Pragmatics and Nonverbal Communication*. Cambridge: Cambridge University Press.

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

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

Wilson, D., & Sperber, D. (2012). Explaining Irony. In D. Wilson & D. Sperber (Eds.), *Meaning and Relevance* (pp. 123-145). Cambridge: Cambridge University Press.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, 38, 1557-1579.

Wilson, D., & Sperber, D. (1993). Linguistic form and relevance". *Lingua* 90, 1/2, vol. 2, pp.1-25.

Wilson, D. & Sperber, D. (1992). On verbal irony. *Lingua*, 87, 53-76.

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

Winner, E. Windmueller, G., Rosenblatt, E., Bosco, L., Best, E., & Gardner, H. (1987). Making Sense of Literal and Nonliteral Falsehood. *Metaphor and Symbolic Activity*, 2(1), 13-32.

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

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

Yus, F. (2003). Humour and the search for relevance. *Journal of Pragmatics*, 35(9), 1295-1331.

Yus, F. (1997). La teoría de la relevancia y la estrategia humorística de la incongruencia-resolución. *Pragmalingüística*, 3-4, 497-508.

8. Appendices

Appendix A

English translations of the discourse contexts used for Experiments 1, 2, and 3 with (a) literal-biased and (b) ironic-biased contextual paths (Experiment 3) and (c) an ambiguous contextual path (Experiments 1 and 2).

Discourse context 1

John and Peter are about the same age and live in the same apartment block. Today they have met on entering the building. John calls the lift. While waiting, they make small talk. Both express the hope that someday that week a gas technician will come to review the gas pipes in the building, since all the tenants have experienced disturbing problems with the gas in the last few weeks.

(a) Literal-biased condition

When the lift arrives, both of you see a note attached to the mirror which says that this afternoon from 3PM to 4PM a technician will come to check the gas pipes. John looks at Peter and says:

Perfect

(b) Ironic-biased condition

When the lift arrives, they both see a note attached to the mirror which says that the gas company has communicated to the president of the tenants' association that they will not be able to check the gas for another three weeks. John looks at Peter and say:

Perfect

(c) Ambiguous condition

Then John says to Peter that he would like to be at home when the gas man comes, but depending on what time he comes he may not be at home. When the lift arrives, they both see a note attached to the mirror which says that the gas man will be coming tomorrow between 3PM and 4PM to check the pipes. John takes out his agenda, consults his plans, and tells Peter:

Perfect

Discourse context 2

Laura and Julia live on the same street and are about the same age. They know each other only by sight. Today they have met by chance at the theater. Having greeted each other, they are now waiting for the show to start, seated side by side. While waiting, they make small talk.

(a) Literal-biased condition

Before the play starts, a theater employee announces over the PA system that because today is International Theater Day, at the end of the show everyone in the audience will receive a free ticket for another play. Laura looks at Julia and says to her:

Fantastic

(b) Ironic-biased condition

Before the play starts, a theater employee announces over the PA system that unfortunately the performance has to be canceled because the leading actress has lost her voice. Laura looks at Julia and says to her:

Fantastic

(c) Ambiguous condition

Laura tells Julia that she loves this play and this theater company but that she is very sad because of the illness of the leading actor. Julia tells Laura that today she has heard on a radio broadcast that tonight he might be able to perform, but the company won't confirm it until a few minutes before the play starts. Then a signal bell rings, and the definitive cast for tonight's performance appears on the two flanks of the stage. Laura looks at the stage, then looks at Julia and says:

Fantastic

Discourse context 3

Mark and Robert are both members of the same gym and always meet each other at spinning class on Tuesday evening. Today is Tuesday, the class is about to start and they are mounted on static bikes side by side. While waiting for the arrival of the teacher instructor, they are making small talk.

(a) Literal-biased condition

When the instructor arrives, he greets everyone and puts on a CD of brisk, thumping music. He then sits down on his static bike, and starts pedaling vigorously, while urging the class into action. Robert looks at Mark and says:

Today we are really going to sweat

(b) Ironic-biased condition

When the instructor arrives, he greets everyone and puts on a CD. However, unlike what he usually plays, the music today is soft and melodious. The instructor sits down on his static bike and begins to pedal very slowly. Robert looks at Mark and says:

Today we are really going to sweat

(c) Ambiguous condition

Marc comments to Robert that he doesn't know which instructor will be leading the spinning session today: Andrew, whose sessions tend to be light and easy-going, or Michael, who makes them sweat a lot. Robert tells Marc that he knows which instructor they are going to have today because he has just met him in the changing room. Intrigued, Marc asks Robert, "So? Are we going to get a real workout today?" Robert looks at Mark and says:

Today we are really going to sweat

Discourse context 4

Teresa and Martha are neighbors. Today they have met in the lobby of the building, and Martha calls the lift. While waiting for it to arrive, they make small talk.

(a) Literal-biased condition

When the lift arrives, they enter, and as the lift ascends they begin to reminisce about the old days before the lift was installed and they had to walk up the stairs, which kept them in better physical condition. Teresa looks at Martha and says:

I used to like walking up the stairs

(b) Ironic-biased condition

When the lift arrives, they enter, and as the lift ascends they begin to recall what a nuisance it was before the lift was installed and they

had to walk up the stairs carrying heavy shopping bags. Teresa looks at Martha and says:

I used to like walking up the stairs

(c) Ambiguous condition

When the lift arrives, they both go inside and encounter their neighbor Susan, who is coming up from the parking garage. They greet each other and then Susan comments to them that she loves lifts, since having to walk up all the stairs laden with shopping bags, as they used to have to do, was terrible. Martha shakes her head in disagreement with Susan, since climbing all those stairs used to keep them healthier and more fit. Susan turns smiling to Teresa and asks, “How about you, Teresa? Did you use to like walking up all those stairs?” Teresa looks at Martha and Susan and says:

I used to like walking up the stairs

Appendix B

Discourse Contexts 2 and 3 used in the DCT (for Discourse Context 1 see Table 9 in section 5.2.1). The original Catalan version of the script is shown in italics with the English translation below.

Discourse Context 2

Situational prompt (a negative event)	Prosodic-gestural condition	Target sentence (a positive comment)
<p><i>Estàs jugant un partit de futbol i un amic teu llença un penal i el falla.</i></p> <p>You are playing football with a friend of yours. He/she misses a penalty kick.</p>	<p><u>Matching</u></p> <p><i>Llavors, li dius al teu amic/ga amb entusiasme exagerat:</i></p> <p>You say to your friend with exaggerated enthusiasm:</p>	<p><i>‘Que ben fet!’</i></p> <p>‘Well done!’</p>
	<p><u>Weakly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga amb emoció continguda:</i></p> <p>You say to your friend with restrained emotion:</p>	
	<p><u>Strongly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga ofensivament:</i></p> <p>You say to your friend in a critical manner:</p>	

Discourse Context 3

Situational prompt (a negative event)	Prosodic-gestural condition	Target sentence (a positive comment)
<p><i>Estàs prenent un refresc amb un amic teu. De sobte, li cau el got a terra i es trenca.</i></p> <p>You are having a drink with a friend of yours. Suddenly, his/her glass falls and smashes on the floor.</p>	<p><u>Matching</u></p> <p><i>Llavors, li dius al teu amic/ga amb entusiasme exagerat:</i></p> <p>You say to your friend with exaggerated enthusiasm:</p>	<p><i>‘Que ben fet!’</i></p> <p>‘Well done!’</p>
	<p><u>Weakly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga amb emoció continguda:</i></p> <p>You say to your friend with restrained emotion:</p>	
	<p><u>Strongly mismatching</u></p> <p><i>Llavors, li dius al teu amic/ga ofensivament:</i></p> <p>You say to your friend in a critical manner:</p>	

Appendix C

Series of drawings used in PowerPoint slides illustrating the 12 discourse contexts in the irony detection task. Slide 1 introduces the two characters, slides 2 and 3 accompany the situational prompt, and slide 4 includes an embedded video in which a person gives the reaction comment which the child subject must judge as ‘nice’ or ‘mean’. In the first nine sequences, the situation described has a negative outcome. The reactions to the first six are ironically congratulatory, while the following three reactions are congruently critical. The last three sequences depict situations with positive outcomes. The reactions to all three of these events are appropriately positive.

Slide 1: presentation	Slides 2 and 3: situational prompt		Slide 4: utterance
