



Universidad de Valladolid



**PROGRAMA DE DOCTORADO EN ESTUDIOS INGLESES
AVANZADOS**

TESIS DOCTORAL:

**HUMOUR PRODUCTION IN FACE-TO-FACE
INTERACTION: A MULTIMODAL AND COGNITIVE
STUDY**

**Presentada por Marta Buján Navarro para optar al
grado de
Doctor/a por la Universidad de Valladolid**

**Dirigida por:
Dra. Marisol Velasco Sacristán**

To Grandma, always.

“Humor is the great thing, the saving thing after all. The minute it crops up, all our hardnesses yield, all our irritations, and resentments flit away, and a sunny spirit takes their place.”

Mark Twain – *What Paul Bourget Thinks of Us*

«Para combatir tantos males que afectan al mundo tan sólo dispongo de un arma, que es el humor».

Miguel Gila - *Yo muy bien, ¿y usted?*

Acknowledgements

This thesis would not have been possible without the input and support of many people. First and foremost, I would like to thank my supervisor, Dr. Velasco-Sacristán, for her always insightful comments and for giving me the freedom to pursue my own research interests. My appreciation also goes to the colleagues at IMFAHE Foundation and their International Mentoring Programme, for giving me the opportunity to benefit from the advice of top-class researchers and for awarding me the grant for my first international research stay at UT Austin. I would like to express my gratitude to all those that generously offered to read, review and proofread my work, to make suggestions and improvements, or simply to support me throughout in any way they could: Josep, Daniel, Pati, Lara, Paco, Olaia, Ernesto, Germán, and Félix.

Needless to say, I owe special thanks to my family. To Santi, my dear husband, who bore with me through this long and difficult process, even when he did not really see the point of me putting so much effort into it. And to my loving son, Santi, who has given up so much holiday time to be dragged across the world by mummy for her research stays.

Finally, I have no words to express my utmost gratitude to Dr. Juan José Colomina Almiñana. To say that he has played a key role for me to complete this journey would be an understatement. I would really not have seen this thesis through had it not been for his invaluable and unwavering patience, guidance, and support. Not only did he enhance my dissertation with his insightful and rigorous views; he also walked me through the ups and downs of the process, always having the right thing to say. Most importantly, he and his lovely family—Nicole and little Ausiàs—went out of their way to make me and my son feel at home in Austin. I am forever indebted to them.

Resumen

El humor es una de las formas de comunicación más complejas que existen (Veale, Brône, & Feyaerts, 2015). Algunas teorías lingüísticas sobre el humor tienen un enfoque semántico-pragmático, como la *Semantic Script Theory of Humour* (Raskin, 1984) o la *General Theory of Verbal Humour* (Attardo, 2001). Otras se inscriben en la Teoría de la Relevancia (Yus, 2016) y las hay también con una perspectiva más cognitiva (Giora, 1991, 2015; Coulson & Okley, 2005; Veale, Feyaerts, & Brône, 2006). Se han realizado varios estudios sobre los marcadores multimodales de la ironía o el sarcasmo, cuyos resultados son dispares (Attardo, Eisterhold, Hay, & Poggi, 2003; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011). Sin embargo, el humor no irónico ha sido objeto de menos estudios. Además, la mayor parte de los análisis se circunscriben al humor ensayado, con pocos estudios sobre el humor producido de forma espontánea (Bryant, 2010, Feyaerts, 2013; Tabacaru, 2014, etc.) y menos aún que conjuguen la perspectiva multimodal con la cognitiva.

En esta tesis se analizan 14 entrevistas extraídas de *The Late Show with Stephen Colbert* con vistas a explicar la comunicación espontánea del humor desde el punto de vista multimodal y cognitivo. Los enunciados se han identificado como humorísticos cuando el público reaccionaba riendo. El análisis multimodal se ha realizado en ELAN, con cinco niveles de anotaciones: transcripción, tipo de humor (Feyaerts et al., 2010), mecanismo conceptual subyacente (Croft & Cruse, 2004), gestos y prosodia. El estudio prosódico se ha llevado a cabo con Praat, a fin de determinar si había un mayor contraste prosódico en enunciados humorísticos. Los resultados muestran que los mecanismos multimodales y cognitivos analizados en el presente estudio no difieren entre enunciados humorísticos y no humorísticos.

Abstract

Humour is one of the most complex instances of communication (Veale et al., 2015). Different linguistic humour theories can be found in the literature, from Raskin's Semantic Script Theory of Humour (1984), or Attardo's General Theory of Verbal Humour (2001), adopting a semantic-pragmatic approach, to Yus's Relevance Theory account of humour (2016), and falling under a more cognitive perspective (Giora, 1991, 2015; Coulson & Oakley, 2005; Veale et al., 2006). In addition, many studies have been conducted on the markers of irony or sarcasm, with conflicting results (Attardo, Eisterhold, Hay, & Poggi, 2003; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011). Less attention has been devoted to non-ironical humour and most of the literature is limited to staged humour, with just a handful of studies focusing on spontaneous humour (Bryant, 2010, Feyaerts, 2013; Tabacaru, 2014, etc.). To my knowledge, no previous study combines a multimodal and cognitive perspective.

This dissertation revolves around the analysis of humorous instances selected in 14 interviews from *The Late Show with Stephen Colbert*. The aim of the analysis is twofold: to account for spontaneous humorous communication from a multimodal and cognitive perspective. Humorous instances have been identified based on laughter elicited in the audience. The interviews have been uploaded to ELAN and humorous utterances annotated on five tiers: transcription, humour type (Feyaerts et al., 2010), underlying construal mechanisms (Croft & Cruse, 2004), gestures and prosody. Prosodic features have been extracted with Praat in order to look into prosodic contrast between humorous and non-humorous utterances. The humorous utterances identified in the sample rely on the same cognitive mechanisms and multimodal resources analysed than non-humorous utterances.

Table of contents

Acknowledgements	iii
Resumen.....	iv
Abstract.....	v
Table of contents	vi
List of tables.....	xv
List of figures.....	xvi
List of acronyms.....	xviii
Chapter 1	1
Introduction: Humour in face-to-face interaction. A multimodal and cognitive study	1
1.1 Scope of the study and rationale	1
1.2 Humour studies.....	2
1.3 Multimodality.....	3
1.3.1 Multimodal markers of humour	4
1.4 Cognitive linguistics	5
1.5 Research questions	6
1.6 Structure of the dissertation	7
Chapter 2	12
Review of humour studies.....	12
2.1 Introduction	12
2.2 The definition of humour	13

2.3 Traditional humour studies.....	13
2.3.1 Incongruity theories	14
2.3.2 Superiority theories.....	15
2.3.3 Tension release theories	15
2.4 Humour and multimodality	16
2.4.1 Prosody, gestures, and humour.....	18
2.4.1.1 Prosody and humour.....	18
2.4.1.2 Prosody and irony	20
2.4.1.3 Contrast as a marker.....	25
2.4.1.4 Gestures and humour.....	28
2.4.1.5. Markers, indices, and indicators	30
2.5 Linguistic humour studies	32
2.5.1 Semantic-pragmatic approach.....	37
2.5.1.1 Raskin's Semantic Script Theory of Humour.....	37
2.5.1.2 Attardo's General Theory of Verbal Humour	39
2.5.2 Relevance theory applied to humour studies.....	41
2.5.3 Humour and cognitive linguistics	46
2.5.3.1 Humour in interaction from a cognitive perspective.....	52
2.6 Conclusion	57
Chapter 3	60
Multimodality	60
3.1 Introduction	60

3.2 Multimodality.....	61
3.3 Multimodality and semiotics.....	62
3.3.1 Semiotics vs semiology.....	64
3.3.2 Multimodal social semiotics.....	65
3.4 Multimodality and pragmatics.....	66
3.5 Multimodality and cognitive linguistics.....	70
3.6 Multimodality in face-to-face interaction.....	71
3.6.1 Prosody.....	72
3.6.1.1 Prosody and its parts.....	73
3.6.1.1.1 Pitch.....	74
3.6.1.1.2 Pauses.....	75
3.6.1.1.3 Pitch accent and stress.....	76
3.6.1.2. Prosody and its uses.....	77
3.6.1.3 Prosody and gestures.....	80
3.6.2 Gestures and speech.....	82
3.6.2.1 The growth point hypothesis.....	86
3.6.2.2 Types of gestures.....	87
3.6.2.2.1 Early classifications.....	87
3.6.2.2.2 Kendon's continuum.....	88
3.6.2.2.3 McNeill's classification.....	90
3.6.2.3 Functions of gestures.....	91
3.7 Conclusion.....	93

Chapter 4	96
Cognitive linguistics.....	96
4.1 Introduction	96
4.2 Key concepts in cognitive linguistics	98
4.2.1 Mental spaces	100
4.2.2 Conceptual mapping.....	102
4.2.3 Image schemas	102
4.2.4 Construal mechanisms	104
4.2.4.1 Attention / salience	106
4.2.4.1.1 Selection	106
4.2.4.1.1.1 Profiling.....	106
4.2.4.1.1.2 Metonymy	108
4.2.4.1.2 Scope of attention	109
4.2.4.1.2.1 Scope of predication	109
4.2.4.1.2.3 Accessibility	110
4.2.4.1.3 Scalar adjustment	110
4.2.4.1.4 Dynamic attention	111
4.2.4.1.4.1 Fictive motion.....	111
4.2.4.2 Judgement / comparison	112
4.2.4.2.1 Categorization.....	112
4.2.4.2.2 Metaphor.....	113
4.2.4.2.3 Figure-ground alignment.....	114

4.2.4.3 Perspective / situatedness	114
4.2.4.3.1 Viewpoint	115
4.2.4.3.2 Deixis	115
4.2.4.3.3 Subjectivity / objectivity	116
4.2.4.4 Constitution / Gestalt	117
4.2.5 Conceptual metaphor	118
4.2.5.1 Multimodal metaphor	120
4.3 Current Discourse Space	121
4.4 Cognitive linguistics and semiotics	123
4.5 Cognitive semiotics and multimodality	125
4.6 Conclusion	128
Chapter 5	131
Material and methods	131
5.1 Introduction	131
5.2 Sample	131
5.3 Annotation tools	150
5.3.1 ELAN	150
5.3.2 Praat	155
5.3.3 Segmentation and types of utterance	157
5.4 Humour types	159
5.4.1 Humour types	160
5.4.1.1 Situational humour	160

5.4.1.2 Narrative joke	161
5.4.1.3 Puns	161
5.4.1.4 Irony	162
5.4.1.5 Sarcasm	163
5.4.1.6 Understatement.....	163
5.4.1.8 Teasing	164
5.4.1.9 Hyper-understanding.....	165
5.4.1.10 Misunderstanding	166
5.4.1.11 Parody.....	166
5.4.1.12 Register humour.....	168
5.4.1.13 Anecdote	168
5.4.1.14 Absurd humour.....	169
5.4.1.15 Insider humour	169
5.4.1.16 Joint fantasy	170
5.4.1.17 Stereotype humour.....	170
5.4.1.18 Meta-humour	171
5.4.1.19 Self-mockery	171
5.4.1.20 Self-glorification.....	172
5.4.1.21 Gender humour	172
5.4.1.22 Sexual humour	172
5.4.1.23 Inter-textual humour	173
5.4.2 Overview of humour types in sample.....	175

5.5 Prosodic analysis.....	177
5.5.1 Overview of results.....	186
5.6 Gestures.....	187
5.6.1 Head movements.....	188
5.6.2 Face gestures.....	192
5.6.3 Overview of results.....	196
5.7 Construal mechanisms.....	198
5.7.1 Overview of results.....	200
5.8 Conclusion.....	202
Chapter 6.....	204
A multimodal analysis of humour.....	204
6.1 Introduction.....	204
6.2 Facial expressions.....	207
6.2.1 Raised eyebrows.....	208
6.2.1.2 The meaning of raised eyebrows.....	208
6.2.1.2 Eyebrows and prosody.....	210
6.2.1.3 Raised eyebrows in sample.....	211
6.2.2 Smile and laughter.....	220
6.3 Head movements.....	230
6.3.1 Head tilts.....	232
6.3.2 Head nods.....	238
6.3.3 Head shakes and head turns.....	245

6.4 Multimodal density: a prospective test.....	251
6.5 Conclusion.....	256
Chapter 7.....	259
A cognitive analysis of humour.....	259
7.1 Introduction.....	259
7.2 Construal mechanisms identified in sample.....	260
7.2.1 Construal mechanisms and humour.....	261
7.3 Cognitive analysis of humorous instances.....	264
7.3.1 Current discourse space.....	264
7.3.2 Construal mechanisms and humour types.....	269
7.3.2.1 Anecdote and viewpoint.....	270
7.3.2.2 Irony and profiling.....	275
7.3.2.3 Parody and viewpoint.....	279
7.3.3 Mental spaces.....	282
7.3.4 The space structuring model.....	291
7.4 Convergence with other approaches to humour.....	295
7.4.1 General Theory of Verbal Humour.....	295
7.4.2 Conceptual subversion.....	299
7.4.3 Layering model and pretence space.....	302
7.5 What makes something humorous?.....	303
7.6 Conclusion.....	311
Chapter 8.....	314

Conclusion	314
8.1 Introduction	314
8.2 Summary of the dissertation	314
8.3 Main findings.....	318
8.3.1 Answering the research questions.....	318
8.3.1.1 Are there multimodal cues that contribute to signal the humorous nature of an utterance?	318
8.3.1.2 Can they be considered markers, i.e., are they consistently associated with humour and do they help to predict it?.....	320
8.3.1.3 What are the cognitive mechanisms involved in the production of humour?	320
8.3.1.4 Where does the humorous nature of an instance of communication lie?	321
8.3.2 Contribution to knowledge	323
8.4 Suggestions for further research.....	324
8.5 Conclusion	326
References	328
Annex 1: Prosodic analysis	373

List of tables

Table 1: List of construal mechanisms.....	105
Table 2: Interviews in the sample.....	133
Table 3: Utterances studied.....	158
Table 4: Humour types found in the sample.....	176
Table 5: Prosodic data per type of utterance in each interview	183
Table 6: Gestures found in the sample	197
Table 7: Type and number of occurrences of construal mechanisms found in the sample	201
Table 8: Head and face gestures per type of humour.....	206
Table 9: Face and head gestures per type of construal.....	207
Table 10: Functions of face gestures and head movements identified in the sample.....	250
Table 11: Construal mechanisms underlying the most frequent types of humour.....	269
Table 12: Mother's viewpoint.....	285
Table 13: Cristela Alonzo's viewpoint.....	286
Table 14: Stephen Colbert's conceptual integration network.....	287
Table 15: Susan Sarandon's conceptual integration network.....	288
Table 16: Stephen Colbert's conceptual integration network.....	289
Table 17: Condola Rashad's conceptual integration network.....	289
Table 18: Conceptual integration network leading to subversion of category.....	301

List of figures

Figure 1: Daniel Kaluuya mocking Stephen Colbert	27
Figure 2: Representation of a usage event	121
Figure 3: Overview of Alec Baldwin interview.....	134
Figure 4: Overview of Alison Janney interview.....	135
Figure 5: Overview of Amy Schumer interview.....	136
Figure 6: Overview of Condola Rashad interview.....	137
Figure 7: Overview of Cristela Alonzo interview.....	138
Figure 8: Overview of Daniel Kaluuya interview.....	139
Figure 9: Overview of Elon Musk interview.....	140
Figure 10: Overview of VP Joseph Biden interview.....	141
Figure 11: Overview of John McWhorter interview.....	142
Figure 12: Overview of Michael Hayden interview.....	143
Figure 13: Overview of Riz Ahmed interview.....	144
Figure 14: Overview of Sheryl Crow interview.....	145
Figure 15: Overview of Sigourney Weaver interview.....	146
Figure 16: Overview of Susan Sarandon interview.....	147
Figure 17: Gestures performed by men.....	148
Figure 18: Types of humour (men).....	148
Figure 19: Gestures performed by women.....	149
Figure 20: Types of humour (women).....	149
Figure 21: Control window in ELAN.....	151
Figure 22: Annotation tiers in ELAN.....	152
Figure 23: View of annotations by list of construal mechanisms.....	153
Figure 24: View of annotations by humour type.....	153
Figure 25: FASTsearch function.....	154
Figure 26: Annotation statistics function.....	155
Figure 27: Full soundtrack of an interview in Praat and excerpt selected	156
Figure 28: Sound excerpt open in annotation window.....	157
Figure 29: Sheryl Crow kicking.....	160
Figure 30: Daniel Kaluuya mimicking Stephen Colbert.....	167
Figure 31: Humour types found in sample.....	177
Figure 32: Total length of each type of utterance.....	183
Figure 33: Head movements.....	188
Figure 34: Tilt and turn.....	189
Figure 35: Tilt.....	189
Figure 36: Turn.....	190
Figure 37: Upstroke.....	190
Figure 38: Nod.....	191
Figure 39: Protrusion.....	191
Figure 40: Retraction.....	191
Figure 41: Smile.....	193
Figure 42: Frowning.....	194
Figure 43: Raised eyebrows.....	194
Figure 44: Blank face.....	195

Figure 45: Gestures found in sample.....	198
Figure 46: Classification of construal mechanisms.....	199
Figure 47: Distribution of construal categories found in sample.....	200
Figure 48: Construal mechanisms found in sample.....	202
Figure 49: Action Units for the brow/forehead.....	209
Figure 50: Raised eyebrows in example 54.....	213
Figure 51: Raised eyebrows in example 55.....	215
Figure 52: Raised eyebrows in example 56.....	217
Figure 53: Raised eyebrows in example 57.....	218
Figure 54: Raised eyebrows in example 58.....	219
Figure 55: Smile in example 59.....	223
Figure 56: Smile and frowning in example 60.....	225
Figure 57: Smile in example 61.....	227
Figure 58: Smile in example 62.....	228
Figure 59: Smile in example 63.....	229
Figure 60: Tilt in example 64.....	233
Figure 61: Tilt in example 65.....	235
Figure 62: Tilt and shake in example 66.....	236
Figure 63: Head tilts in example 67.....	237
Figure 64: Nod in example 68.....	240
Figure 65: Nod in example 69.....	241
Figure 66: Nods in example 70.....	242
Figure 67: Nods in example 71.....	244
Figure 68: Head shake in example 72.....	246
Figure 69: Turns in example 73.....	247
Figure 70: Shake and turn in example 74.....	249
Figure 71: Current discourse space.....	265
Figure 72: Viewing frame.....	267

List of acronyms

BF	Bona fide
CDS	Current discourse space
CP	Cooperative principle
dB	Decibels
F0	Fundamental frequency
GTVH	General Theory of Verbal Humour
H	Hearer
Hz	Hertz
KR	Knowledge resources
LA	Language
LM	Logical mechanism
NBF	Non bona fide
NS	Narrative strategy
RT	Relevance Theory
S	Speaker
SD	Standard deviation
SI	Situation
SO	Script opposition
SSTH	Semantic Script Theory of Humour
TA	Target
UE	Usage event

Chapter 1

Introduction: Humour in face-to-face interaction. A multimodal and cognitive study

1.1 Scope of the study and rationale

Humour is an inherent human quality, a form of play, and an aesthetic experience, insofar as it is an end in itself, carried out for the mere sake of enjoyment without the need for further justification (Morreal, 1983). It is a cognitive tool, enabling us to see something from a different perspective and to relish things which do not conform to our expectations (Forabosco, 2008; Dynel, 2018). Humour is also a social instrument, which serves to reinforce ties with other human beings and to acknowledge like-minded people (Hay, 2001; Dynel, 2009). Finally, it can be a coping strategy, as it helps us distance ourselves from a particular event, in order to approach and process it differently (Morreal, 1983). Hence, humour is part of what makes us human, a quality that sets us apart from other animals, along with language and rational thinking. That, in itself, makes it a fascinating object of study.

From a linguistic point of view, humour is arguably one of the most complex instances of communication, both in terms of production and comprehension (Veale, Brône, & Feyaerts, 2015). Furthermore, although most people would be able to recognise humour, its very nature has proved elusive to be defined and grasped scientifically (Attardo, 1994). Any communicative event is grounded in discourse (Langacker, 2001), and can be approached from two points of view, namely, how it is conceptualised and how it is expressed. With that in mind, the

purpose of my study is to gain an insight into how humour is multimodally expressed and cognitively grounded, drawing both on the multimodal paradigm, in order to account for how humour is conveyed in face to face interaction, and on cognitive linguistics to explain how humour is construed.

1.2 Humour studies

Traditional humour theories fall under three main categories: superiority theories, release theories and incongruity theory, the latest being the most widespread paradigm applied in verbal humour studies (Attardo, 1994). Different humour theories can be found in the literature, from Raskin's Semantic Script Theory of Humour (1985), or the General Theory of Verbal Humour (Attardo & Raskin, 1991; Attardo, 2001), adopting a semantic-pragmatic approach, to Yus's Relevance Theory account of humour (2016), along with those falling under a more cognitive perspective (Giora, 1991; Veale, Feyaerts, & Brône, 2006; Coulson, 2015; Giora, Givoni, & Fein, 2015, etc.). I believe these approaches differ in perspective, but they share core principles which can be brought together and accounted for resorting to widely studied notions in cognitive linguistics, such as Fauconnier and Turner's conceptual blending (2002) or Langacker's current discourse space (2001).

The notion of incongruity is largely behind almost every account of humour in linguistic theories (Raskin, 1985; Giora, 1997; Attardo, 2001; Veale et al., 2006; Yus, 2016, etc.). Incongruity arises when we are confronted to something that breaks our expectations and does not fit in our usual worldview. Incongruity is a necessary but not a sufficient condition to humour, as for incongruities to be

enjoyed, a certain humour-prone mood, devoid of immediate pressing needs is needed (Morreal, 1983).

1.3 Multimodality

Communication is inherently multimodal. A multimodal analysis of discourse and human communication brings in information conveyed through different modalities, such as gesture, gaze, prosody, posture, etc. (Kress & van Leeuwen, 2001; Norris, 2004a; Forceville, 2014; Adami & Kress, 2014; Adami, 2016). Modality refers to the each of the various semiotic resources with meaning potential, i.e., affordances (Adami, 2016). A multimodal perspective assumes that all these modalities, and crucially, how they interplay, contribute to communication (Kress & van Leeuwen, 2001; Norris, 2004a). Therefore, a comprehensive account of communication can only be achieved through careful exploration of how these modalities operate and relate to each other. That is the paradigm shared by multimodal studies, which do not constitute a theoretical framework as such, but rather take a multidisciplinary approach drawing on different theoretical sources. Rooted in social semiotics (Kress & van Leeuwen, 2001; Adami, 2016), multimodality has gradually adopted a more cognitive perspective (Forceville, 2016), especially when it comes to the study of co-speech gestures or prosody. Hence the relevance of combining both a multimodal and cognitive perspective in this thesis. Messages can be encoded verbally or non-verbally, but speakers can communicate more effectively taking advantage of their full multimodal expressive potential to reinforce, complement, underline or even contradict information conveyed through different modes.

With regards to multimodality in face-to-face interaction, both gestures and prosody have been found to have discursive value (Wennerstrom, 2011; Müller, 2013a). In addition, gestures can be referential (Kendon, 2004), and prosody, in turn, carries affective meaning (Cruttenden, 1986). Recent research has explored the extent to which recurrent form-function pairings in co-speech gesture occur in face-to-face communication, and to what degree these are conventionalised. Less research has been devoted to the interplay between prosody, gestures and spoken language (Muller, 2013b).

1.3.1 Multimodal markers of humour

Research has also taken an interest in studying how —or if— humour is multimodally marked (Pickering et al., 2009; Attardo, Pickering, & Baker, 2011; Urios-Aparisi & Wagner, 2011; Attardo, Pickering, Lomotey, & Menjo, 2013; etc.). The aim of these studies was to determine whether there are certain gestures, face expressions, head movements, changes in gaze, intonation or prosody patterns invariably associated with humour. In addition, research has focused on establishing whether multimodal cues co-occurring with humour can be considered as markers. That is, if they help to predict humour, or if they are simply used as metamessages as in other forms of non-humorous communication (Attardo, Pickering, & Baker, 2011).

Many studies have been conducted on the markers of irony or sarcasm, with conflicting results (Rockwell, 2000; Attardo, Eisterhold, Hay, & Poggi, 2003; Bryant, 2010; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011; Tabacaru, 2014, etc.). Less attention, though, has been devoted

to non-ironical humour. Furthermore, most of the literature is limited to staged humour, with just a handful of studies focusing on spontaneous humour (Archakis & Tsakona, 2005; Attardo, Pickering, & Baker, 2011; Feyaerts, 2013, etc.).

1.4 Cognitive linguistics

Cognitive linguistics considers language to be a window to the mind, providing clues on processes whereby meaning is constructed. Meaning making in cognitive linguistics is based on usage and experience (Croft & Cruse, 2004). Meaning is conceptualisation, rooted in discourse, and conceptualisation is embodied (Johnson, 1987; Croft & Cruse, 2004; Geraeerts, 2008). Language structure is also to a great extent determined by our experience of the world (Langacker, 2001; 2008). This cognitive approach to language blurs the boundaries between traditional dichotomies in linguistics such as langue/parole (Saussure, 1993), performance/competence (Chomsky, 1965) or semantics/pragmatics, thus allowing for a more comprehensive account of communication.

The world and our experience is conceptualised by means of construal mechanisms (Croft & Cruse, 2004). Examples of construal mechanisms are viewpoint, metaphor, metonymy, etc. Construal mechanisms encompass elements in conceptualisation that do not stem from the object or event being construed as such, but from the conceptualiser's experience, e.g., viewpoint, perspective, subjectivity, attention, etc. Consequently, they can be valuable analytic tools to determine how a given humorous utterance is construed in order to render it humorous. Conceptual blending (Fauconnier & Turner, 2002) is a further cognitive linguistic theory that has been recruited to account for humour

(Coulson, 2005a, 2005b, 2005c; Dynel, 2011). Blending refers to the process of merging various input mental spaces, which are partial representations of an event, in order to create new meaning. Therefore, it can yield insights into how humour is produced.

1.5 Research questions

In my view, a survey of the literature clearly points to a gap in current research. Further studies are needed to explore spontaneous humorous communication, as opposed to staged humour, from a multimodal and cognitive perspective, in order to gain an insight into how humour is multimodally signalled and cognitively motivated. The purpose of this dissertation is, mainly, to take one step towards filling that gap. I believe the focus on spontaneous communication is relevant, as humour is based on familiarity (Flamson, Bryant, & Barret, 2011). Due to the fact that posed humour needs to reach a wide audience, it may be delivered in an exaggerated manner, which can yield different multimodal resources than those employed in naturally-occurring, non-scripted humorous utterances (Rockwell, 2000; Urios-Aparisi & Wagner, 2011).

In light of the above, the aim of my research is twofold. On the one hand, I look into how prosody, gestures, and speech interplay in the production of spontaneous humorous utterances in English, focusing specifically on co-speech spontaneous and non-representational gestures and intonation patterns. On the other hand, I analyse humorous instances in an attempt to unearth the main underlying construal mechanisms at play in their production. I will try to answer the following questions: (i) Are there multimodal cues that contribute to signal the humorous nature of an utterance? If so, (ii) can they be considered markers, i.e.,

are they consistently associated with humour and do they help to predict it? (iii) What are the cognitive mechanisms involved in the production of humour? And finally, (iv) where does the humorous nature of an instance of communication lie? In addition, I provide a cognitive account of humour relying on the notions of current discourse space (Langacker, 2001) and conceptual blending (Fauconnier & Turner, 2002; Coulson, 2005a, 2005b, 2005c). Furthermore, I bring together current cognitive theories of humour, highlighting their similarities and trying to fit them within a simple, and broader account based on the two notions referred to above.

1.6 Structure of the dissertation

Chapter 2 offers a review of humour studies, with special emphasis on linguistic and multimodal humour studies. First, a broad definition of humour is provided (Raskin, 1985; Attardo, 1994), before giving a brief account of traditional humour theories as background for the later development of multimodal and linguistic approaches to humour. The chapter includes a detailed review of multimodal studies of humour, listing findings on the prosody of irony, sarcasm, and humour (Haiman, 1998; Rockwell, 2000; Attardo et al., 2003; Cheang & Pell, 2009; Bryant, 2010, etc.), along with other multimodal cues in humorous communication (Attardo et al., 2003; Attardo, Pickering, & Baker, 2011; Tabacaru, 2014; Tabacaru & Lemmens, 2014; etc.).

The chapter then delves into linguistic humour studies. From a semantic-pragmatic approach, both the Semantic Script Theory of Verbal Humour (SSTH) (Raskin, 1985) and the General Theory of Verbal Humour (GTVH) (Attardo & Raskin, 1991) are reviewed, as two highly influential theories, focused on humour

competence (as opposed to performance). Then, a full account of an approach based on Relevant Theory (Curcó, 1997; Yus, 2016) is included. Finally, the survey encompasses humour theories with a more cognitive stance (Giora, 1991, 1997; Clark, 1996; Brône & Feyaerts, 2003; Kyratzis, 2003; Coulson, 2005a; Coulson & Oakley, 2005; Tabacaru, 2014; Giora, Givoni, & Fein, 2015, etc.), before concluding with a review of cognitive analyses of humour in interaction.

Chapter 3 revolves around multimodal studies. It lays out the major assumptions shared by this wide-ranging analysis approach (Kress & van Leeuwen, 2001; Norris, 2004a; Adami, 2016). It reviews how this paradigm is shifting from a purely semiotic perspective into a more cognitive view (Forceville, 2009). In addition, the chapter explores the link between multimodality and other disciplines, such as cognitive linguistics, providing the grounds and rationale for combining both paradigms in a comprehensive study of humorous communication. Finally, the chapter offers a survey of the literature on gestures and prosody, as the two main modalities co-occurring with speech in face-to-face interaction. The different parts of prosody, as well as its value and functions are reviewed (Brazil, 1984, 1997; Bolinger, 1986; Cruttenden, 1986; Wennerstrom, 2001). After examining what is known to date about the relation between prosody and gestures (McClave, 1998; McNeill, 2000; Kendon, 2004; etc.), the chapter moves onto exploring the interplay between gestures and speech (Kendon, 2002; 2004; McNeill, 2005; 2008; 2013; Muller, Ladewig, & Bressemer, 2013; Muller, Cienki, Ladewig, McNeill, & Teßendorf, 2013; Ladewig, 2014a, 2014b; Muller, Ladewig, Cienki, Fricke, Bressen, & McNeill, 2014). It also includes a brief overview of major classifications of gestures, widely acknowledged in the

literature (Efron, 1941; Ekman, 1979; Kendon, 1980; McNeil 1992). Additionally, an overview of the uses and functions of gestures is provided.

Chapter 4 addresses cognitive linguistics. Its basic tenets and notions are briefly reviewed to lay the ground for the subsequent analysis of humorous utterances. It endorses the claim that people with different bodies have different experiences (Casasanto, 2013), which means that they may think differently, as meaning and conceptualisation are rooted in our experience of the world (Johnson, 1987; Croft & Cruse, 2004; Geraeerts, 2008). The chapter offers an overview of Langacker's current discourse space (Langacker, 2001), as the framework where a multimodal and a cognitive analysis of communication can converge. Furthermore, it goes through the list of construal mechanisms employed as a core element of the cognitive study conducted for this dissertation, providing definitions illustrated by examples from the analysed sample when possible.

The results of the empirical study are presented in Chapter 5. I have conducted a qualitative and quantitative analysis of a sample of 14 interviews extracted from the *Late Show with Stephen Colbert* (videos freely available on YouTube). The videos feature 7 women and 7 men at different ages and from different ethnic and professional backgrounds. 109 humorous instances have been identified in the sample, using laughter in the audience as the criterion to judge on their humorous nature. They have been transcribed, imported, and annotated in ELAN on five tiers: transcription, humour type, construal mechanism, gestures (head and face movements), and prosodic values. Mean pitch and mean intensity values have been obtained in Praat for every humorous

instance identified. In addition, a set of non-humorous utterances have also been selected to seek whether there is prosodic contrast between humorous and non-humorous utterances. A control study comparing only non-humorous utterances has also been conducted.

Chapter 6 focuses on an analysis of face expressions and head movements in humorous utterances, on the basis of the gestures with the highest number of occurrences in the sample. I explore the relationship between these multimodal cues and the construal mechanism and humour type identified in the instance in which they occur. The chapter ends with a critical review of the study conducted, clearly stating challenges encountered, as well as setting the results against the backdrop of previous studies in the literature.

Chapter 7 offers a cognitive analysis of humour. I bring together current theories of humour (Attardo, 2001; Coulson, 2015; Veale, 2015; Giora, Givoni, Heruti, & Fein, 2017; Gibbs, Sammermit, & Karzimak, 2018) under the umbrella of cognitive linguistics. In addition, I provide a model to account for humorous communication based on Fauconnier and Turner's (2002) conceptual blending, and Langacker's (2001) current discourse space. On the basis of the quantitative results presented in Chapter 5, in Chapter 7 I provide detailed analyses of several examples of humorous utterances extracted from the sample, accounting for their main construal mechanism and the humour type employed in each case. This chapter concludes with an attempt to disentangle where the humorous nature of an utterance lies.

Finally, Chapter 8 includes a summary of the conclusions reached in each chapter, along with the major findings of my study with regards to the research questions included in this introduction. Suggestions of possible future research lines are made to put an end to the chapter.

Chapter 2

Review of humour studies

2.1 Introduction

Humour is an inherently human quality (Morreal, 1983; Moran, Wig, Adams, Janata, & Kelley, 2004) and an extremely complex phenomenon (Veale et al., 2015). Despite being an intuitive notion, as most people could recognise humour even when they do not find it funny, it has proved difficult to be scientifically defined and explained (Raskin, 1985; Attardo, 1994).

Humour studies go back to Plato and Aristotle, and have been approached from various disciplines: education, philosophy, literature, anthropology, psychology, linguistics, etc. (Attardo, 1994). In this chapter, I provide a brief overview of major theories and inputs in the field of linguistics. This overview is focused on those studies and authors that have had a major influence on current research on verbal humour (for a comprehensive review of verbal humour studies see Attardo, 1994, Morreal, 2009a, and Tabacaru, 2014).

I briefly touch upon traditional general humour studies before moving on to multimodal and linguistic approaches to verbal humour. First, an outline of multimodal approaches to humour is included, along with the use prosody and gestures as possible markers of humour. As for linguistic theories, the semantic-pragmatic approaches are reviewed first, followed by the application of Relevance Theory (RT) to verbal humour, before looking into humour studies from a cognitive linguistics perspective. The chapter ends looking into humour in interaction from a cognitive perspective.

2.2 The definition of humour

Morreal (1983) considered humour to be an aesthetic experience, pursued or enjoyed for the sake of the experience itself. Raskin (1985) broadly defined humour as anything that triggers laughter or is considered amusing or funny. Attardo (1994) pointed to the difficulty of pinpointing the notion of humour. He reviewed various attempts throughout history to conclude that a more detailed definition or categorization is impossible to provide. He accepted Raskin's above-mentioned broad definition as the best starting point to undertake linguistic studies on verbal humour.

2.3 Traditional humour studies

Plato is widely considered to be the first theorist of humour. He considered humour to be a mixture of pleasure and pain (Attardo, 1994). Plato would therefore be the first to comprehend humour as arising from two opposing or contrasting feelings, thus setting a long tradition which continues to our days whereby humour is based on ambivalence or incongruity (Morreal, 2009a).

Aristotle also delved into this link between humour and incongruity. He considered humour to be a rhetorical tool to serve a speaker's power of argumentation, and drew a line between inappropriate jokes and those mechanisms suitable for the purposes of the speaker, such as irony (Attardo, 1994).

Cicero also proposed his own theory of humour, largely drawing from Aristotle's. He was the first one to establish a difference between referential humour, i.e. humour about what the text is referring to, and verbal humour, i.e.

humour about what is being said (Tabacaru, 2014). Two other elements of Cicero's approach to humour stand out from the viewpoint of linguistics: a taxonomy of humour based on the above-mentioned distinction, and the use of translation as a test to determine whether a given instantiation of humour can be classified as verbal or as referential. According to Cicero, if a humorous text can be translated, then the humour lies in the semantic value —referential humour— (Attardo, 1994). If, on the contrary, the text does not withstand translation without losing its humorous effect, it can be argued that humour relies on the form of the sign —verbal humour— (Morreal, 2009a; Tabacaru, 2014).

As opposed to the classical theories mentioned above, which attempted at providing a global account of humour, modern theories of humour are invariably created within a certain discipline (Attardo, 1994). It is widely accepted in the literature that theories of humour fall into three major categories: (a) incongruity theories (cognitive perspective); (b) superiority theories (social perspective); (c) tension release theories (psychological perspective).

2.3.1 Incongruity theories

Kant (Morreal, 2009a) and Schopenhauer (Raskin, 1985) are credited as the first authors to lay down an incongruity theory of humour (Attardo, 1994), although this notion goes back to Greek times, as explained before. They related humour to a sudden breach of expectations, a momentary deception of the audience, and a contrast between two opposing ideas (Morreal, 1983; Attardo, 1994). Incongruity-based theories are the current most widespread paradigm in

linguistic theories of humour (Raskin, 1985; Giora, 1997; Attardo, 2001; Veale, et al., 2006; Yus, 2016).

2.3.2 Superiority theories

This type of theories is also referred to as aggression or hostility theories. This notion goes back to the earliest theories of humour (Plato, Aristotle), and was forcefully advocated by Hobbes, who argued that laughter and humour implied a sense of superiority of the creator of humour with regards to the object of humour (Attardo, 1994; Morreal, 2009a). As shown in example (1) below, analysed in Veale et al. (2006, p.315), aggressive humour is therefore exclusive, as opposed to humour which is used to reinforce bonding between those individuals participating in it. Superiority theories place the focus on the social function of humour.

(1) G. B. Shaw (S): Here is an invitation to the opening night of my new play. Bring a friend, if you have one.

Churchill (H): I'm afraid I can't make it on the opening night. But I may attend on the second night, if there is one.

As noted by Veale et al. (ibid.), humour in (1) was used both by G.B. Shaw and Churchill to attack each other.

2.3.3 Tension release theories

Release theories rely on the psychological aspect of humour. Tension release theories establish that humour is used to release tension, energy or to free oneself from inhibitions, norms and conventions (Morreal, 1983, 2009a).

Freud was the most influential proponent of this theory (Attardo, 1994). Release theories are interesting from a linguistic point of view, as the idea of liberation applies also to the rules of language, i.e. creative use of language, puns, word-play, etc. (Morreal, 1983; Attardo, 1994).

Morreal (1983), focusing on laughter —not necessarily humorous laughter—, blended all three approaches and concluded that for laughter to occur a sudden shift in psychological state is needed, which can be cognitive —related to the detection and resolution of incongruities, therefore humorous in nature—, affective —to release energy or as an expression of superiority— or both, in the case of hostile humour.

2.4 Humour and multimodality

Research has taken a recent interest in studying how and if humour is multimodally marked in order to determine if there are certain gestures, face expressions, changes in gaze, intonation or prosody patterns which are invariably associated with humour, and, if so, whether they are markers, i.e. they predict humour, or they are simply co-occurring with humour (Attardo et al., 2003; Pickering et al., 2009; Bryant, 2010; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011; Attardo et al., 2013; Tabacaru & Lemmens, 2014, etc.).

Attardo, Wagner, and Urios-Aparisi (2011) drew a line between markers, i.e., intentional signs used by the speaker to communicate metamessages and mark the humorous, ironical, etc. nature of the utterance, and indices, understood as involuntary indicators of humour. They acknowledged, though, that some

behavioural cues such as laughter can be either a marker or an index. Nevertheless, it is extremely difficult, if not impossible, to differentiate between what is intentional or involuntary when it comes to gestures or prosody (Ekman, 1979). Although at times it is straightforward to point at a gesture as intentional, such as with gestures used in mimicking or emblems, which are highly conventionalised gestures; e.g. the thumbs up gesture (Kendon, 2004), most instances would be far more ambiguous (Ekman, 1979).

If we admit that raised eyebrows, for example, may be a recurrent gesture paired with spontaneous humour —although not exclusive to humour, as it can also signal surprise, for example (Ekman, 1979)—, it is difficult to determine whether this gesture is intentional or not. We may raise eyebrows because we mean to signal the humorous nature of our utterance, but it can also just be a deeply embedded gesture coming out naturally (either as a result of a multimodal metaphor or a highly conventionalised cultural gesture). As for prosodic cues, we may intentionally decide to raise our voice to warn someone of danger, but I hold questions about the extent to which we can intentionally monitor and produce every element in our prosodic delivery to showcase a certain aspect of our speech.

Many studies have been conducted on the markers of irony, with conflicting results (Attardo et al., 2003; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011). Less attention has been devoted to non-ironical humour and most of the literature is limited to staged humour, with just a handful of studies focusing on spontaneous humour (Attardo, Pickering & Baker, 2011; Flamson, Bryant, & Barret 2011).

Attardo, Wagner, and Urios-Aparisi (2011) linked multimodality in humour to underlying cognitive processes, such as metaphor or blending. In fact, under cognitive linguistics, any instance of communication is related to those cognitive processes and our conceptualisation of the world as the bases to produce linguistic units with meaning. Multimodal communication should be no different. It just brings in other channels of communication, i.e. modes or modalities, more explicitly, apart from the segmental channel (see Langacker, 2001).

2.4.1 Prosody, gestures, and humour

2.4.1.1 Prosody and humour

To date, no consistent prosodic markers associated with humour have been found (Attardo et al., 2003; Pickering et al., 2009; Attardo, Pickering, & Baker, 2011; Bertrand & Priego-Valverde, 2011; Urios-Aparisi & Wagner, 2011; Attardo et al., 2013). Pickering et al. (2009) found that punch lines in humour narratives were delivered at a lower pitch, but concluded that it was due to the fact that they occurred at the end of a paratone, thus following usual narrative intonation patterns. However, a similar study conducted on jab lines (Attardo, Pickering, & Baker, 2011) defined as one-liners occurring anywhere in the conversation and not as the final part of a narrative structure, concluded that they were not prosodically marked. Furthermore, they found no statistically significant differences between narrative, ironical, and conversational humour in terms of pitch or volume. Punch lines, however, seem to be delivered at a lower speech rate, while ironical humour is delivered at a slightly higher pitch. In the same study, Attardo, Pickering, and Baker (2011) pointed to smiles and laughter as

possible indices of humour, and as differential elements between punch and jab lines. They found that whereas punch lines were delivered with a deadpan face and acknowledged with a smile or laughter (Hay, 2001), jab lines were always delivered with a smile or laughter.

On the contrary, Archakis, Giakomelou, Papazachariou, and Tsakona (2010), who looked into the prosodic cues used to mark humorous turns, concluded that jab lines in humorous narratives are marked by pauses, slower pace and higher intensity. However, methodological issues arise from their study (Attardo et al., 2013; Gironzetti, 2017), as they used pause-based intonational units to segment speech. In addition, they set the lower limit for pauses at 0.3 seconds, which is defined as very short pauses (Brown et al., 1980), thus not significant enough to mark a conspicuous silence.

Bird (2011) found certain prosodic features associated with riddles that tell them apart from non-humorous conversational Wh-questions. Prosody can also be used for humorous effects (Wennerstrom, 2001, 2011; Wennerstrom & Siegel, 2003). For example, using changes in key and paratones, or using lower pitch to play down the importance of certain elements to later maximise the humorous effect. The intonation of contrast and the intonation of given information contribute to the humorous effect of jokes, as cohesive devices of discourse which enable the listener to track down cohesive links that may not be obvious in the purely linguistic form of the utterance (Wennerstrom, 2011). Punch lines are funnier when there is a mismatch between linguistic and intonational cohesion signals, which contributes to heightening the incongruity underlying humour.

2.4.1.2 Prosody and irony

A survey of the literature, then, shows that most studies have found no consistent markers of humorous speech, as compared to serious discourse. This is a counterintuitive notion, especially given the abundance of studies claiming that irony, for instance, is associated with certain intonation patterns (Rockwell, 2000; Attardo et al., 2003; Cheang & Pell, 2009; González-Fuente, Escandell-Vidal, & Prieto, 2015; etc.). These studies have yielded a wide range of often conflicting results, whereby irony is associated with flat (Haiman, 1998), rising intonation (Schaffer, 1982), higher (Rockwell, 2000) and lower pitch (Haiman, 1998; Anolli, Ciceri, & Infantino, 2000), heavy exaggerated pitch (Adachi, 1996) and relatively monotonous intonation (Haiman, 1998), etc. Given this varied array of results, the question arises as to what extent we can consider any of those intonation patterns to be a marker of irony. In fact, Attardo et al. (2003) claimed that there is no such thing as an ironic intonation, but rather that pitch and changes in prosody are just contrastive markers.

Cheang and Pell (2009) conducted a study on prosodic markers of sarcasm in Cantonese and English. They found differences in speech rate, voice quality, frequency, and resonance between sincere and sarcastic utterances. However, prosodic patterns were not the same in both languages, which led the authors to claim that prosody is key to show non-literal intention in speech, but conventions vary among languages. They also argued that certain sarcastic expressions were so entrenched in the language/culture —what Haiman (1998) refers to as enantiosemaic expressions— that they became markers of sarcasm themselves without the need of any underlying prosodic cue.

González-Fuente et al. (2015) found that some ironic utterances are punctuated by audiovisual behaviour that they call “gestural codas”, which greatly facilitate the comprehension of the utterance by the hearer as ironic. They considered irony to be a phenomenon beyond the traditional view based on the opposition between what is said and what is meant, as there are instances in which the speaker means what they say and still wants to be ironic. Therefore, they endorsed a more comprehensive approach to irony as a phenomenon in which discrepancies between expectations and reality are showcased. Their study bore out former claims made by Attardo et al. (2013) pointing to the lack of markers of irony, which can be signalled by speakers relying on different verbal and non-verbal clues, not just specific to irony. González-Fuente et al. (2015) frame the use of these cues, and in particular of gestural codas, within a Relevance Theory account of humour (Yus, 2003, 2016), as tools to reduce the cognitive effort required from the hearer to interpret the ironic nature of the utterance. Prosody and gesture, according to them, are just pragmatic facilitators. Along this line, Tabacaru (2014) found a correlation of prosodic patterns with humorous, mostly sarcastic, utterances, whereby words allowing for a number of implicatures are stressed.

Haiman (1998) explained that sarcasm is a form of metalanguage, as it conveys the speaker’s feelings about the utterance rather than its propositional content. A series of what he called indices are used to signal that the utterance has to be interpreted as sarcastic. He listed the following mechanisms as possible cues to underline the sarcastic metamessage of a given utterance (1998, pp. 30 – 33): sneers and laughter (linked to nasalisation); inverse pitch obstruction, i.e.

the stressed syllable uttered at a lower pitch than the surrounding material; intonational misfits, defined as inappropriate intonation, such as exaggeration or caricature, falsetto, heavy sarcasm —combination of stress and relatively monotonous intonation—, separation by heavy pauses or sing-song intonation. Haiman acknowledged, though, that these indices are not exclusive to sarcasm; they point to the insincerity of what is being said, or to the discord or distance between the speaker's real feelings and the utterance, so we could assume that they could be indices for any utterance in which the propositional content cannot be taken at face value, therefore, they could apply to non-ironical humour as well.

Rockwell (2000) found out that sarcastic utterances are delivered at a lower pitch, slower tempo and higher intensity than non-sarcastic utterances. It is interesting to note, though, that while in her study participants were able to recognise posed sarcasm just by means of vocal cues, they were unable to tell spontaneous sarcasm apart from non-sarcasm. This suggests that posed, i.e. staged, sarcasm is delivered with exaggerated features, which may be the case also for gestures. Hence the importance of working with spontaneous, natural speech despite the obvious difficulties. Finally, she saw sarcasm as an expression of mild negative emotions, arguing that for stronger negative emotions speakers may resort to stronger verbal and vocal tools, e.g. insults, starker changes in F₀, etc. She also recognised that studies linking emotion to prosodic features are often inconsistent regarding their findings.

Attardo et al. (2003) established three broad categories of prosodic patterns in ironic utterances:

- a) Strong within-statement contrast: characterised by consecutive phrases with extremely high and extremely low pitch range respectively. In the literature, extreme contrast is linked to high emotional meaning, so this pattern could be explained so as to convey the feelings of the speaker rather than the literal meaning of the utterance. Following Rockwell (2000), this could point to stronger feelings in the speaker requiring more exaggerated features.
- b) Compressed pitch pattern: little change in pitch; flat intonation. Wennerstrom (2001) associates compressed intonation to parenthetical remarks, so it could be used to show that the sarcastic utterance is not really part of the normal flow of speech, and therefore its communicative value lies elsewhere.
- c) Pronounced pitch accents: all content words heavily stressed by means of higher pitch. As high pitch is generally accepted to be used to underline certain words, it could be argued that this intonation pattern would signal that special attention is required from the listener to process the sarcastic utterance.

As hitherto explained, Haiman's (1998) hypothesis, fully in line with Attardo et al.'s (2003) findings, is borne out by empirical studies regarding irony and sarcasm, as the common thread of the various intonation patterns found in the literature point to the notion of exaggeration and fake. Nevertheless, no such patterns have empirically been associated to non-ironical and non-sarcastic

humour, which most linguistic theories base on some form of incongruity; that is, a clash between the expectations held on an utterance and its actual instantiation (Attardo, 1994; Giora, 1998; Attardo, 2001; Brône & Feyaerts, 2003; Forabosco, 2008; Veale, 2015; Dynel, 2018, etc.). For example, in Alec Baldwin's interview, upon taking his seat, right after being welcome by the host and by the audience with a very big round of applause, he thanks the audience and stresses that they are very nice people. Then he utters: "It's nice and chilly in here", which causes a bout of laughter in the audience. I posit that humour in this utterance arises from a clash in expectations about what he was supposed to say, e.g. "it's nice, what a warm welcome", etc., and the fact that he actually states that it is "chilly". Furthermore, 'chilly' is reinforced with higher pitch, a nod and a smile.

One possible explanation of the absence of markers in humour as opposed to irony may be that prosodic cues are used only as metalanguage showing affect, that is, the position and feelings of the speaker with regards to the utterance. In the case of humour, it can be argued that there is no such detachment between the speaker and the humorous text. Both sarcastic/ironical and humorous utterances are manipulated by the speaker, but in two distinct ways. Sarcastic/ironical utterances are manipulated to show what the speaker thinks about the utterance. Humorous speech, on the other hand, is manipulated to mislead the hearer to a false interpretation to be subsequently proved wrong in order to achieve the humorous effect (Tabacaru, 2014), as in example (2) below, in which General Hayden gives the impression that he is going to tell a big secret only to reverse that idea and make clear that he will not. If that is the case, we are confronted to two different phenomenon and there is therefore no reason

to expect they are conveyed resorting to the same kind of metalinguistic or multimodal mechanisms (Bryant, 2010).

(2) General Hayden: Can you keep a secret?

Stephen Colbert: Turn off the cameras...I can keep a secret.

General Hayden: Me too.

Another explanation put forward to account for the difference between ironic and non-ironic humour in terms of multimodal marking associates the lack of markers to signal humour with an in-group expression of bonding on the part of the speaker, as relying on the common ground assumed to be shared with the interlocutors, and necessary for humour to be comprehended, thus demonstrating the affinity between participants (Tabacaru, 2014).

2.4.1.3 Contrast as a marker

Bryant (2010) rejected the existence of consistent prosodic patterns for ironic utterances. First, he established that since different subtypes of irony convey different affective states, e.g. sarcasm – negative; jocularly – playful, etc., it is only to be expected that prosodic features will be different for each of those types. He linked irony utterances to prosodic contrast, regardless of what acoustic variables change. Prosodic contrast was defined as “a statistically reliable shift between adjacent phrasal units in at least 1 of 5 acoustic dimensions (mean fundamental frequency, fundamental frequency variability, mean amplitude, amplitude variability, and mean syllable duration)” (Bryant, 2010, p. 545). His data suggested that ironic utterances are delivered with greater contrast from adjacent phrases, but that contrast is varied with regards to the parameters changed and

the direction of those changes. The only consistent feature linked to ironic utterances is that they are delivered at a significantly slower rate, which may be explained as a way to facilitate processing by the listener, as more cognitive effort is needed to interpret this type of utterances (Sperber & Wilson, 2004).

Nevertheless, Bryant (2010) acknowledged that contrast is not exclusive to irony. It is a prosodic tool at the speaker's disposal to help disambiguate meaning. Prosody may be used to convey or stress linguistic meaning (focus, stress) and affective meaning (the speaker's attitude). But in any speaking context, the interplay between verbal and vocal communication is highly complex, and trade-offs may be needed, as the prosodic features necessary to mark linguistic and affective information may conflict with each other.

The notion of contrast (Attardo et al., 2003; Bryant, 2010) was also advocated by Urios-Aparisi and Wagner (2011) as the motivation underlying the use of prosody for humorous purposes. They argued that there are not prosodic markers of humour as such, but rather that prosody is used for the performance of humour, as opposed to competence, and therefore cannot be dissociated from its pragmatic value. I argue that all communication is inherently multimodal and that both the production and comprehension of utterances, whether humorous or not, inevitably hinge upon a wide range of multimodal, cognitive, and communicative tools at hand: language, prosody, gestures, background knowledge, context, and discourse (Langacker, 2001). An illustration can be found in example (3) below, where Daniel Kaluuya is teasing Stephen Colbert by mimicking a previous remark made by the host. The humorous nature of (3) can only be understood in the context of the interview, knowing what Stephen Colbert

had said first, why it had been picked up by Daniel Kaluuya to mock the host, i.e. because it showcased a certain awkwardness due to racial differences, being aware that racial issues was the main topic in the film they are discussing, starred by Daniel Kaluuya. Crucially, only by seeing and listening to Daniel Kaluuya's speech and multimodal behaviour (mimicking gestures, smile, etc.), can the humorous intent be fully apprehended.

(3) Daniel Kaluuya: It's like...What would I say...If I was white...What would I...?



Figure 1. Daniel Kaluuya mocking Stephen Colbert.

Interestingly, Flamson et al. (2011) argued that as humour comprehension is influenced by context, the more background information is shared by the participants in the interaction, the less marking would be necessary for humour to be interpreted. In other words, the larger the intended audience of the

humorous utterance, the more salient this humour will need to be made in order to ensure it is successfully conveyed (Attardo et al., 2003).

This is precisely one of the reasons why I believe that staged humour cannot always be taken as a reliable proxy of spontaneous humour. Humour needs to be marked ostentatiously to make sure the audience will not miss it. Furthermore, actors do not improvise humour, they enact rehearsed and carefully scripted situations. Therefore, the cognitive processes involved in this enactment of humour are arguably not the same as the ones involved in the production of spontaneous humour, therefore drawing on a different kind of multimodal devices.

In light of the above, there seems to be no consistent markers of humour. Instead, prosodic and gestural cues, not specific to humour utterances, are sometimes used to communicate humour more effectively. The patterns and salience of the indices involved will eventually depend on the pragmatic context in which humour is conveyed. As Bryant (2011, p.306) puts it, “prosodic production [...] can only be understood in the larger framework of language use, vocal signalling, and the evolution of communication and cognition”.

2.4.1.4 Gestures and humour

As mentioned before, Attardo et al. (2003) pointed to the blank face as a possible marker of irony or sarcasm. However, they placed special emphasis on drawing a distinction between marker and phenomenon: an ironic utterance will be ironic even in the absence of any marker: blank face or prosodic cues, etc. In

other words, the markers associated with irony in the literature do not confer the ironic character to the utterance.

Attardo et al. (2003) further claimed that there seems to be a hierarchy of cues leading to successful communication: behavioural, e.g. laughter, smile, gestures in general; intonational, and semantic. However, this hierarchy has not yet been conclusively borne out or rejected by research (Attardo et al., 2003, p. 4). They added that these cues have metacommunicative, i.e. communicate about the utterance, or paracommunicative, i.e. communicate besides the utterance, value, insofar as that they alert the hearer about a certain non-regular interpretation of what is being said (ironic, humorous). They posited that both intonational cues and gestures, such as a blank face, function as contrastive markers signalling irony.

Attardo, Wagner, and Urios-Aparisi (2011) compiled different ironical gestural cues appearing in the literature (Muecke, 1978; Attardo et al., 2003): eyes (wide open, squinting, rolling, winking); nodding; lip tightening; smiling; blank face, prosody-face incongruity; gaze aversion. It is not clear whether these are intentional or involuntary signals. Only intentional –albeit maybe unconscious as entrenched conventionalised forms of communication– signs can be considered as markers (Attardo et al., 2003). Non-intentional signals are defined as leaks (Ekman, 1979). Leaks are just giveaways showing the speaker's attitude and feelings. Furthermore, the mere co-occurrence of gestures/prosodic patterns with humorous utterances is not enough to consider them markers, especially if they also co-occur with non-humorous communication (Attardo, Pickering, & Baker, 2011). For example, in Attardo et al. (2013), laughter and smile are used

to frame humorous turns or chunks of speech, but they cannot be considered markers, as they are not consistently associated with humour nor do they integrate with the humorous part, namely, the punch line. They are just used by the interlocutors as switches from non-humorous to humorous communication.

Tabacaru and Lemmens (2014) argued that raised eyebrows are gestural triggers prompting the hearer to take the utterance as humorous, ironic, or sarcastic. Gestural triggers co-occur with humour and contribute to meaning making. They placed gestural triggers among the discourse gestures defined by Müller (2004), used to punctuate speech, favouring the pragmatic analysis over the semantic view that this gesture would be associated with an expression of human emotion, namely surprise, as far as raised eyebrows are concerned (Guaiatella et al., 2009).

2.4.1.5. Markers, indices, and indicators

A terminological clarification is needed with regards to the different terms used in the literature to refer to the multimodal cues used to possibly signal humour, irony, sarcasm, etc. Analysing multimodal markers of irony and sarcasm, Attardo (2000b) first established a difference between factors, as necessary constituents of irony the lack of which implies there is not irony, e.g. contextual inappropriateness, the presence of two distinct meanings, etc. (Attardo, 2000a; Attardo et al., 2003), and markers, which can help to signal irony and facilitate its understanding, but that are not essential to the phenomenon. In other words, irony will still occur even if those markers are removed.

Markers mainly serve to communicate the metamessage that a certain utterance or turn is intended as ironical, or humorous (Attardo, Pickering, & Baker, 2011). Multimodal cues can only be considered as markers if they are intentional, which does not necessarily mean that they are made consciously. Attardo, Wagner, and Urios-Aparisi (2011) argued that markers need to be “at least implicitly intended by the speaker to facilitate the recognition of the humorous/ironical intention” (p. 196). That is, a signal can be made unconsciously to mark humour as a result of deeply entrenched conventionalisation, but it would still be considered intentional in the weak sense just described, as it is precisely its use as a marker of humour what has been conventionalised and embedded in our behaviour.

Unintentional cues, such as spontaneous laughter, are named indicators (Attardo et al., 2011, p. 197). The authors considered that this is a very significant difference, as indicators would not fall within ostensive communication from a pragmatic point of view. A further distinction is drawn between indices, as unintentional leaks (Ekman, 1979) of humour, and indicators, the difference being that the latter would always co-occur with humour (Gironzetti, 2017).

In my view, this classification does not bode well with a multimodal perspective. First, following the definition and the few examples of factors offered by Attardo (Attardo, 2000; Attardo, Pickering, & Baker, 2011), I argue that they are semantic-pragmatic in nature, consequently not relevant for a strictly multimodal analysis of humour. In addition, no indicators have been found in the literature so far (Gironzetti, 2017), so their existence is not warranted empirically. With regards to the distinction between markers and indicators/indices, I find that

it is extremely difficult, if not impossible, to judge on the intentionality of the multimodal cues produced by speakers, leaving aside highly symbolic and language-like signs, such as emblems or sign language (cf. Chapter 3).

The line between intended but unconscious markers and unintentional indices on the basis of deeply entrenched conventionalisation is too thin. No criteria to differentiate what is the product of embedded culture from what is fully spontaneous, to the extent that is unintended, are provided. I have found none in the literature. Finally, if a certain multimodal behaviour to signal humour had become entrenched and conventionalised precisely for that particular purpose, one could expect to find such behaviour recurrently associated with humour. The fact that no multimodal cue has to date been consistently and unequivocally linked to humour leads me to at least question Attardo's premise.

As a matter of fact, I claim that the intended/unintended distinction is irrelevant for a multimodal analysis (Ekman, 1979), as unintentional signs do not need to be excluded from the study of communication. Regardless of whether multimodal signs are intended or unintended, they contribute to facilitating communication and comprehension of the message by the hearer, i.e. meaning is attributed to these signals whether they have been intentionally produced or not (Flecha-García, 2010).

2.5 Linguistic humour studies

Linguistic humour studies have been undertaken from different perspectives. What follows is an account of those theories and authors that have

had a major influence on current research, either in support or against such approaches. A critical analysis is provided as the theories are presented.

Before explaining the most influential theories more in detail, I would like to touch upon a recurrent topic in linguistic theories of humour, as it is dealt with and referred to in many of the approaches which will be subsequently explained: humour as a violation of Grice's maxims (1989).

In his paper "Logic and Conversation" (1989), Grice described certain parameters (maxims) governing conversation, understood as a rational process, falling under a general Cooperative Principle (CP). He stated that as rational beings, humans could be expected to want to actually communicate –be cooperative– and follow certain guidelines when speaking to each other, which can be summarised as follows:

- a) Maxim of quality: be truthful.
- b) Maxim of relation: be relevant, thus informative.
- c) Maxim of quantity: say no more nor less than what is required to be informative.
- d) Maxim of manner: be clear, avoiding ambiguity and obscurity.

He also claimed that speakers could flout, i.e. intentionally violate, some of those guidelines, in order to point to the need of inferring non-literal meaning, thus creating implicatures. He further classified implicatures into conventional implicatures, i.e. inherent to the utterance, detachable but not cancellable, and

conversational implicatures, i.e. cancellable but not detachable, as no other way of saying the same thing could be found.

A widespread assumption in the literature is that humour involves the violation of one or more of Grice's maxims (Raskin, 1985; Attardo, 1994; Yus, 2016). Grice established that in order to maintain the cooperative principle and achieve communication, a maxim can be 'flouted' by the speaker to lead to a given implication. But maxims are violated only when the speaker's communicative intention is not cooperative. However, humour scholars argue that humour does imply the violation, and not just mere flouting, of Grice's maxims (Raskin, 1985; Attardo, 2001). This would therefore result in the paradox that humour is not cooperative in terms of communication, so this apparent paradox in humour communication must be accounted for (Attardo, 1994).

One explanation put forward by Sperber and Wilson (1986), followed-up later by Yamaguchi (1988), is the mention theory, claiming that in the case of humorous texts, the violation of Grice's maxims is not real, but enacted (Attardo, 1994). According to this theory, the speaker would pretend to be violating the cooperative principle by just mentioning the utterance or echoing the violation committed by a character in the joke. However, this mention theory does not withstand many examples of jokes, such as the following, extracted from Attardo (1994, p. 285):

(4) Have you heard the latest?

No? Well, neither have I.

In this example it cannot be said that the speaker is just mentioning the utterance or echoing a character in the joke, as there is no assumption of a role or clear referent to be mentioned. Therefore, the violation of Grice's maxim of relevance is not enacted (Attardo, 1994).

Another attempt to resolve the paradox comes from Raskin (1985), supported by Attardo (1994). They set a difference between *bona fide* (BF) communication, to which Grice's Cooperative Principle (CP) applies, and *non bona fide* (NBF) communication, with its own CP and set of maxims, under which humour falls.

- a) Maxim of quality: say only what is compatible with the world of the joke.
- b) Maxim of quantity: give exactly as much information as is necessary for the joke.
- c) Maxim of relation: say only what is relevant to the joke.
- d) Maxim of manner: tell the joke efficiently.

NBF differs from Grice's flouting in that in NBF a maxim is not violated to implicate information, but because a different CP is followed.

As further explained in section 2.5.2, Yus (2003, 2016) rejected this claim and went around the paradox of the seemingly non-cooperative nature of humour appealing to Relevance Theory (RT). He argued that the humorous intention is mutually manifest to the interlocutors, who then cooperate to achieve relevant communication. Relevance, in this case, lies on the eventual positive interpretation of the humorous text and the amusement, laughter, etc. entailed.

He further claimed that inferential strategies must be the same for humorous and non-humorous texts alike, as humans can resort to one single ability to interpret and comprehend any given input.

Using RT terminology, I argue that communication occurs even when it is not ostensive. We cannot help but to communicate by means of gestures, facial expressions, gaze, etc. even if we do not mean to. Therefore, any account of communication including only those instances in which interlocutors clearly show they want to communicate is leaving aside many instances of communicative situations in which that desire to communicate is neither ostensive nor manifest.

In fact, this is precisely where cognitive linguistics and semiotics intersect, understanding semiotics as the science exploring meaning-making resources, intentional or not. However, whereas semiotics places emphasis on how meaning is produced and interpreted through signs, cognitive linguistics is more concerned with how meaning is created as a result of conceptualisation (Croft & Cruse, 2004; Evans & Green, 2006). Signs are just the visible results of embodied and experiential cognitive processes allowing us to apprehend the world around us. We conceptualise such world in domains leading to meaning-units, eventually shaped into multimodal signs, which can ultimately be multimodally, pragmatically, and semantically interpreted by an audience.

Grice's account of communication applies to prototypical rational conversations between mainly cooperative beings. If we admit that humour is produced and enjoyed as a liberating aesthetic experience (Morreal, 1983), we may assume that it is not entirely rational —understood as governed by the laws

of logic—. Consequently, Grice's theory may not be the best suited to account for humorous communication.

2.5.1 Semantic-pragmatic approach

2.5.1.1 Raskin's Semantic Script Theory of Humour

The Semantic Script Theory of Humour (SSTH) (Raskin, 1985) was designed to explain the speaker's humour competence in an idealised communicative (humorous) situation. Therefore, it is rooted in Chomsky's generative grammar (Attardo, 1994).

SSTH is based on the notion of 'script', that is, a set of organised prototypical contextual and lexical information in the mind of the speaker and addressee. Yus (2016) argued that scripts are merely sequences of actions, but Attardo's accounts of SSTH (1994, 2001) included information such as place, time, and related vocabulary linked to those actions. Attardo (1994, p.199) equated the notion of script to that of Fillmore's 'frames', among other terms used to refer to this kind of cognitive structures, stating that the difference is merely terminological. Scripts link to form semantic networks including lexical and non-lexical scripts as well as all the links. Semantic networks contain all the information a speaker has about their culture (Attardo, 1994).

SSTH applies solely to single jokes, not to larger texts or other types of humour. Raskin's theory predicts humour will occur if the following conditions are met: (a) the joke text is compatible with two different scripts, which fully or partially overlap; (b) both scripts are opposite.

Scripts may be combined in different ways leading to various meanings. Coherent interpretations yielded by these combinations will be stored as the meaning of the text, which is then considered to be well-formed (Attardo, 1994). Occasionally, texts may be compatible with different scripts. When the scripts overlapping are opposed to each other, humour occurs, as in example (5), taken from Raskin (1985, p.11):

- (5) “Is the doctor at home?”, the patient asked in his bronchial whisper.
“No”, the doctor’s young and pretty wife whispered in reply. “Come right in”.

In (5) two possible scripts are elicited by the joke. One, more salient (Giora, 1991) script, in which the patient goes to the doctor’s house to seek medical assistance. Another, less salient script, in which the patient is seeking to spend time alone with the doctor’s wife. In the latter, intimacy is conveyed through the whispering tone of the conversation. The patient whispers because he has chest infection; the doctor’s wife whispers as a sign of intimacy. The hearer is compelled to switch scripts by a section in the text which acts as a script-switch trigger. In (5), this trigger would be the “Come right in” invitation by the doctor’s wife.

Script-opposition may be of three different categories: actual/non-actual; normal/abnormal; possible/impossible. Raskin went one step further and broke down these three categories into 5 types of instantiations: good/bad; life/death; obscene/non-obscene; money/no-money; high/low-stature (Raskin, 1985). I agree with Attardo (1994) when he claims that this is a rather limited list of

instantiations, and that it may be expanded and changed depending on cultural differences. Furthermore, Raskin (1985) argued that jokes only seem to violate Grice's maxims, whereas they followed a different set of maxims defining non-bona fide communication (NBF), whereby speakers are not committed to telling the truth.

Attardo (2003) claimed that SSTH fully integrates semantics and pragmatics, but I argue that it is essentially a semantic theory, as the humorous nature of the text —namely, the joke— is eventually placed on the text itself and not on the context or the communicative situation. Also, as one of its main premises is to account for the speaker's humorous competence and its object of study is an idealised communicative situation, it explicitly leaves out the pragmatic value of context in the production of humour.

As I see it, the complex nature of humour makes any analysis based on a single approach incomplete. The SSTH constitutes a valid first-step analysis. Humour may lie, among other things, in the semantic value of texts (or utterances). The SSTH provides a valid framework to start to grasp how humour is created. In fact, Attardo (2001) stressed the semantic-pragmatic value of SSTH in his own revision of the theory, his General Theory of Verbal Humor (GTVH), explained in section 2.5.1.2, pointing that text-processing occurs simultaneously at syntax, semantic and pragmatic level.

2.5.1.2 Attardo's General Theory of Verbal Humour

Attardo (1994) portrayed his General Theory of Verbal Humour (GTVH) (Attardo & Raskin, 1991; Attardo, 2001) as a revision of Raskin's SSTH, in order

to account for any kind of humorous texts, including larger texts. GTVH goes beyond semantics to become a linguistic theory at large.

GGTV considers script opposition to be just one of 6 knowledge resources (KR) necessary to produce humour (Attardo, 1994, p.27):

- a) SO – Script Opposition: taken from SSTH.
- b) LM – Logical Mechanism: the logical relation between both scripts (false analogies, juxtaposition, etc.). This is the cognitive mechanism used to solve the incongruity leading to humour.
- c) TA – Target: the target of the joke (in aggressive humour; non-aggressive humour have a 0 value in this parameter).
- d) NS – narrative strategy: the structure of the text (narrative, dialogue, etc.).
- e) LA – language: the information necessary to place language elements in the right place and correctly verbalise the text.
- f) SI – situation: the situation in which the text is placed (participants, settings, etc.).

According to GTVH, humour arises from different combinations of the values assigned to KR. Attardo's theory establishes a hierarchy whereby each parameter is constrained by those above and determines those below:

SO → LM → SI → TA → NS → LA

Veale et al. (2006), however, argued that KRs cannot be analysed in isolation and that they interplay all along the humour interpretation process. For

Brône and Feyaerts (2003), GTVH is essentially a cognitive theory, as it aims at explaining humorous use of language drawing on parameters of different kinds, among which the LM is cognitive in nature. They argued that the broader cognitive semantic notion of *mapping* can take Attardo's LM even further in order to account for humour. Attardo (2001) states that the LM presupposes a playful logic which may not exist outside the joke, and that it embodies the resolution of such joke (ibid. p.25). Finally, although GTVH aims at being a comprehensive linguistic theory of verbal humour, expanding its scope beyond canned jokes to include larger texts, Attardo himself acknowledged that "its application to conversational humour is less than straightforward" (ibid. p. 68).

2.5.2 Relevance theory applied to humour studies

Yus (2016) defined Relevance Theory as a cognitive pragmatic theory of human communication, aiming at explaining the mental process undertaken by speakers in the production of utterances, as well as the inferential strategies allowing addressees to process and interpret those utterances. He argued that the mental mechanisms at stake are universal, and are therefore applicable also to humorous communication. This claim is tantamount to saying that humorous discourse is not interpreted by means of special procedures, but on the basis of a "single cognitive criterion" (Yus, 2016 p. xv).

Curcó (1997) also framed her study on humour within Relevance Theory. She disagreed with the traditional claim that humorous discourse is somehow deviant from the standard use of language. She argued that texts are not humorous *per se*, and that the humorous effect lies in the mental processes

undertaken by the hearer during the interpretation phase. Interpretation of an utterance hinges on two stages: decoding the lexico-semantic component and inferences made on the basis of ostensive communication. The intended meaning of an utterance may be made ostensive by the speaker through various verbal and paraverbal means, such as multimodal cues (gestures, intonation, gaze, etc.).

In humorous communication the speaker clearly marks some sort of detachment from what they are uttering. While acknowledging the incongruity-based nature of humour, Curcó (ibid.) argued that humour does not stem from incongruity-resolution, but from the inferential process triggered by its production and the cognitive effects on the hearer.

Relevance Theory (RT) (Sperber & Wilson, 1986) claims that human cognition is set to maximise relevance; that is, to obtain as much information as possible from inputs received in a communicative situation, verbal or otherwise, with as little cognitive effort as possible.

In humorous communication, though, it is clear that jokes or other forms of humorous utterances are not necessarily very informative, as in (6), already presented above:

(6) Have you heard the latest?

No? Well, neither have I.

Furthermore, humour comprehension does take additional cognitive effort on the part of the hearer. This additional effort, though, is compensated by the humorous effect: pleasure in incongruity resolution, laughter (tension release),

group bonding, etc. RT is based on three major theoretical assumptions: that all humans seek the maximum relevance in communication, that inputs in ostensive communication are presupposed to be optimally relevant, and that all utterances are less informative than what they are meant to communicate (Sperber & Wilson, 1986). In other words, the pragmatic value of an utterance always goes beyond its semantic value.

RT also claims that humans are endowed with a mind-reading ability (Yus, 2016) enabling speakers to make predictions about how their utterances will be interpreted. This is crucial for the production of humour, as the speaker can foresee what interpretation is more likely to be favoured by the hearer, i.e. the most relevant, to use it as a bait to mislead the hearer toward it before producing the humorous effect, associated with a less likely —a priori less relevant— interpretation.

For RT, the process of coding and decoding utterances, humorous or not, is just a first step in communication, which relies mainly on inferences; that is, the interpretation of that utterance, or, as Yus (2016, p. xvii) put it: “turning the schematic coded discourse into meaningful and relevant interpretations”.

We can also state, using standard RT terminology, that interpretation of an utterance involves for the addressee first to work out the explicature, decoding the semantic value, and, when necessary, move on to infer the implicature of that utterance, i.e. what is not explicitly said but implicated by the speaker. A further distinction is drawn between implicature —intended to be communicated by the

speaker, albeit not explicitly— and implication, understood as what can be inferred but was not intended to be communicated.

Yus (2016) pointed to two universal cognitive systems leading to two different types of inferences: inferential, to process the utterance, and social, to compare inferences made with stored cultural information. He claimed both systems are activated simultaneously and argued that the social system is culture-specific.

To me, our conceptualisation of the world —which cannot be dissociated from our experience of it, and is therefore necessarily embodied and culture-bound— affects how we communicate in general: how we produce and interpret communicative inputs, verbal or otherwise. In my view, there is only one cognitive system allowing us to process communication through the combination of semiotic, semantic, pragmatic and cognitive elements.

RT is only concerned with the pragmatic analysis of ostensive communication; that is, communication in which there are both a communicative intention on the part of the speaker and an informative intention, referring to the actual information that is meant to be conveyed. Accidental, non-ostensive communication is excluded from RT analysis (Yus, 2016). However, I argue that humorous communication may sometimes be non-ostensive, and still achieve to convey a humorous effect retrieved by the hearer. An example is an utterance whose humorous effect is only apparent to the interlocutors after it has been produced. In this case, the context renders that particular utterance humorous.

As a result, the utterance is a valid object of study for pragmatics. Let's consider (7) below:

(7) Susan Sarandon [in the middle of the interview, as she was talking about one of her films]: And then, various other scripts...Oh, my god, look at your ceiling!

In this case, there is no ostensive intent to produce humour on the part of the speaker; Susan Sarandon had been really taken aback by the ceiling when she accidentally looked up and saw it, but the effect of such statement ends up being humorous for the audience, which reacts to it laughing, as a result of the incongruity of making an out-of-place remark on the ceiling while in the middle of the interview. To me, the humorous effect in this case is grounded in the context, and not on the speaker's ostensive humorous intention.

In his Relevant Theory account of humour, Yus (2016) endorsed the incongruity-resolution theory, whereby the speaker manipulates the hearer to guide them to take a series of steps toward a given interpretation. This is possible because the speaker, as mentioned above, can predict what inferential strategies will be used by the hearer. That path is eventually reversed and the incongruity exposed in order to produce a humorous effect.

As explained above, different humour theories (Raskin, 1985; Giora, 1991; Attardo, 1994, among others) assume that the incongruity arises from the clash between two opposing interpretations of the humorous utterance or text, depending of the theory. Those interpretations, in turn, are based on a certain mental representation of the situation elicited by the humorous input, which draws

on previously stored lexical and encyclopaedic information. Furthermore, Yus (2016) categorised humorous incongruity into two types: discourse-based and frame-based. Discourse-based incongruities relate to the inferential strategies applied to the input as such in order to retrieve the explicature, whereas frame-based incongruities link to the situation mentally represented by the hearer on the basis of that input.

Yus (2016) proposed an Intersecting Circles Model to account for humorous communication, whereby all humorous input involves predictions made by the reader about possible interpretations made by the audience on the basis of three areas or circles, namely utterance interpretation, make-sense frames and cultural frames. Although this new approach has the merit of doing away with a clear cut divide between verbal and cultural jokes found in previous research, I do not see the need to establish those distinct, possibly overlapping areas or circles. In my view, at any given time there is only one mental representation elicited by the communicative input including inseparable linguistic and cultural elements. I agree with Yus, though, that the cognitive processes involved in the production and comprehension of any kind of communication are the same, whether the purposes of communication are humorous or not.

2.5.3 Humour and cognitive linguistics

Meaning making in cognitive linguistics is based on usage and experience (Evans & Green, 2006; Geraeerts, 2008). Language structure is to a great extent determined by our experience of the world. Cognitive linguists consider that language is embodied and grammar is conceptualisation (Croft & Cruse, 2004).

Language and cognition, therefore, cannot be separated, and are influenced by the interplay between social, cultural, psychological, communicative and functional elements (Brône & Feyaerts, 2003).

In cognitive linguistics, artificial boundaries between syntax, semantics and pragmatics disappear, and humour is considered as a phenomenon in which the same cognitive-semantic strategies underlie its different instantiations: verbal and non-verbal humour, puns, etc. (Brône & Feyaerts, 2003). Cognitive linguistics provides a holistic umbrella for the study of humour, which cannot be considered as an isolated phenomenon arising from the non-prototypical use of language (or other modes of communication), but as part of a cognitive system in which humour is processed beyond mere linguistic mechanisms. In cognitive linguistics meaning stems from conceptualisation. A given situation may be conceptualised in different ways, which in turn leads to different semantic structures. Meaning is attained through construal operations (Croft & Cruse, 2004; Langacker, 2008). Conceptualisation is based on our experience of the world, which is categorized in frames. Frames are mental representations of the kind of experience elicited by a given sign or code (i.e., language, gestures, etc.) allowing us to produce and comprehend language mainly by means of analogy and comparison of the input received against that frame. In light of the above, humour related-incongruity involves two different mental representations or frames, compatible with the same input. This is the same idea underlying in Raskin's SSTH and Attardo's GTVH explained in previous sections, in which humour arises from the switch between frames or scripts in Raskin and Attardo's terminology.

Following up on the same notion, Giora (1991) introduced the concept of salience in her theory of humour, in which the shift occurs from a more salient, foregrounded —hence, prototypical, more conventionalised— interpretation of the humorous text to a second unexpected one. She developed the graded salience hypothesis (Giora, 1997), whereby there is always a more salient interpretation to an utterance, which is activated and accessed first. Only when that salient interpretation is not appropriate given the communicative context is the less salient meaning elicited. Humour arises from the ambiguity between the more salient interpretation and the actual, less salient, one leading to the humorous effect.

Giora (1991) stated that the humorous input must be sufficiently innovative to surprise, but at the same time allow for an automatic retrieval of the less salient interpretation (optimal innovation hypothesis). Although this could be applied to many instances of jokes and humorous utterances, it does not account for absurd humour, which precisely draws on completely unintended interpretations. Therefore, it can hardly fall within a more or less salient representation of any given situation. Furthermore, in her studies on negative sarcasm she acknowledged that her graded salience hypothesis cannot account for the direct, or default in her own terms, interpretation of such utterances, e.g. ‘Punctuality is not his forte’, without the need to access the most salient, non-sarcastic interpretation first (Giora, Drucker, Fein, & Mendelson, 2015; Giora, Givoni, & Fein, 2015). Giora’s (1997) hypothesis can be linked to incongruity-resolution theories of humour through her optimal innovation hypothesis (Giora et al., 2017).

She argued that humour must imply a balanced processing difficulty, i.e. maximum possible relevance with the least possible cognitive effort.

Meaning in cognitive linguistics is dynamically construed (Brône & Feyaerts, 2003). Construal mechanisms are used in humour, usually distorting some element in order to produce the humorous effect. According to these authors, humour research will be enhanced when considering construal mechanisms widely accounted for in cognitive linguistics literature, such as framing, metaphor, metonymy, conceptual integration, figure-ground patterns, etc.

Coulson (2005a) and Coulson and Oakley (2005) focused on blending as one major process underlying humour. Blending, also called conceptual integration, involves the combination of elements from different cognitive representations or mental spaces. Typically, there is at least two input spaces, a generic space—a representation of an underlying structure common to all spaces—and the blended space, with elements from all input spaces and which can develop a semantic structure of its own (Fauconnier & Turner, 1998, 2002).

Coulson (2005b) argued against Sperber and Wilson (1986) echoic mention to explain sarcasm, whereby sarcastic utterances are not used, but mentioned, thus signalling the distance between the speaker and what is being said. Speakers are just recalling a previous event, statement, etc. to convey their attitude towards it. Coulson claimed, however, that the echoic mention cannot account for all instances of irony, such as insincere questions or overpolite requests. She then moved on to assess pretence theory: previous statements,

norms, etc. are not echoed but alluded. The pretence space is shared by interlocutors and communication is based on Clark's (1996) model of layered meaning: one layer being the actual communication space, the other the pretence space. She finally proposed a variant of pretence theory: the space structuring model (Coulson & Oakley, 2005), which accounts for irony resorting to blending, considering that Clark's layers are very similar to Fauconnier's mental spaces.

Sarcasm and irony are explained by the fact that the listener is confronted to a blend which needs to be separated into the two input spaces: an expected reaction and a counterfactual trigger space (Coulson, 2005b). This is similar to Kyratzis (2003) notion of de-blending referring to metaphors. Coulson (2005b) further argued that humour and sarcasm are indirect forms of communication which trigger multiple mental spaces at the same time. Humorous or sarcastic blends usually foreground less salient constituents of the input spaces. This disanalogy is based on what's expected/unexpected in humour, and what is normative/non-normative in sarcasm. Coulson claimed that this interplay clearly points to the fact that language use is not limited to the transfer of linguistic signs, but is rather based on human experience and interaction. Humorous, i.e. insincere, NBF communication, is possible because meaning is dynamically construed in large conceptual structures.

The relation of metaphor and humour has also been explored under cognitive linguistics. Metaphor is a conceptualisation mechanism, whereby we assign features of a familiar phenomenon to a less well-known one in order to better comprehend the latter, a process known as mapping (Lakoff & Johnson, 1980). Source and target domains are therefore blended, whereas in humour the

opposite —often, but not always— occurs, and attention is precisely drawn to the boundaries between both domains in a process of de-blending (Kyratzis, 2003). Analysing GGTV from a cognitive perspective, Attardo et al. (2003) considered structure mapping to establish similarities between opposing scripts as the key element allowing incongruity resolution.

Other construal operations such as metonymy or viewpoint can also be resources drawn upon in order to produce humour. Different viewpoints imply different conceptualizations, therefore different mental spaces (Brône & Feyaerts, 2003). Hyper-understanding and misunderstanding are hailed as examples of humour based on a shift in viewpoint. At the same time, these phenomena are accounted for as instantiations of figure-ground reversal mechanisms. Metonymy, as yet another construal mechanism underlying humour, is thoroughly explained in Brône and Feyaerts (2003). Their analysis relied on the “image-schematic value of metonymy as a prominent reference-point phenomenon” (p. 22). They drew on Langacker’s (2008) reference-point notion as a salient easily coded concept used to access another less salient one, as well as on the non-bona-fide nature of humorous communication to identify four different strategies whereby metonymy-based reasoning is used for humorous communication: profiling of non-salient referent points, suppression of salient reference points, compression of two conceptual entities into a single reference point, and distortion of the prototypical causal logic.

This brief overview shows how the notion of meaning construal mechanisms, along with key notions in cognitive linguistics such as mental spaces, frames, blending, mapping, etc., can be used to shed light on the

production and comprehension of humour providing an overarching framework in which previous theories of humour can be brought together. Regardless of the approach taken to tackle humour studies, certain recurrent shared notions emerge:

- a) The need for ambiguity based on the interplay between two different possible interpretations of the humorous text. This ambiguity may rely on lexical, semantic or pragmatic processes, but can ultimately be accounted for resorting to the notion of mental spaces, triggered by any previously-listed process. Varied terminology is used depending on the theoretical approach taken: scripts (Raskin, 1985), frames (Fillmore, 1976), image-schemas (Cienki, 2007), make-sense frame (Yus, 2016), reference-points (Langacker, 1987), etc. Despite the varied terminology used, the underlying principle is basically the same (Raskin 1985; Giora, 1991; Attardo, 2001; Brône & Feyaerts, 2003, etc.).
- b) One of those spaces is more salient, i.e. more easily accessible or more prototypical (Giora, 1991, 1997, 2002; Coulson, 2005b).
- c) Incongruity arises from the mapping of elements from the salient space into the less salient one (or vice-versa), or as a result of foregrounding disanalogies in an emergent blended space in which the incongruity is resolved (Brône & Feyaerts, 2003; Coulson, 2005b, 2005c; Veale et al., 2006; Yus, 2016).

2.5.3.1 Humour in interaction from a cognitive perspective

The social function of humour is widely acknowledged in the literature (Morreal, 1983; Raskin, 1985; Attardo, 2001; Hay, 2001; Dynel, 2009). Although utterances and texts may be created with a humorous intent, this humour will ultimately be realised only upon perception as such by the hearer (Hay, 2001). The interactional value of humour has received much attention in recent literature (Holmes, 2000; Hay 2001; Baxter, 2002; Holmes & Marra, 2002; Archakis & Tsakona, 2005, etc.). Baxter (2002) classifies broad categories of humour in interaction, which she refers to as conversational joking types: joking (jab lines), teasing and self-denigrating joking. Different types of humour in interaction have been studied. What follows is an account of the most relevant literature on interactional humour from a cognitive perspective.

Hyper-understanding (Veale et al., 2006) and misunderstanding (Brône, 2008) are hailed as two major kinds of interactional humour. Hyper-understanding stems from the speaker's ability to reverse the intending meaning of the previous speaker's utterance, exploiting weak points found in it to achieve a humorous effect. Misunderstanding was defined by Brône (2008, p. 2027) as a "genuine misinterpretation of a previous utterance by a character in the fictional world". I hold the view, however, that misunderstanding may arise and have an (unintended) humorous effect in any given situation and not necessarily in the fictional world. Both phenomenon imply a conflict between different viewpoints.

Brône (2008), Feyaerts (2013), and Tabacaru and Lemmens (2014) resorted to Clark's layering model of communication (1996) to account for these two phenomena. It is argued that Clark's model can be used to explain any type of non-serious language, such as fiction, irony, sarcasm, over- and

understatements, rhetorical questions, etc. Clark's model depicts discourse as multiple layers sharing a surface represented by the communicative situation.

Brône combined this layered-meaning structure of discourse with Fauconnier's Mental Space Theory (Fauconnier, 1994). Fauconnier's mental spaces are "small conceptual packets constructed as we think and talk, for purposes of local understanding and action" (Fauconnier & Turner, 2003, p. 58). They allow us to structure discourse. Brône argued that the mental spaces elicited in humorous communication are distinct but connected across the discourse layers. In Brône's view, Clark's layer 1 (the communicative situation) is equivalent to the discourse base space in Mental Space Theory.

Both models are combined to account for the conflict of viewpoints in hyper-understanding and misunderstanding. The humorous effect of misunderstanding is explained by means of the clash between the salient interpretation favoured by contextual elements, and the misinterpretation (non-salient interpretation) of one of the participants in the interaction. Hyper-understanding, in turn, is explained resorting to a figure-ground reversal mechanism (Croft & Cruse, 2004), whereby weak points in the previous speaker's utterance are exposed to achieve a humorous effect. Both phenomena are metarepresentations of discourse: hyperunderstanding involves the creation of a pretence space in which the speaker dissociates him or herself from what has been said by their counterpart. Misunderstanding, on the contrary, relies on erroneous mapping of features to the interlocutor's own space.

Coulson (2005a) resorted to blending and mental spaces to account for conversational humour. She argued that humorous blends are shaped by the demands of conversation: the need to maintain relevance implies that speakers must take at least one element of the input blend created by the previous speaker. On the other hand, the requirement to contribute new information is met by bringing in a new input within the activated shared larger cognitive structure to produce a new blend.

Feyaerts (2013) tackled spontaneous conversational humour from a socio-cognitive perspective as the background for the collection of the CORINTH corpus of interactional spontaneous humour in Flemish (Feyaerts, Speelman, Brône, & Oben, 2010). He argued that meaning is not just a process of conceptualisation but is also the result of interactive negotiation between interlocutors in conversation. He focused on the intersubjective aspects of meaning, defined by the interaction between interlocutors.

He defined intersubjectivity as the ability to figure out the mental spaces represented by our interlocutors. Meaning construction in interaction will largely depend on the common ground shared by the interlocutors (Clark, 1996). The mind-reading ability advocated by Yus (2016) is reminiscent of this idea.

Veale et al. (2006) argued that humour is first and foremost a social-interactional phenomenon. They focused on adversarial humour, studying it from a cognitive perspective. They defined trumping as a series of humorous exchanges subverting the use of language, whose interpretation involves thorough knowledge of the interlocutors, therefore representing a highly complex

form of interactional humour. They drew on previously studied mechanisms to account for adversarial humour: Attardo's GTVH figure-ground reversal mechanism (2001) and Langacker's conceptual profiling (Langacker, 1991). These mechanisms are general meaning construal mechanisms, not only applying to humour.

The vast shared knowledge required for adversarial humour to occur is explained by means of the Current Discourse Space (CDS), defined by Langacker (2001, p. 144) as "the mental space comprising those elements and relations construed as being shared by the speaker and hearer as a basis for communication at a given moment in the flow of discourse"; that is, the background setting containing all shared information and cognitive structures that will be rearranged as communication flows leading to the dynamic creation of meaning through different construal mechanisms.

Meaning, therefore, is dynamically created relying on the relation between the profile (figure) and the base (ground) of the conceptual structure linked to a linguistic unit. Profiled, more salient aspects of a given conceptual structure can be manipulated by speakers to initially mislead the hearer before achieving a humorous effect surprising them with the non-profiled elements of the structure (ground).

On a more social-based perspective, Archakis and Tsakona (2005) considered humour to be an expression of identity, understood as a dynamic process linked to discursive strategies whereby speakers construct their social identity in that particular situation (Holmes & Marra, 2002). Holmes (2000), in

turn, argued that humour has a social function of marking solidarity, and can also serve either to strengthen or level out power imbalances. She studied humorous interaction at the workplace.

Whenever interaction takes place among interlocutors, a negotiation process is entailed whereby turn-taking is organised and each speaker's discourse structured for communication to be successful. Humorous communication is no exception, all the more so as for humour to be realised, it must be somehow acknowledged by the counterpart in the conversation. The strategies implemented to acknowledge humorous utterances are called humour support (Hay, 2001). Hay listed the following humour support strategies: (a) laughter, (b) echoing the humorous utterance, (c) contribute to humour (with a new humorous turn), (d) offer sympathy or contradict self-deprecating humour, (e) overlap and heighten involvement in the conversation.

I endorse the view that cognitive linguistics provides the best framework to account for humorous communication (Veale et al., 2006). In my opinion, most of the work on humour reviewed above can be brought together and explained resorting to the following cognitive linguistics notions: (a) Langacker's Current Discourse Space (CDS) (Langacker, 2001), (b) Fauconnier and Turner's Conceptual Integration Theory (Fauconnier & Turner, 2002), (c) Construal mechanisms (Croft & Cruse, 2004).

2.6 Conclusion

In this chapter, I have reviewed the literature and current state of the art on verbal humour studies. First, a brief overview of classical humour theories has

been provided, outlining the three major approaches to humour: incongruity, superiority and release theories (Attardo, 1994). Then, a review of multimodal humour studies is also included (Attardo et al., 2003; Pickering et al., 2009; Bryant, 2010; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011; Attardo et al., 2013; Tabacaru & Lemmens, 2014, etc.), in which a clear gap emerges with regard to the study of multimodal markers —intentional signals— or indices —unintentional cues— of humour in spontaneous communication, as opposed to staged humour.

With regards to the multimodality of humour in interaction, the literature offers conflicting results, with no multimodal cues consistently associated to humour found to date, despite the intuitive belief that humour should be delivered as a highly marked form of discourse (Gironzetti, 2017). Having said that, the idea of contrast, signalled through prosody or gestures in order to mark humour, emerges from different studies as a way to showcase certain utterances, as compared to the speaker's baseline, in order to convey the metamessage that they must not be taken at face value, and that they are ironical, humorous, etc. (Attardo et al., 2003; Bryant, 2010; Urios-Aparisi & Wagner, 2011).

Subsequently, linguistic theories of humour are reviewed. Raskin's (1985) Semantic Script Theory, with a clear semantic-pragmatic perspective, is described. Two main notions underpin this theory: that of script, as a pre-structured knowledge unit, and the concept of script-opposition, which in turn draws on the idea of incongruity-based humour. The chapter then explores Attardo's (2001) General Theory of Verbal Humour, which expands Raskin's SSTH to include a more pragmatic and cognitive perspective, laying down six

knowledge resources around which humorous texts revolve: script opposition, target of the joke, narrative strategy, language, situation, and logical mechanism. The latter is considered to be the link to a more cognitive-based approach of humour (Veale et al., 2006).

The idea of opposition, ambiguity in interpretations or incongruity-resolution is also present in other theories of humour mentioned, both from a Relevance Theory point of view (Yus, 2016) and with a more cognitive stance (Giora, 1991; Brône & Feyaerts, 2003; Coulson, 2005b; Veale et al., 2006). This cognitive perspective is favoured in the subsequent review of studies concerning humour in interaction (Holmes, 2000; Hay, 2001; Holmes & Marra, 2002; Archakis & Tsakona, 2005; Veale et al., 2006; Feyaerts, 2013; Tabacaru & Lemmens, 2014), as the standpoint taken by this thesis, with the aim to conduct a multimodal and cognitive analysis of spontaneous humorous communication.

Chapter 3

Multimodality

3.1 Introduction

Multimodality refers to the combination of different semiotic resources or modes/modalities in communicative events (Adami, 2016). A multimodal analysis of discourse and human communication brings in information conveyed through different modalities, such as gesture, sound, proxemics, etc. All these modalities, and crucially, how they interplay, contribute to communication, which is never monomodal. In addition, each modality has its own affordances. Therefore, a comprehensive account of communication can only be achieved through careful exploration of how these modalities operate and relate to each other (Kress & van Leeuwen, 2001; Norris, 2004a; Adami & Kress, 2014; Forceville, 2014; Adami, 2016).

Multimodality is not a theoretical framework as such, but rather an approach or perspective applied to the study of communication at large, e.g. language, pictures, films, multimedia resources, etc. The main basis of this paradigm is that all modalities through which information is conveyed must be accounted for in a comprehensive analysis of communication. Furthermore, overall multimodal meaning is larger than the sum of the meaning communicated by each single modality, as their interplay also plays a role in meaning production and interpretation.

3.2 Multimodality

Kress and van Leeuwen (2001, p. 20) defined multimodality “as the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined”. They defined modes as “semiotic resources which allow for the simultaneous realisation of discourses and types of interaction” (Kress & van Leeuwen, 2001, p. 21). However, multiple and divergent definitions of mode (or modality) have been proposed in the literature. Adami (2016, p. 451) alluded to modes as resources “to make meaning and to express and shape values, ideologies, and power relations”. Given the lack of widespread agreement on a single definition of mode —or modality, as both terms are used interchangeably in the literature— multimodal studies draw on categories of modes/modalities (Forceville, 2009, 2014), among which written and spoken language (e.g.; Pennock-Speck & Del Saz-Rubio, 2009; Bryant, 2010; Attardo, Wagner, & Urios-Aparisi, 2011), visuals (e.g. Del Saz-Rubio & Pennock-Speck, 2009; Forceville, 2009, 2014, 2015), gestures (e.g. Muller & Cienki, 2009; Wagner, Malisz, & Kopp, 2014), olfactory stimuli (Velasco-Sacristán & Fuertes-Olivera, 2006), etc.

Multimodal studies share four major assumptions (Adami, 2016 p. 451):

- a) Communication is always multimodal.
- b) Language-only analyses cannot fully explain meaning.
- c) Each mode has specific affordances, stemming from the nature of the mode (“their materiality”) and the socio-cultural background against which they have develop their communicative value.

- d) All co-occurring modes simultaneously contribute to meaning-making.

Although rooted in social semiotics and currently leaning towards cognitive linguistics, multimodal analyses are also linked to conversation analysis of social interaction (Goodwin & Goodwin, 1992; C. Goodwin, 2013; Goodwin & Cekaite, 2013), within the framework of conversational analysis (Depperman, 2013a, 2013b). A multimodal interaction analysis focus on how interlocutors communicate, making and expressing semantic and pragmatic meaning resorting to different modes, as raising eyebrows to mark sarcastic speech (Tabacaru & Lemmens, 2014), changing prosody to signal irony or sarcasm (Cheang & Pell, 2009; Bryant, 2010), punctuating speech with head movements (Hadar, Steiner, Grant, & Clifford, 1983; Hadar, Steiner, & Grant, 1984; Hadar, Steiner, & Clifford, 1985), using their hands (Kendon, 2004; McNeil, 2014), etc.

3.3 Multimodality and semiotics

Adami (2016) defined multimodality as the combination of various semiotic modes or meaning-making resources in a given communicative event (still or moving images, colour, speech, gestures, etc.). Semiotics can be defined as the general science of representation (Bergman, 2007), whereas semiology “occupies the part of semiotics which relates either to conventional communication, or intentional communication” (Daylight, 2012 p. 37). Syntax, semantics and pragmatics can be viewed as three branches of semiotics, although this discipline, in the Peircean tradition, is not restricted to language (Bergman, 2007). It includes any sign with communicative potential, and from this perspective, it could merge with multimodal studies of communication. That is

precisely the approach taken by Bezemer, Jewit, Diamantopolou, Kress, and Mavers (2012). With regards to the relation between linguistics and semiotics, most semioticians accept the view held by Saussure (1993), placing linguistics as a branch of semiotics, with “language as the semiotic system *par excellence*” (Lévi-Strauss, 1972, p. 48), which leads to semiotics considerably drawing on linguistic concepts to analyse and study other semiotic systems (Chandler, 2007).

Multimodality was traditionally grounded in semiotics before gradually adopting a more cognitive linguistics perspective (Zlatev, 2012; Cienki & Muller, 2014; Forceville, 2015). Feng and O’Halloran (2015) drew an interesting distinction in their study of multimodal metaphor, claiming that a cognitive linguistic approach takes it as the expression of metaphorical thinking, whereas semioticians focus on how these metaphors are visually expressed.

If we understand modes or modalities as realms in which different types of meaning-motivated signs with communicative potential are made, there is a clear intersection between semiotics, cognitive linguistics and multimodality, where each of these disciplines would apply a different approach to analysing a communicative event. Semiotics is concerned with the nature of the signs, whereas cognitive linguistics looks into how meaning conveyed through those signs is construed and motivated, and multimodality addresses how meaning is conveyed through the various modes through which signs are expressed and how they interplay.

3.3.1 Semiotics vs semiology

Semiotics, as understood in the Peircean tradition, is the general science of representation, including as potential objects of study “all sensory stimuli that could create another idea in the receiver’s mind.” (Daylight, 2012 p. 37). Saussure’s semiology (1993), on the other hand, is concerned primarily, albeit not exclusively, with the study of language as a code system, and more broadly with intentional signs. In addition, semiology limits itself to the study of signs representing idealised concepts, i.e. in Saussure’s terms *langue*, as opposed to *parole*¹, whereas semiotics is grounded in our experience of the world, delving into how signs found in the environment are perceived consciously (Daylight, 2012).

Another difference lies in the structure of the sign. Whereas signs in semiology are dual, made of a signifier —the expression of the sign— and a signified —the concept represented by the signifier—, in Peircean semiotics (Peirce, 1965) signs have a tripartite configuration, made of a) the representamen: the expression of the sign; b) the interpretant: the interpretation of the sign by the receiver; c) the object: what is represented by the representamen.

Its abstraction of meaning from the real world and our own experience of it excludes semiology as a valid approach for my study, as I intend to disentangle

¹ Saussure defines *langue* as “a social product of the faculty of speech and a collection of necessary conventions that have been adopted by a social body to permit individuals to exercise that faculty” (1966. p. 9), whereas *parole* refers to the various manifestations of speech (ibid.). For Saussure, only *langue* can be studied, as the “norm of all other manifestations of speech” (ibid. p.9). Saussure argues that speech, on the contrary, is too heterogeneous to constitute a valid object of study, as its unity cannot be discovered.

how humour is multimodally conveyed in actual spontaneous communication, thus falling within the realm of *parole*. As for semiotics, while some of its fundamental tenets mentioned above are fully compatible with the theoretical framework applied in this dissertation, I cannot share the need for a tripartite structure of the sign, insofar as the interpretant presupposes the existence of a receiver. This could be accommodated if, as claimed by Brandt and Brandt (2005), communication includes those situations in which the receiver and the producer of the sign are the same person; that is, instances of inner thoughts. Some authors, though, do not consider this situation to involve semiotics and advocate for the need of three separable constituents of the sign (Deely, 1990; Dalylight, 2012).

3.3.2 Multimodal social semiotics

Bezemer et al. (2012) defined multimodal social semiotics as follows:

It is a theoretical perspective that brings all socially organized resources that people use to make meaning into one descriptive and analytical domain. These resources include modes such as image, writing, gesture, gaze, speech, posture; and media such as screens, 3 D forms of various kinds, books, notes and notebooks. (p. 1)

Together with Kress (2015), they are major advocates of applying this multimodal social semiotic approach to the study of communication. A social semiotic approach considers signs to be socially shaped and motivated. Signs are not arbitrary, and as such they provide insight into the sign-maker motivations, views, context, etc. (Adami, 2016).

Multimodal social semiotics draws on Halliday's (1978) systemic functional linguistics (Adami, 2016), where language is conceived as a network of related systems, subject to choice, serving three metafunctions: ideational meaning, interpersonal meaning, and textual meaning (Feng & O'Halloran, 2015). The ideational metafunction is related to the way we conceive and represent the world; the textual meaning refers to verbal expression (Halliday, 1978; 2004); finally, the interpersonal, or inter-subjective, metafunction lies on the relationship between interlocutors and their communicative role (Zlatev, 2008). Social semioticians argue that these principles apply to non-linguistic resources as well.

Kress (2015) considered each modality as a partial means of communication with its own affordances, which leads him to claim that language (speech and writing) is also partial and does not always meet all communication needs alone. He therefore resorted to multimodality as the best approach to account for different forms of meaning-making. He further argued that as meaning-making is driven by semiotic entities, processes and relations, the theories and tools necessary to study this phenomenon can be drawn from social semiotics. In his vision, multimodality is a domain of study or application of the larger framework provided by social semiotics (Adami, 2016; Bezemer et al., 2012).

3.4 Multimodality and pragmatics

Wilson and Sperber (2004 p. 607) defined Relevance Theory as follows:

Relevance theory may be seen as an attempt to work out in detail one of Grice's central claims: that an essential feature of most

human communication, both verbal and non-verbal, is the expression and recognition of intentions (Grice 1989: Essays 1-7, 14, 18; Retrospective Epilogue). In developing this claim, Grice laid the foundations for an inferential model of communication, an alternative to the classical code model. According to the code model, a communicator encodes her intended message into a signal, which is decoded by the audience using an identical copy of the code. According to the inferential model, a communicator provides evidence of her intention to convey a certain meaning, which is inferred by the audience on the basis of the evidence provided.

Arguably the most comprehensive attempt so far to apply Relevance Theory—hereinafter RT—(Sperber & Wilson, 1986) to multimodal communication has been undertaken by Forceville, who has mainly focused on visual communication (see Forceville, 2014), although he believes his reasoning can be extended to other modes or combinations thereof (Forceville, 2014). He argued that RT can account for multimodal communication provided the specific affordances of each modality are taken into account.

Although Sperber and Wilson's Relevance Theory (1986) aims at accounting for any instance of ostensive communication, it is primarily based on the verbal modality. If we accept face-to-face communication as prototypical, it seems clear that inferences to arrive to explicatures and implicatures will be derived not only from verbal messages, but also from other modalities such as prosody, gestures, facial expressions, etc. When communication occurs in less prototypical settings, e.g. telephone, email, voicemails, etc., the modes through

which inferences are made are significantly reduced. Things are even more complex in cases of mass communication, where there are at least two simultaneous addressees as an audience (Forceville, 2014). Each addressee will draw their own inferences according to their cognitive environment, which may differ from one addressee to another.

Forceville (2014) proposed to discard cases of given-away information from the phenomenon of communication, as they are not ostensive, and just happen to convey information without the sender being aware of it. However, I hold the view that unintended messages do not necessarily have diminished communicative value. I argue that non-intentionally conveyed information is indeed communication, and that excluding it as an object of study undermines our comprehension of the phenomenon as a whole, all the more so when studying multimodal communication, whose aim, by definition, is to provide a holistic analysis of the phenomenon of communication regardless of the modality in which it occurs.

In Forceville's (2014) view, multimodal communication falls within RT insofar as it is ostensive —produced intentionally by the maker of the picture, sound, etc.—, and it reflects the desire of the communicator to convey information at no great cognitive effort on the part of the audience. Therefore, it is a case of relevant communication.

He further claimed that, in the case of visuals, explicatures can be inferred, as there is a number of propositions arising from the non-verbal elements encoded in a given picture which can be judged as true or false. Implicatures, in

turn, arise from the context provided by the genre to which a particular modality is ascribed. Addressees will derive implicatures matching the visual with the expectations and knowledge they have of the appropriate genre (Forceville, 2009, 2014; Forceville & Clark, 2014).

Although he made a strong case regarding the potential of pictures to yield both implicatures and explicatures, I find his initial claim that his analysis could be further extended to other modes more problematic. I strongly believe that gestures and prosody do play a major role in allowing for the inference of implicatures, as when a change in voice quality or a gesture may signal a humorous or ironic statement (Cheang & Pell, 2009; Bryant, 2010; Tabacaru & Lemmens, 2014; etc.). As for explicatures, if we understand them as propositions, as explained by Sperber and Wilson (1986), I wonder to what extent we can derive them from intonational cues, for example. Or whether a gesture or facial expression can be said to give rise to true/false propositions.

Given that each modality has different affordances, and therefore different communicative value, I doubt that general claims applicable across the board can be made. Having said that, Forceville's (Forceville, 2009, 2014; Forceville & Clark, 2014) pioneer work raises valid and interesting questions for multimodal research. Clearly, further work is needed to disentangle the role played by modalities based on their respective affordances and how that role is transformed when several modalities simultaneously interplay.

3.5 Multimodality and cognitive linguistics

As seen before, multimodality is concerned with the analysis of communication from the point of view of how different semiotic modes interact (Pinar Sanz, 2015). A multimodal study of interaction addresses what is expressed and perceived in a communicative event through different modes, e.g. gaze, posture, gestures, etc. (Norris, 2013). Norris further claimed that a full account of human interaction is not possible if the mind is not considered, as “a person always thinks, perceives, and/or feels something when interacting with others, and at least some of these thoughts, perceptions, and/or feelings are communicated through a person’s actions” (Norris, 2004a, p. xi). Given that cognitive linguistics addresses the “interrelation between thought, meaning, and linguistic structure” (Lee, 2001 p.1), the relevance of conducting both a cognitive linguistic and multimodal analysis for my present research on spontaneous communication is warranted.

Although a bias towards the study of conceptualisation and categorization, to the detriment of discourse or interaction has been raised concerning cognitive linguistics (Nuyts, 2007), there is a clear link between thought patterns in the mind and their linguistic and discursive representations, based on the dynamic construction of meaning and conceptual frames in language-use events (Cap, 2017).

Furthermore, Langacker (2001) himself rejected the claim that a cognitive approach to language involves a lack of concern for social interaction and discourse (p.185). He argued that, in fact, a cognitive approach to language presupposes a close relation between linguistic units and discourse, as the

former are “abstracted from usage events, retaining as part of their value any recurring facet of the interactive and discourse context” (p.143). He continued to mention that conceptualisation occurs through multiple channels, including vocalization, which, again, to my view, supports my stance of combining a multimodal and cognitive linguistic analysis for the purposes of this dissertation.

One of the major tenets of cognitive linguistics is that meaning is conceptualisation, and that conceptualisation stems from our experience of the world, which is necessarily embodied (Johnson, 1987; Croft & Cruse, 2004; Geraeerts, 2008). Therefore, it can be assumed that people with different bodies will have different experiences, and they will consequently think differently (Casasanto, 2013). In fact, McNeill (2013, p.29) claimed that it is possible to “see someone’s thought in gesture”. Gestures (face and body movements), along with prosody and speech, conform the various modalities that interplay in face-to-face interaction, and that we can resort to in order to try to ensure successful communication (Stivers & Sidnell, 2005; Poggi, 2013).

3.6 Multimodality in face-to-face interaction

Gestures and prosody are two main modalities interplaying in face-to-face communication. Two major influences in gesture studies are Adam Kendon (2002, 2004, 2013) and David McNeill (1985, 1992, 2000, 2005, 2013, 2014). Both argued that gestures point to a representation of meaning at multiple levels simultaneously, expressed by means of co-speech gestures and speech. McNeill (2013) claimed that gestures and speech are co-expressive, representing the same idea but not necessarily the same aspect, each drawing on its own semiotic value. Gestures are the expression of imagistic thinking, while language draws

on analytic thought (McNeill, 2005). Intonation, in turn, provides observers with the possibility to gain an insight into the mind, feelings, and attitudes of speakers (Cruttenden, 1986; Wennerstrom, 2001). As Bolinger (1986, p. vii) put it, “intonation is the spoken counterpart of facial expression and physical gesture”.

In this section, I provide a review of the literature pertaining studies into prosody and gestures, placing special emphasis on their interplay with speech.

3.6.1 Prosody

Prosody can be defined as the set of vocal features such as intonation, pause, stress, tempo, loudness, rhythm, and voice quality (Wennerstrom, 2001). According to this definition, intonation, i.e. changes in pitch as we speak, is therefore just one of the features of prosody, but both terms are often used interchangeably in the literature. Hirst and Di Cristo (1999) defined intonation—in its broadest sense; that is, understood as prosody—as the most universal, yet language-specific, feature of human language. For them, its universality lies in the fact that intonation is a characteristic of all languages, whose functions seem to be roughly the same across a wide spectrum of very different languages. Wennerstrom (2001) argued that it would not be surprising to find common features in the intonation of different languages, as intonation may be the “grammaticalisation of the prosodic features of the cries of our pre-linguistic ancestors” (ibid., p. 7).

The literature widely acknowledges prosody as having discursive value, and as primarily conveying the speaker’s feelings (Bolinger, 1986; Brazil, 1997; Wennerstrom, 2001, etc.). This function is rooted in our biological origins, in our

need to read other people's feelings and behavioural cues in order to avert potential danger, but it has reached a certain degree of conventionalisation. Prosody may also have a syntactic function (Halliday, 1976), albeit subordinate to the emotional one (Bolinger, 1986, 1989; Wennerstrom, 2001). Brown, Currie, and Kenworthy (1980), however, pointed to the difficulty to systematically attribute affect-meaning to prosody, when attitude is expressed by means of lexical words as well. Furthermore, considering intonation in its restricted sense, they claimed that voice quality, i.e. a set of prosodic features other than pitch patterns, whereby the voice may change into coarse, whisper, etc., is a far more effective way of conveying attitude.

No significant differences have been found in the pragmatic use of prosody related to dialect or gender. Beyond the obvious differences in terms of pitch range between men and women or dialectic prosodic patterns, the fundamental principles of the pragmatic value of prosody stand more or less universally, even across languages (Bolinger, 1989).

3.6.1.1 Prosody and its parts

As said before, prosody can be defined as a set of suprasegmental oral features stretching over one or more than one consecutive utterances (Cruttenden, 1986). The prosodic features most frequently used for linguistic analysis are pitch, defined as the varying height of the voice tone over one or more syllables; loudness, i.e., changes in volume and intensity over one or more syllables; and length, understood as the relative duration of one or more than one consecutive syllables as compared with their immediate environment or with that

set of syllables in other context. Pauses, tempo, rhythm and voice quality are also considered prosodic features (Cruttenden, 1986).

3.6.1.1.1 Pitch

Pitch is the most salient feature of intonation. The most widespread acoustic measure used for pitch is fundamental frequency (F0), resulting from the repetitions of waveforms created by the vibration of the vocal cords. This implies that voiceless sounds do not produce fundamental frequency, although this does not affect the overall perception of the pitch of what is being said (Cruttenden, 1986). Intonation stems from upward and downward changes in pitch, and it is responsible for the melody of speech (Bolinger, 1986).

Pitch-accented words stand out from the utterance baseline. Humans are very sensitive to changes in frequency, so pitch changes need to be abrupt but not necessary long to be perceived. Also, they can occur in either direction, with higher or lower F0 values, as what matters to place emphasis is the pitch obstruction, i.e. the shift, rather than its directionality. Pitch is the most prominent characteristic of intonation patterns. In fact, intonation arises from the succession of pitch patterns (Cruttenden, 1986; Wennerstrom, 2001). The curve resulting from changes in pitch over a period of time is called pitch contour. The pitch range used at the onset of an utterance is named key. It may be high, to draw attention or to express emotion; low, showing boredom and lack of interest, or mid key, where there is no change in pitch (Wennerstrom, 2001). Of particular interest is the concord tone, which refers to a key, i.e. onset pitch, at the same F0 as the

terminal pitch used by the interlocutor's previous utterance. It signals agreement, understanding, engagement, or lack thereof (Wennerstrom & Siegel, 2003).

Intonational features form intonation groups, which loosely correspond to constituents of the utterance. These, in turn, link to form paratones, which can be defined as intonational paragraphs. Wennerstrom and Siegel (2003) distinguished between high paratones, to mark changes in topic, and low paratones, signalling digressions or side remarks. Both paratones and intonation groups tend to start with rising intonation and end in falling intonation. Finally, intonation phrases include a more or less continuous pitch contour with at least the following features: onset key, pitch boundary and a number of pitch accents. Intonational phrases decline in pitch, and are slower towards the end. They are also known in the literature as tone units, intonational units, or intonation group. With regards to the study of spontaneous speech, it is highly difficult to define boundaries between intonation phrases (Brown et al., 1980; Wennerstrom, 2001).

3.6.1.1.2 Pauses

One element used in the literature to delimit intonation phrases is pause (Brown et al., 1980; Archakis et al., 2010). Brown et al. define substantial pauses (0.6 – 0.8 sec) for boundaries as opposed to short pauses (0.4 – 0.6) to allow time for mental processing. Pauses can be filled, e.g. with sounds like “uh”, or unfilled, i.e. silent. According to Cruttenden (1986), pauses tend to occur in three different instances: at major constituent boundaries, e.g.: between subject and predicate; before words of high lexical content —Cruttenden linked these pauses

to the need for speakers to stop to find a word—, and right after the first word in an intonation group, due to errors in performance. In order to establish the boundaries of intonation groups, a number of elements have to be simultaneously taken into account: pause, syntactic and semantic utterance constituents, and length of syllables, as the last syllable of an intonation group tends to be lengthened. A revealing indicator of boundaries between intonation groups is any change in pitch level and direction when occurring in non-stressed syllables. Boundaries at the end of intonation phrases are characterised by falling pitch and longer final syllables. Rising boundaries usually yield turn. Pitch at boundaries is known as terminal pitch (Bolinger, 1986; Cruttenden, 1986; Wennerstrom 2001).

3.6.1.1.3 Pitch accent and stress

While some authors (Bolinger, 1986; Cruttenden, 1986) argued that there is usually just one prominent pitch accent in an intonation phrase, usually occurring towards the end of the sentence, Wennerstrom (2001) claimed that several pitch accents can occur in one intonation phrase. Accented words are associated with new information, whereas the baseline represents previously-known information. The nucleus is the most prominent pitch accent in a sentence, where the speaker wants to place the focus. It corresponds to the pitch accented syllable of the most prominent word in an intonation phrase. Accent is different from stress, which refers to the prominence given to an element in an utterance not through pitch, but by means of loudness, or changes in speech rate (Cruttenden, 1986). Furthermore, certain qualifiers may contribute to nuance messages even more, such as using whispering for confidence, a husky voice for sexual appeal or tremolo voice for uncontrolled emotions (Bolinger, 1986).

3.6.1.2. Prosody and its uses

Wennerstrom (2001) borrowed from the work of three major previous references in prosody studies (Halliday, 1976; Pierrehumbert, 1980; Brazil, 1984; 1997) to conclude that prosody is compositional, given that it encompasses different features; cohesive, for it is used as a main tool to structure both information and discourse in an oral text; interactional, as it is a mechanism employed in conversation for turn-taking, endorsing or rejecting other speaker's speech; and expressive, because it is a conduit through which we express and acknowledge feelings.

Prosody mainly conveys meaning at discourse level, and it also carries affective meaning (Cruttenden, 1986; Wennerstrom, 2001; Pennock-Speck & Del Saz-Rubio, 2009). Prosody, as gestures or body movements, is used to foreground part of what is being communicated (Rockwell, 2000; Bryant, 2010; Muller, 2013a). It is also related to grammar, although more on the basis of correlation than on a univocal relation or dependence. It often contributes to establishing segments of speech, marking the limits between constituents, and underlining major constituents apart from secondary ones, thus helping to structure information (Wennerstrom & Siegel, 2003).

Falling pitch, pauses and lower speech rate are usually indicators of the end of an utterance, although they are not in themselves defining elements of such boundaries. In fact, pauses may occur when speakers stop to retrieve a certain word, and falling pitch may be used to draw attention to a particular element of the utterance. Only in combination with other cues, e.g. syntactic

and/or semantic, or gestures, can they be taken as markers of a speech-segment end (Brazil, 1997; Wennerstrom, 2001).

Cruttenden (1986) associated prosody to the pragmatic, rather than the literal, meaning of an utterance. However, Wennerstrom (2001) wondered whether prosody has in fact pragmatic or semantic value, as it is encoded in the verbal message, although it does not affect propositional meaning. In my opinion, semantic and pragmatic meanings are not two distinct separable entities. I believe meaning and conceptualisation are grounded in discourse and our experience of the world (Langacker, 2001), which blurs the boundaries between semantics and pragmatics (Croft & Cruse, 2004).

Prosody is one modality that can be recruited for the purposes of communication (Forceville, 2009), which serves to reinforce a certain layer or aspect of meaning in a given utterance, such as when a husky voice is used to signal sexual appeal (Bolinger, 1986). The same way that, according to Relevance Theory (Sperber & Wilson, 1986), humans are endowed with the ability to read minds in terms of assuming what inferences are meant by the communicator or will be made by the listener (Yus, 2003, 2016), our ability lies in processing the different multimodal inputs available in a given communicative situation to ultimately understand what is being communicated, through inferences, taking utterances at face value or (un)consciously decoding multimodal signs such as prosody or gestures (Coulson & Wu, 2014).

One of the most straightforward pragmatic uses of intonation is to ease the cognitive effort needed from the listener to process what is being said by

showcasing the most relevant part of the utterance (Bolinger, 1989; Wennerstrom, 2001). Low pitch is associated with known or previously given information, requiring less saliency as more easily retrievable by the listener. Wennerstrom and Siegel (2003) considered this feature to be a window onto the mind of the speaker to determine what assumptions are made during communication in terms of what knowledge is shared with interlocutors.

Intonation also contributes to increasing the illocutionary and perlocutionary effects of speech. When searching for this effect, stress words and syllables tend to be terminal, that is, to occur at the end of an utterance, as last impressions are remembered best (Cruttenden, 1986). As for the affective value of prosody, rising intonation is linked to emotion, and excitement; it has an emphatic value. Falling intonation, on the contrary, is related to rational, controlled, unemphatic statements and boredom (Wennerstrom, 2001). A wide pitch range points to a fluctuation of feelings (Bolinger, 1986; Brazil, 1997; Wennerstrom, 2001).

Bolinger (1986) explained the emotional function of intonation as evolving from biological needs. He claimed that “intonation is a part of a gestural complex whose primitive and still surviving function is the signalling of emotion” (1986, p. 195). Therefore, both intonation patterns and gestures serve to convey and recognise behavioural cues. Given the link between intonation, gestures and language, though, the question that arises is to what extent intonation patterns and gestures have become conventionalised. Since it is possible to mimic and fake intonation and gestures to resemble spontaneous communication, it seems clear that despite their biological origin, both gestures and intonation have also

evolved into culture-bound conventionalisations, at least in part (Bolinger, 1986; Attardo, Wagner, & Urios-Aparisi, 2011). I believe that exploring the interplay between prosody, verbal language and gestures can lead to a better understanding of human communication (Pennock-Speck & Del Saz-Rubio, 2009; Mondada, 2013).

3.6.1.3 Prosody and gestures

According to Forceville (2009), prosody is one of the many modalities through which messages are conveyed. Although recent studies have shown that each mode plays an important role in communication, the interplay between different modes remains largely unexplored (Muller, 2013b). In Chapter 2, I reviewed studies delving into the relationship between prosody, gestures, and humorous communication. This section deals with literature covering the link between prosody and gestures at large, e.g. head movements, face expression, body posture, etc.

Kendon (2004) and McNeill (2000) have both studied how prosody and gestures align, finding a close relationship in terms of timing in the realisation of gestures and prosody. Gesture strokes, for example, defined as the phase involving the highest effort and an information density peak (Kendon, 2004), are completed either shortly before or at the same time that stressed syllables (Loehr, 2004; Cvejic, Kim, & Davis, 2010; Flecha-García, 2010). Bolinger (1986) went as far as to suggest that both gestures and intonation move in parallel, following the up and down metaphor, but this claim has not been supported empirically by later studies (McClave, 1998; Loehr, 2004).

McClave (1998) showed that gestures and intonation may correlate in terms of direction to mark salience of a certain point in the utterance, but that such correlation is not automatic. He did find that the stroke of propositional gestures tends to coincide with the most prominent pitch-accented syllable of a tone unit, i.e. the nucleus. McClave's findings are in line with Loehr's (2004; 2012), who concludes that intonation and gesture patterns, regardless of their nature, tend to co-occur. He found that gestural apex, i.e. the point of maximal gestural excursion (Wagner et al., 2014) tend to coincide with pitch accents, and that gestural phrases typically align with intermediate intonational phrases, with gestural boundaries slightly preceding intonational ones. Furthermore, Loehr posited that gesture and intonation occasionally express the same pragmatic meaning and are rhythmically linked. However, yet again no correlation was found between gesture and intonation type, or between gesture and pitch upward or downward movement. Loehr further claimed that these findings support the theory of a common origin for gesture and speech, co-occurring as expressions of an idea unit, similar to Kendon's (2004) idea unit or McNeill's (2014) growth point. Ultimately, speakers have several modalities at their disposal for communication, which may not always surface or co-occur. When they do, though, they are connected in terms of timing, structure and meaning.

The close link between prosody, face gestures, and head movements was also showcased by Cvejic et al. (2010), who argue that people can draw prosodic information from face and head movements alone, and found out that even just nods and tilts were reliable cues for prosodic information. Cavé et al. (1996) considered that gestures and prosodic features cannot be semiotically analysed

as independent items. The authors claimed that they form a global semiotic modality leading to intertwined meaning conveyed by both modes. Their findings are consistent with more recent studies into the correlation between, for example, eyebrow movements and prosody (Guaiatella, Santi, Lagrue, & Cavé, 2009; Flecha-García, 2010; Tabacaru, 2014).

Loehr (2014) showcased what the literature has so far shown regarding the interplay between prosody and gestures. First, both are synchronised at certain unit levels, e.g. pitch accent with gestural apex (Kendon, 2004), intonational phrases with gesture phrases (Loehr, 2004), etc. Furthermore, both modalities serve to regulate discourse and interaction. The interaction between prosody and gesture involves many parts of the body, such as hands, head, face, torso, etc. (Loehr, 2014). To conclude, language, gestures and prosody clearly seem to coordinate, although to date, no recurrent, one-to-one pattern governing relations between these modalities has been found (Bohle, 2014).

3.6.2 Gestures and speech

Gestures and speech are the moment-by-moment thinking that takes place while one speaks (McNeill, 2005); different modes of thinking are expressed in both modalities: imagistic in gestures, and analytic in language (McNeil, 2008). For McNeill (2005), it is possible to see someone's thought in gesture. He defines gesture as "an unwitting, non-goal-directed action orchestrated by speaker-created significances, having features of manifest expressiveness" (McNeill, 2013, p. 29). In other words, gestures are expressive actions which enact imagery. Notice that he excludes emblems, i.e. highly

conventionalised gestures; e.g. the thumbs up gesture, from his definition. Gestures and speech are co-expressive, but not necessarily conveying the same aspect of one particular idea. In their respective semiotic ways, though, they express the same underlying notion (McNeill, 2007). In any communicative situation, speakers choose characteristics that fit the conceptualisation of the event and that are encodable by language (Slobin, 1987, p. 435), which also has an impact on their gestures (McNeill, 2008).

In order to avoid confusion as a result of the polysemy of the term gesture, Kendon (2004) resorted to the term visible utterance action, to include the various ways in which bodily movement can contribute to the creation of utterances. In the same vein, Lapaire (2011) referred to gestural actions. Kendon (2013) related gestures to cognitive linguistics, as they are considered representations of image-like meaning. Visible utterance actions, therefore, point to a mental multileveled representation of meaning, organised simultaneously and expressed by way of speech and co-speech gestures. Utterances, therefore, are expressed both through speech and visible bodily action. Larger segments of discourse are typically marked by sustained changes in posture or head movements, while shorter segments are associated with shorter movements (Kendon, 2013).

Delving into the relation between speech and gestures, Müller and Cienki (2009) claimed that gestures are used for spatial expressions, re-enactment, shape and motion, whereas speech is favoured to refer to complex situations and to convey abstract ideas and arguments. In addition, they posited that gestures may be used to explicit the pragmatic value of verbal speech. Magno-Caldogneto and Poggi (1997) listed several functions governing the relation between speech

and gesture as follows: (a) repetition or redundancy, when different modalities contribute to the same meaning, e.g. a raised eyebrow coinciding with a stress word; (b) substitution, when a signal is used in place of another, as when a raised eyebrow marks a question that has not been syntactically formulated; (c) contradiction, when co-occurrent modalities express opposing meanings, such as shaking the head while saying 'yes'; (d) addition, when the signals add meaning to each other, e.g. pointing in a given direction while saying "there". Ladeweig (2014a) established a continuum of integrability between gestures and speech depending on the type of integration, the distribution of information over different modalities, and the order in which speech and gestures are deployed. Referential gestures are the most often used to replace speech (Muller & Cienki, 2009). In sum, gestures can take either a syntactic or semantic function; they can be either co-expressive or redundant, and complementary to the verbal message.

While the relation between form and meaning in language is conventionalised and arbitrary, gestures are motivated signs, whose meaning is based on form (Müller, Ladewig, & Bressemer, 2013). Cognitive linguistics is the framework in which that motivation may be thoroughly explained, seeing gestures as part of construal mechanisms such as viewpoint, metaphor, metonymy, etc. (Mittelberg, 2008; Zlatev, 2008; Cienki, 2013a; Cienki & Müller, 2014; Gibbs et al., 2018). Cienki (2013a) argued, for example, that speakers may access image schemas to produce and understand gestures. A higher level of specification in the motivation of gestures is found in Zlatev mimetic schemas (2008), defined as "preverbal concepts which possess a number of properties which can help explain the emergence of language as a conventional-normative semiotic

system” (ibid. p. 123), and which involve actions such as EAT, CRAWL, SIT, etc. Zlatev propounded that they serve to structure our thought, and therefore can also underlie gestures as their motivation.

Most gesture studies concerned with the relation between gestures and speech fall into two broad perspectives: cognitive-psychological or functional-communicative (Lapaire, 2011). This also implies the following distinction: for some authors, gestures are produced for the benefit of the hearer, to facilitate comprehension (Kendon, 1980; Clark, 1996; C. Goodwin, 2013; etc.), while other scholars favour the view that gestures are mainly produced for the benefit of the speaker, and that they point to the mental processes undergone (Alibali, 1999; Alibali, Kita, and Young, 2000, Hostetter & Alibali 2008, etc.). Lapaire (2011) argued that both strands can be combined into a cognitive-functional approach, taking gestures as having representational as well as discourse and pragmatic functions. McNeill (2008) posited that this dichotomy is unfounded, and that gestures serve both purposes. He considered an individual social duality to be inherent to gestures, which are a link between individual cognition and social interaction (McNeill, 2008, p. 54). Every thought goes through a social filter, so they are two inseparable dimensions.

With regard to the interaction between speech and gesture, the hand-in-hand hypothesis claims that gestures reflect what is simultaneously expressed in speech (So, Kita, & Goldin-Meadow, 2009). Others consider, on the contrary, that gestures and speech are complementary in terms of meaning production, and that the less costly channel in terms of production is favoured to convey a given idea (de Ruiter, Bangerter, & Dings, 2012). The proponents of the Lexical

Retrieval Hypothesis (Kraus & Hadar, 1999) posit that gestures help retrieve words (Wagner et al., 2014), whereas in an alternative view, the Information Packaging Hypothesis (Kita, 2000; Alibali et al., 2000), authors argue that speech and gestures interact earlier, at a stage in which information is packaged and distributed across different modalities. In other words, in the IPH gestures are part of the process of conceptualising something for speaking (Wagner et al., 2014). Another cognitive model of the speech-gesture relation links conceptualisation to the creation of multimodal messages (Kopp, Bergmann, & Kahl, 2013; Bergmann, Khal, & Kopp, 2013). On this account, visuo-spatial and propositional representations are both simultaneously activated and dynamically linked while thinking for speaking.

3.6.2.1 The growth point hypothesis

McNeill approached gestures from the point of view of psychology. He conceived gestures as “ingredients in an imagery-language dialectic that fuels speech and thought” (McNeill, 2008, p. 3), where the dialectic occurs between two different modes of thinking: analytic (language) and synthetic (gestures). The difference lies in the fact that while language combine semantic units to produce a whole meaning, gestures are not divisible but constitute a single symbolic unit.

For McNeill, language is both static and dynamic, and the dialectic between these dimensions give rise to an unstable state where an idea is represented in two different forms simultaneously. That is what he called the growth point (McNeill, 2013), which must be unpacked into a linguistic form to resolve the dialectic. A growth point must explain speech and gesture synchrony,

and co-expressiveness (McNeill, 2013). In his hypothesis, McNeill presupposes that speech and gestures express the same idea in different semiotic modes. They are synchronous precisely at the exact point where they are co-expressive (McNeill, 2008), which for him proves that the mind is doing the same thing in two different ways.

He further argued that the semiotic value of gestures only comes into play when an observer is taken into account. In the absence of such an observer, gestures are “part of the speaker current cognitive being” (McNeill, 2008 p. 42). That is, they are a window to the mind showcasing the speaker’s mental processes. In McNeill’s view, communication occurs when the context in which the speaker is embedded is shared with an interlocutor, who will in turn also draw from that context. He proposed the growth point hypothesis to explain how utterances are created and expressed by means of different modalities, namely gestures and speech. Recurrent-gesture features point to recurrent imagery in the speaker’s thought and they show cohesive links in discourse (McNeill & Levi, 1993).

3.6.2.2 Types of gestures

3.6.2.2.1 Early classifications

The first classification of gestures goes back to Efron (1941), later expanded and developed by Ekman and Friesen (1971), and Ekman (1979). Efron drew a line between discursive and referential gestures, which has since been widely acknowledged in the literature (Ekman, 1979; McNeill, 1992; Kendon, 2004; Bohle, 2014, etc.). For Efron, gestures can be logical-discursive,

such as baton-like or ideographic gestures, which are not referential and are linked to the conceptualisation process. The second broad category of gestures established by Efron is that of objective gestures, i.e. referential, which include deictic, iconographic, and emblematic gestures. These three types, in turn, differ in their degree of conventionality, with emblems as the most conventionalised gestures with the status of stand-alone symbols.

Ekman and Friesen (1971), and Ekman (1979) later expanded and refined Efron's classification to set up one of the most influential gesture categories in gesture studies (Bohle, 2014). They divided gestures into five broad types: a) emblems, as highly conventionalised and symbolic gestures, which can be produced and understood without speech; b) illustrators: referential, non-conventionalised, always co-occurring with speech; v) regulators, with a discourse-management function; d) affect displays, including mainly facial expressions; e) adaptors, with no communicative meaning, such as touching one's hair or nose, manipulating an object, etc.

3.6.2.2.2 *Kendon's continuum*

Kendon (1980) identified different types of gestures along a continuum, which McNeill would later call the Kendon's continuum (McNeill, 1992). As we move along the continuum, there is less need for accompanying speech, and gestures increasingly show analytic properties as those typical of language.

- Gesticulations

Gesticulations are co-speech gestures made unwittingly by the speaker. They are by far the most common type of speech in conversation. They are

synchronous or slightly precede the co-expressive part of speech they match. Gesticulations have universal and language-specific features. Typically, raised eyebrows, head tilts or deictic hand movements fall into this category.

- Speech framed gestures

Speech framed gestures are part of the sentence; they fill a grammatical slot. In other words, speech framed gestures complete the structure of that sentence. As opposed to gesticulations, they do not synchronise with co-expressive speech, but rather fill a gap (McNeill, 2008). For example, if we call someone and then motion for them to approach with our hands, we would be performing a speech framed gesture.

- Emblems

Emblems are conventionalised, culture-bound gestures, which are interpretable without speech by a certain cultural or social group (Teßendorf, 2013). Although these gestures are meaningful without the need for speech, they can co-occur with it too (McNeill, 1992). E.g., the thumbs-up gesture (Kendon, 2004).

- Pantomime

Gestures produced without speech for the purposes of telling a story, a narrative, as those performed by mimes during a show.

- Signs

Signs as used in signed language. They are lexical words with their own structure, grammar, morphology, etc. (McNeill, 2008).

3.6.2.2.3 McNeill's classification

McNeill's (1992) classification of gestures is widely acknowledged in the literature (Bohler, 2014). He categorised gestures on the basis of three criteria: form, meaning, and communicative function. McNeill considered that gesture types are dimensions (2008) rather than categories, as a gesture can belong to more than one type simultaneously without any hierarchy involved. All these gestures are either gesticulations or speech-frame gestures.

- Iconic gestures

Iconic gestures are referential and imagistic; they represent physical entities or concrete actions. They are semantically and pragmatically co-expressive with speech, being not just redundant, but complementary to it (McNeill, 1992, pp. 12-13). An example could be a hand gesture signalling a downward movement when speaking of a fall, or shaping a certain volume, etc.

- Metaphoric gestures

They are similar to iconic gestures insofar as they are referential and imagistic, but metaphoric gestures represent abstract ideas (ibid., p.14). E.g., a tight-fist gesture to refer to someone as stingy.

- Deictic

Deictic gestures can be both referential and discursive; they are used to refer to physical entities and abstract ideas (ibid., p. 18), such as when pointing to an object or backwards and forward to signal past and future.

- Beats

Beats have discursive value. They are used to punctuate or emphasise parts of speech in terms of their pragmatic-discourse value (ibid., p. 15), as when hitting a palm with the other hand to stress a particular idea or word.

- Cohesive gestures

They are used to link parts of the conversation which are on the same theme but may be temporally separated (ibid. p 16). E.g. moving a hand away and backwards when drawing on something that had been said before in conversation.

Müller (as cited in Ladewig, 2014b) introduced a new type of gesture: recurrent gestures, different from iconic or metaphorical gestures due to their conventional character and pragmatic function. Conventionalisation of recurrent gestures relies on the pairing of the basic form of that gesture with a semantic value (Kousidis, Malisz, Wagner, Schlangen, & Ladewig, 2014). They are recurrent across different speakers in a given speech community. They differ from emblems, though, in that their meaning is schematic and the form-based motivation of the gesture is still apparent (Ladewig, 2014b). Cyclic gestures (Ladewig, 2014b), the holding away gesture (Bressemer & Müller, 2014), or the palm up open hand gesture (Kendon, 2004), are examples of recurrent gestures.

3.6.2.3 Functions of gestures

As seen in the previous classifications, broadly speaking, gestures can perform the following functions: discursive, referential, conversational, and interactional (Ladewig, 2014a). Kendon (2013) claimed that longer segments of

discourse are associated with changes in posture or head movements, while shorter segments are linked to shorter movements. According to Kendon (2004), gestures may serve to add semantic value, make meaning more specific, represent an object or spatial relationships, or act as deictics. In addition, gestures can have semantic, syntactic, and discursive functions. Referential gestures, for example, can be used to replace speech taking over the syntactic function of the linguistic element replaced (Müller et al., 2013; Lapaire, 2011). Lausber (2013) argued that right-handers use the right-hand for concrete, pictorial representations, while left-hand gestures are linked to prosodic function, emotional processes and metaphoric thinking. This lateralisation would be reversed for left-handers (Casasanto, 2013).

Regarding facial expressions, Poggi and Pelachaud (1998), drawing on previous literature (Ekman 1979, 1982; Castelfranchi & Poggi, 1990) distinguished the following functions: (1) affective display, with seven possible prototypical face expressions for emotions, i.e. happiness, sadness, anger, fear, disgust, surprise and embarrassment; (2) syntactic function, to emphasise words, questions, etc.; (3) dialogic function with regards to turn taking, signalled with eye and head movements; (4) referring function, to allude to an emotion that is not currently being felt by the speaker, with deictic gaze, or facial emblems, even iconic eye movements, e.g. squinting as opposed to wide open; (5) attitude display, to convey social attitude towards the addressee, e.g. anger, ordering, imploring, etc.

In addition, gestures may be used to reinforce the expression of abstract ideas (McNeill, 1985; Cienki & Müller, 2008; Wagner et al., 2014). As seen before,

gestures also contribute to organising speech and to structure the information being conveyed (McNeill, 1992). Very often, the interactional value of gestures will merge with their function as supporting more abstract ideas (Wagner et al., 2014).

3.7 Conclusion

This chapter devoted to multimodality has addressed the basic tenets of this broad perspective of analysis (Adami, 2016), as well as its link to various related disciplines. It begins with an introduction on multimodality, outlining the shift this discipline has sustained from its roots in semiotics (Kress & van Leeuwen, 2001) towards a more cognitive stance (Forceville, 2009). Multimodality is defined as the study of communication that brings together all modes or modalities (Forceville, 2009) through which information is conveyed (Adami, 2016), with modes or modalities—used interchangeably—understood as semiotic channels.

Next, the relation between multimodality and semiotics is explored, taking semiotics as the general science of representation (Bergman, 2007), and concerned with the study of any sign with communicative potential (Kress & van Leeuwen, 2001). The difference between semiotics and semiology is briefly covered, before moving on to explaining the field of multimodal social semiotics, as the approach favoured to study communication by pioneer advocates of multimodality (Bezemer et al., 2012; Kress, 2015).

The chapter moves on to address how multimodality and pragmatics have been brought together in studying communication, where Forceville's (2014;

Forceville & Clark, 2014) research clearly stands out. His work on pictures applies Relevant Theory (Sperber & Wilson, 1986) to account for implicatures and explicatures of the visual modality. This account is followed by a section exploring the link between multimodality and cognitive linguistics, which warrants the relevance of combining both approaches to look into any instance of communication.

In this chapter, I have also surveyed the literature on the multimodality of face-to-face interaction, which, by definition, is inherently multimodal, focusing on prosody and gestures as two of the main modalities at play. I have explored the role of prosody as contributing to information-structure and having discursive and affective value (Cruttenden, 1986; Bolinger, 1986; Brazil, 1997; Wennerstrom, 2001; Pennock-Speck & Del Saz-Rubio, 2009). Furthermore, as explained above, ample evidence points to complex and fine-tuned coordination between language, prosody, and gestures, with users exploiting the different affordances entailed in each modality (McNeill, 2000; Kendon, 2004; Loehr, 2004; Cvejic et al., 2010; Flecha-García, 2010; Ladewig 2014a, 2014b; Tabacaru, 2014, etc.).

Gestures are a window onto the mind (McNeill, 2008), and a review of authors supporting the idea of a close interplay of linguistic and imagistic conceptualisation has been included (Kendon, 2004, 2013; McNeill, 2005, 2008; 2013; Muller, 2009; Lapaire, 2011, etc.). Gestures and language are intertwined in early stages of conceptualisation, as encapsulated in McNeill's growth-point hypothesis (McNeill, 2013). Both speech and gestures activate conceptual representations and hearers can apprehend information from both modalities to

form a cognitive model of the referents included in discourse (Coulson & Wu, 2014).

As with the previous chapter, the overview offered here points to the need to apply a holistic view to analysing face-to-face communication. Human interaction can only be fully comprehended taking into account elements from multimodal discourse pragmatics and the “physical, contextual, and sociocultural anchoring of human cognition and situated meaning-making” (Mittelberg, 2017, p. 203).

Chapter 4

Cognitive linguistics

4.1 Introduction

Cognitive linguistics emerged as a distinct field in the mid-80's (Evans & Green, 2006), as a reaction against Chomskyan Generative Grammar linguistics (Chomsky 1957, 1965, 1986), contested on the basis of two main issues: that language is an independent module separated from other cognitive functions, and that language is best understood at the level of competence (as opposed to performance²), thus taking as an object of study people's abstract, idealised, linguistic knowledge (Geraeerts & Cuyckens, 2007; Ibarretxe & Valenzuela, 2012). Seminal publications and the work of key authors such as Langacker (1987), Lakoff and Johnson (1980), Lakoff (1987), Johnson (1987) and Talmy (2000a, 2000b) pictured language and its study in a very different light. They argued that linguistic structure stems from cognition, in the sense that a given linguistic expression is associated with a particular way in which a situation is conceptualised (Lee, 2001, p.1). In other words, cognitive linguists endorse the idea that thought, meaning, and linguistic structure are closely intertwined. In fact, cognitive linguistics gives meaning a central role in language, as opposed to the

² Chomsky (1965) establishes that "[I]inguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance." (p. 3). He further posits a distinction between *competence* as "the speaker-hearer's knowledge of their language" (p.4) and *performance*, understood as "the actual use of language in concrete situations" (ibid.).

formalist and structuralist view held by Generative Grammar (Talmy, 2000a, 2000b; Geraeerts, 2008).

Cognitive linguists advocate for a bodily and experiential basis of meaning, cognition and language (Ziemky, Zlatev, & Roslyn, 2007); i.e. our way of conceptualising the world is determined by the nature of our body and how we physically interact with our environment (Johnson, 1987). Furthermore, this embodied perspective of meaning and cognition entails a deviation from logical or truth-conditional semantics, whereby meaning is assessed in terms of its true or false representation of an objective reality in the world (Croft & Cruse, 2004). Alternatively, cognitive linguists argue against this objectivist approach, as we can only construe that reality through our bodily experience (Evans & Green, 2006; Geraeerts & Cuyckers, 2007).

Cognitive linguistics is not a theoretical framework, but rather an umbrella linguistic research approach encompassing various sub-fields, theories, and lines of research sharing the previously explained assumptions, including, but not limited to, cognitive semantics (Talmy 2000a, 2000b), conceptual metaphor theory (CMT) (Lakoff & Johnson, 1980), mental spaces, conceptual integration networks, and blending (Fauconnier, 1981; 1994; Fauconnier & Turner 1998, 2002, 2003; Coulson, 2005a; Coulson & Oakley, 2005), cognitive grammar (Langacker, 1987), construction grammar (Goldberg, 2006), frame semantics (Fillmore, 1976), etc. Broadly speaking, all cognitive linguistic areas are held together by the “belief that linguistic knowledge involves not only knowledge of the language but knowledge of the world as mediated by language”. (Geraeerts

& Cuyckens, 2007, p.7). From this assumption, a series of cognitive linguistics basic tenets emerge (Croft & Cruse, 2004; Geraeerts, 2008 pp. 4-5):

- a) Language is part and parcel of our cognitive system and it is therefore not autonomous nor unrelated to other cognitive functions. Linguistic meaning is encyclopaedic, encompassing our bodily experience of the world, as well as our social and cultural identity.
- b) Meaning (ultimately prefiguring language units and structure) is perspectival, as it stems from our conceptualisation of the world, which is necessarily embodied and experiential.
- c) Linguistic meaning is, therefore, necessarily usage-based.
- d) As a result, linguistic meaning is dynamic and flexible.

This cognitive approach to language blurs the boundaries between traditional dichotomies in linguistics such as langue/parole, performance/competence or semantics/pragmatics³.

4.2 Key concepts in cognitive linguistics

Certain basic elements are at the core of cognitive semantics (as the foundation of cognitive linguistics). Clausner and Croft (1999, p. 3) summarised them as follows.

- a) Concept: a concept is a unit of mental representation. For example, 'laughter' would be a concept, as it elicits a mental representation of a

³ The dichotomy holds for traditional views which restrain semantics to the meaning of linguistic units, whereas pragmatics is concerned with meaning in context.

piece of reality we ascribe to the category of laughter. In cognitive semantics, categorization is based on prototypical features of the given category, as boundaries are not always clear cut (Evans & Green, 2006). For example, laughter may be fake or spontaneous, or have different acoustic properties (Bryant & Gibbs Jr., 2015). In spite of the various manifestations of laughter possible, they would be ascribed to the same category and belong to the same concept.

b) Domain: knowledge structure that provides background information allowing us to comprehend concepts, as they do not occur in isolation. Clausner and Croft (1999) equate this term to Fillmore's (1968, 1975) frames or Lakoff's (1987) idealised cognitive model. Langacker (1987) posited that domains may combine in larger, more complex structures called domain matrices. Broadly speaking, domains are mental schematisations of specific situations or experiences (Cienki, 2007; Ibarretxe & Valenzuela, 2012). For example, in order to understand concepts such as 'far' or 'near', we need the larger knowledge structure of SPACE. Drawing on the previous example of 'laughter', to understand it we need to relate to domains such as FACE or SOUND.

c) Construal: operation whereby our experience of the world is conceptualised (Croft & Cruse, 2004). Examples of construal mechanisms are metaphor, metonymy, etc. In fact, so-called figurative language is no longer considered figure of speech, but figure of thought (Lakoff & Johnson, 1980). Construals encapsulate elements of the conceptualisation that cannot be fully grasped exclusively resorting to

the properties of the object conceptualised, e.g. perspective, subjectivity, etc. Construals invariably invoke the agent of conceptualisation as an input to the meaning being construed (Verhagen, 2007).

Consider example (8) below, extracted from the sample under study. Alec Baldwin was speaking about not being able to go to fancy parties due to his tight work schedule. To the remark made by the host, Stephen Colbert, alluding to Baldwin's hard life conditions, he replies construing the situation from the point of view of his wife, not his own, which is clear through the use of 'you' and the vocative 'Alec' to refer to himself. Instead of referring to his wife's comments as indirect speech, e.g. 'my wife would say no one feels sorry for me', he assumes his wife's viewpoint. That introduces a different perspective to the communicative event:

(8) Stephen Colbert: What a hard life you have.

Alec Baldwin: But, but, but my wife would say "Nobody feels sorry for you, Alec".

d) Category: a group of concepts, more or less prototypical, with relations established between the central (prototypical) members —e.g., a mockingbird, for the category "bird"— and those at the periphery (less prototypical) of the category —e.g. a hen or an ostrich, for that same category of "bird"—. (Rosch & Mervis, 1975; Langacker, 1987)

4.2.1 Mental spaces

Mental spaces are defined as "small conceptual packets constructed as we think and talk, for purposes of local understanding and action —they are very

partial assemblies containing elements, structured by frames and cognitive models” (Fauconnier & Turner, 2003, p. 58). Mental spaces are therefore dynamic and change as thought and discourse progress.

Crucially, these spaces can be conceptually blended in order to process thought and language. Blending is a basic mental operation, whereby new meaning is created through the merging and combination of at least two mental spaces. An often-quoted example of blending is (9) below (Grady, Oakley, & Coulson, 1999; Brandt & Brandt, 2005; Fauconnier & Turner, 2002):

(9) This surgeon is a butcher.

In (9), two mental spaces are blended, i.e. one of a surgeon, another of a butcher. To illustrate it briefly, the process of conceptual integration in the blend involves the surgeon’s space, probably with an operation theatre, and the scalpel, with the aim of healing a patient, whereas the butcher’s would have a knife, a butcher’s shop and the objective of cutting meat. Elements from both spaces map onto the new, blended space, with the surgeon trying to pursue their aim with the butcher’s tools. In addition, the blend includes the new emergent feature of incompetence, resulting from a clash in surgeon’s and butcher’s objectives (Grady et al., 1999) not available in any of the input spaces (Fauconnier & Turner, 2002).

Therefore, any blended space is linked to the input spaces through the mapping, or correspondence, of selected source features into the blended space, which may have its new distinct emergent structure. According to Fauconnier and Turner (2002, 2003), mental spaces are blended and connected in integration

networks, containing at least four mental spaces: two input spaces, a generic space with the structure shared by those inputs, and the resulting blended space. Brandt and Brandt (2005, p. 232) argued, however, that the shared structure captured in the generic space “is specified by what is situationally relevant. It is thus not cognitively realistic that this structure exists in the mind as a definite list of entities and relations independent of a goal, a purpose, motivating the conceptualizer to evoke these similarities”.

Blended spaces can in turn serve as input spaces leading to multiple blends and highly complex integration networks. Finally, an important property of blending networks is the Access Principle, whereby an element in a space may be accessed by naming or describing their counterpart in another space (Fauconnier, 1997; Coulson & Oakley, 2005).

4.2.2 Conceptual mapping

The process whereby features of a source domain are assigned to a target domain. Mapping can be achieved by way of metaphor or metonymy, for example, although it can also account for other figures of thought such as irony, hyperbole, etc. (Ruiz de Mendoza, 2014).

4.2.3 Image schemas

Cognitive linguistics posits that our knowledge is grounded in our perceptual interactions, manipulation of objects and bodily experience. This, in turn, structures that knowledge through recurrent patterns, which are experiential gestalts; e.g. our experience of volume, weight, speed, distance, etc. (Hampe,

2005a). Image schemas are those experiential gestalts. The term was coined by Johnson (1987) and Lakoff (1987). Johnson (2005, p.19) defined them as follows:

[I]mage schemas are the recurring patterns of our sensory-motor experience by means of which we can make sense of [our] experience and reason about it, and that can also be recruited to structure abstracts concepts and to carry out inferences about abstract domains of thought.

Image schemas can include information in different modalities, which can be tapped into for abstract conceptualisation and reasoning (Cienki, 2005). They are a basic tenet of cognitive linguistics insofar as they explain how meaning and reasoning are rooted in our embodied experience. Body and mind, i.e. sensory-motor experience and thought, cannot be separated. In other words, our construal of the world around us is shaped by the way we interact with it through our body and perception. Image schemas are schematic pre-conceptual structures stemming from that interaction, i.e. our experience of the world, basically drawing on motion and spatial relations (Talmy, 2000a, 2000b; Hampe, 2005b). The inventory of image schemas identified in the literature is by no means a closed list, but some of the core elements put forward by Lakoff (1987) and Johnson (1987), and supported by later studies (Lakoff & Turner, 1989; Cienki, 1997; Clausner & Croft, 1999) are listed in (10) and (11):

(10) CONTAINMENT/CONTAINER, PATH/SOURCE-PATH-GOAL, LINK,
PART-WHOLE, CENTER-PERIPHERY, BALANCE

(11) FORCE schemas: ENABLEMENT, BLOCKAGE, COUNTERFORCE, ATTRACTION, COMPULSION, RESTRAINT, REMOVAL, DIVERSION
(Hampe, 2005a p.2)

Consequently, examples such as (12), and (13) are grounded on the CONTAINER, and PATH/SOURCE-PATH-GOAL, respectively; i.e.: our bodily and perceptual experience of paths –leading to a destination/goal– and containers –in/out– helps us construe and understand these expressions:

(12) He is in a very difficult situation.

(13) She is far from getting the promotion.

4.2.4 Construal mechanisms

Croft and Cruse (2004) offered a list of construal mechanisms, understood as instantiations of broader cognitive processes, thus highlighting the link between language and cognition (Table 1). Construal mechanisms allow us to conceptualise the world, therefore to create meaning. As mentioned in Chapter 2, many authors resort to construal mechanisms or to notions derived thereof in order to account for humour: figure-ground reversal (Brône & Feyaerts, 2003; Attardo, Pickering, & Baker, 2011), salience (Giora, 1991, 1997), metaphor deblending (Kyratzis, 2003), metonymy (Brône & Feyaerts, 2003), etc. The present study annotated each humorous instance found in the sample on the basis of Croft and Cruse's classification of construal mechanisms listed in Table 1, in order to explain how these instances are cognitively motivated.

Table 3.1 *Linguistic construal operations as instances of general cognitive processes*

I. Attention/salience
A. Selection
1. Profiling
2. Metonymy
B. Scope (dominion)
1. Scope of predication
2. Search domains
3. Accessibility
C. Scalar adjustment
1. Quantitative (abstraction)
2. Qualitative (schematization)
D. Dynamic
1. Fictive motion
2. Summary/sequential scanning
II. Judgement/comparison (including identity image schemas)
A. Categorization (framing)
B. Metaphor
C. Figure/ground
III. Perspective/situatedness
A. Viewpoint
1. Vantage point
2. Orientation
B. Deixis
1. Spatiotemporal (including spatial image schemas)
2. Epistemic (common ground)
3. Empathy
C. Subjectivity/objectivity
IV. Constitution/Gestalt (including most other image schemas)
A. Structural schematization
1. Individuation (boundedness, unity/multiplicity, etc.)
2. Topological/geometric schematization (container, etc.)
3. Scale
B. Force dynamics
C. Relationality (entity/interconnection)

Table 1. List of construal mechanisms (Croft & Cruse, 2004 p. 46)

Whenever we produce an utterance, we “unconsciously structure every aspect of the experience we intend to convey” (Croft & Cruse, 2004, p. 40). A construal can be defined as an operation whereby a speaker creates and delivers a conceptual representation, which, in turn, will have an effect on how that representation is construed by the hearer (Evans & Green, 2006). According to Croft and Cruse (2004, p.45), construal operations are manifestations of four basic cognitive abilities: attention, comparison, perspective, and constitution. In

turn, these cognitive abilities are paired to world properties such as salience, or to the philosophical concepts of judgement, situatedness, and Gelstat.

A definition of each construal type under these categories follows. All definitions are based on Croft and Cruse's book on cognitive linguistics (2004, pp.46 – 73).

4.2.4.1 Attention / salience

Attention is a human cognitive ability in which certain conceptual structures are activated in a neural network. Linked to attention is also salience, understood as the natural property of phenomena and objects in the world which somehow call more attention to themselves.

4.2.4.1.1 Selection

Selection can be defined as our ability to focus only on certain parts of an experience which are relevant at a given time, while ignoring concurrent aspects deemed irrelevant.

4.2.4.1.1.1 Profiling

Profiling is the process whereby a certain concept in a semantic frame is given more prominence. A typical example is that of profiling the arc or radius in a circle (Croft & Cruse, 2004, p. 47). Profiling may also occur within a concept itself, in which only certain aspects of the domain are highlighted, as in example (14) below (ibid., p. 48):

(14) Where is the Sunday Times?

Have you read the Sunday Times?

In (14), the first question profiles the Sunday Times as a physical object, whereas in the second question, it is the text contained in the paper and not the physical object itself which matters.

In the sample under study, profiling is the second most frequent construal mechanism identified. This may be consistent with the widely-held view that humour lies in an incongruity, which may well be produced and resolved by way of shifting the focus from an expected –salient– meaning to a less salient one (Giora, 1997; Giora, Drucker, Fein, & Mendelson, 2015). Consider example (15) below, extracted from our sample.

(15) Amy Schumer: This past winter Canada Goose gave me a coat.

Stephen Colbert: Oh, those are great.

Amy Schumer: And, yeah, super warm. And I was psyched 'cause it was my favourite. It was like...free.

In this example, Amy Schumer has profiled just one feature of the coat she had been given (one “facet of the domain matrix”, to use Croft and Cruse terminology [ibid p.47]), i.e.: that it was free, so she didn’t have to pay for it, as opposed to the coat being warm, trendy, comfortable, etc., which could be hailed as more traditional reasons to like a piece of clothing. The humorous effect stems from the fact that those expectations on the reasons why clothes are appreciated are broken (all the more so coming from a celebrity, which is expected to be able to afford a winter coat).

4.2.4.1.1.2 Metonymy

Broadly speaking, a cognitive definition of metonymy is the ability to select “a different contextually salient concept profile in a domain or domain matrix than the one usually symbolized by the word” – A for B— (Croft & Cruse, 2004, p. 48). That is, the typical concept profile is changed to be used to refer to something else, as in (16) below (ibid. p.48), in which ‘french fries’ [*sic.*] is used by a waiter to refer to the person who ordered such food:

(16) That french fries is getting impatient.

In our sample, the only utterance that could arguably be considered an example of metonymy is (17), which is Elon Musk’s answer to a question by Stephen Colbert on whether a new charger for electric cars to be released by Tesla is safe:

(17) Elon Musk: For the prototype at least, I would recommend not dropping anything when you're near it.

In my view, the profile concept of “dropping anything near it” is metonymic for the consequence, which would allegedly be to be attacked by the new device being tested. That is, the salient meaning of “dropping anything near it” is shifted to include the whole event, which would roughly include two stages: dropping something near the prototype and being attacked by it.

4.2.4.1.2 *Scope of attention*

Scope of attention refers to the area surrounding the focus of attention (what is selected or foregrounded), which contains elements that can be accessible in terms of attention.

4.2.4.1.2.1 *Scope of predication*

The scope of predication means that presuppositions entailed in a concept are more easily accessible than more indirectly related domains. Consider example (18) below.

Amy Schumer is speaking about an incident with animal-welfare activists, who criticised her for wearing a coat with coyote fur.

(18) Amy Schumer: What they do is...To make the hood, they kill coyotes. And I was like...yeah...I was like "Oh, my god". If I made a list of the animals I care about more than coyotes... It would be a list of every animal. I don't care.

In this case, with the counterfactual "If I made a list of the animals...", she establishes a scope of predication in which the domain of an actual list with a given number of animals (potentially a long list, as it was clear in the interview that she did not share the activists' concerns) is easily accessible. As I see it, humour relies on the exaggeration entailed in including every single animal on the list, as a way of making clear that she does not care at all about the use of coyote fur in the coats at stake.

4.2.4.1.2.2 *Search domains*

Search domains refer to achieving the construal of a concept by narrowing down the scope of preceding expressions. This is very clear in locative expressions, in which the order in which they are presented is crucial to pinpoint the concept to be delimited.

- (19) The money is on the top shelf, in the kitchen, under the counter, behind the meat grinder, in the lefthand cabinet. (Croft & Cruse, 2004, p. 51).

4.2.4.1.2.3 *Accessibility*

The last construal operation related to scope is that of accessibility, whereby a domain is accessible by way of a referent point in the scope of attention (that is, part of the concept that is around the focus). Again, no such examples have been found in the sample.

4.2.4.1.3 *Scalar adjustment*

Scalar adjustment involves a change in the scale of attention. Consider example (20) below (Croft & Cruse, 2004, p. 52):

- (20) a. She ran across the field.
b. She ran through the field.

The use of 'across' in (20a) conveys the idea of thickness or volume, as that preposition is used for three-dimensional spaces. In (20b), however, the field is portrayed as two-dimensional by way of the preposition 'through'. Both

sentences describe the same scene, but the way they have been construed and presented directs our attention to different aspects of the field.

Scalar adjustment can be achieved quantitatively or qualitatively. Quantitative scalar adjustment is achieved by means of changing the granularity of the concept, as in example (20) above, in which (20a) offers a more coarse-grained depiction of the field, whereas (20b) is more of a close-up. Qualitative adjustment, in turn, refers to the construal of a concept resorting to a wider category. Whereas quantitative adjustment involves construing along a scale, qualitative adjustment relies on the loss of irrelevant properties.

4.2.4.1.4 Dynamic attention

Attention can lead to static construals of a concept, as with focus, scope and scale, but it can also be dynamic, construing a given scene moving the attention across it.

4.2.4.1.4.1 Fictive motion

Fictive motion involves the dynamic construal of a static scene, as in (21) (Croft & Cruse, 2004, p. 53):

- (21) The road climbs through the valley and then climbs over the high mountains.

Clearly, the road itself is not moving nor going anywhere, but in (21) it is dynamically construed, attributing movement to the static road.

4.2.4.1.4.2 Summary / sequential scanning

A scene can also be construed through scanning, conceptualising it in its entirety (summary scanning), as in ‘The collapse of the building’, or sequentially, e.g.: The building collapsed.

Consider example (22) from the sample, in which Sigourney Weaver is talking about a recent trip to Cuba and about what happened with one of the classic cars they hired.

(22) Sigourney Weaver: Ours [car] stalled and we all had to get out and give a little push...

In (22) the situation is sequentially scanned and it is presented in such a way that we construe the situation over a period of time.

4.2.4.2 Judgement / comparison

Judgement as a philosophical concept is linked to the cognitive process of comparison, which can be performed through the following construal mechanisms.

4.2.4.2.1 Categorization

Categorization consists of “applying a word, morpheme or construction to a particular experience to be communicated” (Croft & Cruse, 2004, p.54). To do so, a process of comparison between the experience at hand and previous experiences is undertaken, and judgement is made to assign the present experience to the same category as the previous ones.

Consider example (23), taken from the sample, in which Alec Baldwin refers to the casual clothes he is wearing for the interview:

(23) Alec Baldwin: I call it the Springsteen look.

In (23), Alec Baldwin is framing the way he looks with the clothes he is wearing by aligning his style to famous American singer Bruce Springsteen's typical casual look. He has construed his look by first comparing it to a well-known singer who usually dresses informally, and then Alec Baldwin has judged it was a legitimate comparison and that today's style can be assigned to Springsteen's usual look.

4.2.4.2.2 *Metaphor*

Metaphor is an operation whereby a certain phenomenon is construed in terms of another –A is B—. A relationship is established between a source domain, which will serve to facilitate the construal, and a target domain, which is the one described by the metaphor. Usually, the source domain is more tangible and easy to understand than the target domain, hence its use as an explanatory tool.

Consider example (24), from Alison Janney's interview. She has been asked how she feels about Trump's Administration, given that, according to the host, she had some relevant expertise gathered when she impersonated CJ, the character of White House Press Secretary in the popular series *The West Wing*. She refers to Mr Sean Spicer's job (Trump's Press Secretary at the time) resorting to metaphor, in which the source domain is that of a plane, where the leading

team is in the cockpit. The plane in the target domain is linked to the country, the cockpit to the Administration, and the pilot to the President of the US.

(24) Alison Janney: It's not easy. And, and, and, I, as CJ, I had Aaron Sorkin in my cockpit. And he [Mr Spencer] has someone I'm not sure knows how to flight a plane.

4.2.4.2.3 Figure-ground alignment

The relation between figure and ground is used to construe spatial relations, whereby the position of an object —figure— is expressed in relation to another —ground—, such as in (25), in which 'boy' is the figure, and 'house', the ground.

(25) The boy is in the house.

Figure-ground relations can also be established between events, as in (24) (Croft & Cruse, 2004, p. 57), where 'sleeping' —the ground— is subordinate to 'dreaming' —the figure:

(26) He dreamed while he slept.

4.2.4.3 Perspective / situatedness

Perspective does not only relate to spatial position. It can also be based on our knowledge, attitudes and beliefs. The philosophical notion of situatedness is linked to perspective, as it refers to being in a particular place, at a given time, with all the spatial and non-spatial elements of the context involved.

4.2.4.3.1 *Viewpoint*

Viewpoint is a focal adjustment based on vantage point and orientation. The vantage point will determine a certain foreground-background setting, whereas orientation refers to the vertical dimension.

Consider example (27) below, taken from the sample, in which British Daniel Kaluuya talks about the extent to which experiences by black people in Britain and in the US are alike:

(27) Daniel Kaluuya: White people say some very weird stuff to you.

(27) is a clear example in which the situation to be conveyed is construed and expressed from the standpoint of the speaker as a black person. It is from his perspective (as a proxy of black people's perspective as a community) that the utterance is produced.

4.2.4.3.2 *Deixis*

Deixis is the operation whereby a speaker will refer to something in the scene or event described, either relative to the speech event or to another deictic centre established by the speaker. The use of pronouns, demonstratives, verb tenses are expressions of deictic construals, as they are always used relative to the speaker.

Deixis can be used to establish spatiotemporal references (28). It can also be based on the knowledge shared by interlocutors, as well as common beliefs and attitudes, in which case we talk of epistemic deixis. In (29), Daniel Kaluuya is making reference to a previous situation in the interview, which in turn referred

to a scene in the film *Get Out*, where a man who would later prove to be extremely racist said to the black main character that he would have voted for Obama a third time. The common ground, i.e., shared knowledge, assumptions, etc., will constrain what is expressed and how it is expressed. In addition, speakers may take the perspective of the interlocutor to construe and express a situation or event by means of applying empathy deixis, such as in (30), where VP Biden takes on the perspective and attitude of the audience (or what he believes will be so).

(28) Susan Sarandon: Oh my God, look at your ceiling.

(29) Daniel Kaluuya: [On Stephen Colbert remark that he had seen *Get Out* three times]. Thank you, thank you so much. I mean, I'm so happy...

(30) VP Joseph Biden: Now, you're all gonna laugh when I say this.

4.2.4.3.3 Subjectivity / objectivity

Subjectivity/objectivity refers to how a situation is conceptualised when the speaker is included. Consider example (31) (Croft & Cruse, 2004, p. 62):

(31) [*said by mother to child*]

a. Don't lie to me.

b. Don't lie to your mother.

(31a) is an example of subjective construal, whereas (31b) implies a process of objectivization in which the mother refers to herself as an entity that is not part of the speech event.

Example (32) below is taken from our sample:

(32) Elon Musk: [on question whether he is a super hero or super villain] I try to do useful things.

Stephen Colbert is trying to portrait Elon Musk as a super hero or super villain, given that he is a billionaire, he develops new technologies and he has the ambition to change the world. Elon Musk replies by playing down all those facts and defining himself just as someone who tries to do useful things. He is positioning himself at the centre of the speech event.

In (33), however, also taken from the sample, Amy Schumer refers to herself as a third person, 'home-girl', therefore placing herself outside the scene as a different person from the one wearing a designer dress for a fancy event portrayed in the picture she is mentioning.

(33) Amy Schumer: But I'll tell you what. Home-girl was sweating.

All construals based on perspective depend on us being in the world in a particular manner (Croft & Cruse, 2004, p. 63). That places ourselves in a given vantage point, both conceptually and spatiotemporally, i.e. deixis. In addition, as participants in a speech event, our role in it will determine the status of the situation to be conveyed —epistemic deixis—, as well as our attitude towards it —empathy—, and how we present ourselves in that situation —subjectivity—.

4.2.4.4 Constitution / Gelstat

Constitution / Gelstat imply construal operations whereby the structure of the entities in a given scene is conceptualised. Such entities are constituted, that

is, given a Gestalt or structure, with the definition of relations of proximity, boundaries, and continuation. It is the most basic level of construing our experience (Croft & Cruse, 2004, p. 63).

4.2.5 Conceptual metaphor

From a multimodal perspective, arguably the most widely researched construal mechanism is metaphor (Velasco-Sacristán & Fuertes-Oliveira, 2006; Forceville & Urios-Aparisi, 2009; Velasco-Sacristán, 2010), with metonymy receiving increasing attention (Brône & Feyaerts, 2003; Velasco-Sacristán, Fuertes-Oliveira, & Samaniego-Fernández, 2005; Velasco-Sacristán, 2010; Tabacaru & Feyaerts, 2016; etc.).

One of the most influential books on metaphor from a cognitive linguistic perspective is Lakoff and Johnson's *Metaphors we live by* (1980), in which they establish that metaphor is a figure of thought, and not of speech. Their Conceptual Metaphor Theory is mainly concerned with conventional metaphors, i.e. metaphors that have become entrenched among a community of language users (Croft & Cruse, 2004).

Metaphor in cognitive linguistics is the ability to understand one concept in terms of another —A is B— (Lakoff & Johnson, 1980; Forceville, 2009). It involves a relationship between the source domain, i.e. the literal meaning of the metaphorical expression, and the target domain, i.e. what is going to be conceptualised by means of such metaphor. Conceptualisation occurs via the mapping of features from the source domain —usually more concrete and easier

to grasp— and the target domain —usually more abstract and difficult to comprehend—.

Let's consider the metaphor LOVE IS A JOURNEY (Lakoff & Johnson, 1980), in which there is a mapping of features from the source domain, i.e. JOURNEY into the target domain, LOVE, which allows us to construe the latter in terms of the former. That means that lovers have a common goal, like travellers going to the same destination; progress in the relationship is equated to progress along the path; there may be obstacles in their relationship, in the fashion of obstacles against motion, etc. (Ruiz de Mendoza & Galera, 2014, pp. 1-2). The mapping between source and target domains include two types of correspondences: ontological and epistemic. Ontological correspondence is established between elements in the respective domains, e.g. destination/goals in the LOVE IS A JOURNEY metaphor, whereas epistemic mapping refers to the relations between elements, e.g.: going in the same direction is good for the travellers / having the same goals is good for the lovers (Croft & Cruse, 2004).

Recent studies distinguish between several types of metaphor, whose different dimensions, such as metaphor in language, in thought or in communication, can be analysed from various approaches. (Steen, 2011; Kaal, 2012). Furthermore, the study of metaphor as it is found in natural discourse environments is receiving increasing attention (Marhula & Rosinski, 2014). One such perspective of analysis is the MIPVU method (Steen et al., 2010), which focus on identifying metaphor-related words in discourse in order to confront their basic and context meaning. Given that MIPVU analyses are mainly concerned with lexical units, not exploring underlying mapping correspondences, this

procedure has been discarded for the present study, as my thesis is mainly interested in conceptual metaphor as a construal mechanism.

4.2.5.1 Multimodal metaphor

Forceville (2009, 2015, 2016) and Cienki (2008, 2013a) argued that metaphors are not exclusive to verbal language, and they can also occur in other modalities, such as gestures. Indeed, metaphoric gestures are one of the categories established by McNeill (1992), defined as gestures whose pictorial content represent an abstract idea. In mixed metaphors, a target domain is understood in terms of two or more source domains (Forceville, 2016). Cienki (2008) provided various examples of multimodal metaphors involving gestures and speech. One such example (p.14) reports a student talking about honesty being a clear-cut concept, in which you are right or wrong (“black and white”, with no grey colours), while making a “chopping gesture” with her hands, demarcating two distinct spaces. This example is interesting because two different source domains are elicited through speech and gestures in order to convey “the stark difference between two kinds of moral behaviour” (ibid.): whereas a colour metaphor is used verbally —black and white—, a spatial metaphor is used in gesture, dividing space into two clearly distinct areas.

As with conceptual metaphors, gestural metaphors can be more or less conventionalised, salient or highlighted in conversation. However, metaphors expressed verbally or by means of gestures do not necessarily coincide, and each modality expresses metaphors in a different way (McNeil 1992; Cienki, 2008, 2013; Cienki & Muller, 2014). This claim is further supported by Forceville

(2016, p.1), who explained it due to the affordances and constraints of each modality, and also to the fact that multimodal discourse has no grammar, albeit it does have structure; therefore multimodal and purely verbal metaphors behave differently.

4.3 Current Discourse Space

Langacker (2001) linked cognitive grammar to discourse claiming that all linguistic units result from the abstraction of usage events, i.e. actual examples of language use. Fig. 2 illustrates the different elements comprised in a usage event:

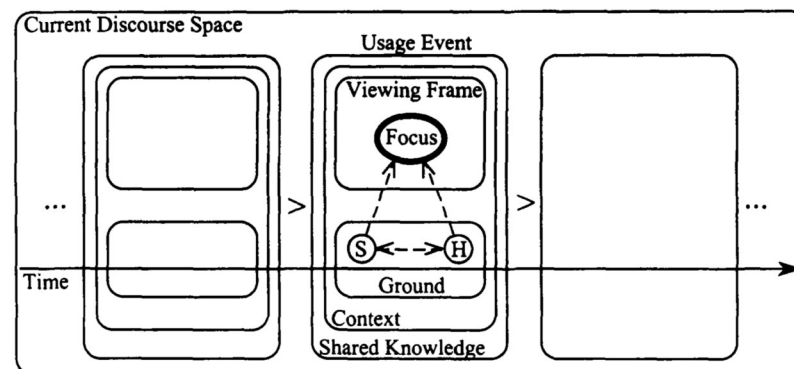


Figure 1.

Figure 2. Representation of a usage event (Langacker, 2001 p. 145)

The ground is the communicative situation itself, including the speaker, the hearer, the time and place of the event, etc. The context of a usage event refers to the immediate context of speech, including “the physical, mental, social and cultural circumstances” (Langacker, 2001, p.145), whereas the shared

knowledge is made of background and encyclopaedic knowledge shared by interlocutors beyond the immediate context. The viewing or discourse frame is the scope of representation and conceptualisation interlocutors have at any given time, which is necessarily limited. Within the activated viewing frame, a certain focus of attention is profiled against the background by multimodal means (verbal, gesture, intonation, gaze, etc.).

The CDS is defined as “the mental space comprising those elements and relations construed as being shared by the speaker and hearer as a basis for communication at a given moment in the flow of discourse” (Langacker, 2001, p. 144). The CDS is therefore the overarching framework in which communication is taking place, including subsequent usage events —and obviously all their components therein—. The CDS evolves as a result of and is transformed by the linguistic structures in usage events, where linguistic must be understood as not merely verbal, but multimodal, comprising gestures, intonation, etc.

To some extent, Langacker acknowledged multimodal inputs as linguistic units in his description of usage events as bipolar events, with a conceptualisation pole and a vocalisation pole. Each pole is enriched by multiple coordinating channels. The conceptualisation pole comprises three channels: information structure, speech management (basically turn taking) and the objective situation (the most prominent channel). In turn, the vocalisation pole would be fed through gestures, intonation, and segmental content. If we refer to these three channels as modes, the clear link with multimodality already present in Langacker’s work becomes explicit.

Compared to Clark's (1996) layering model, advocated by Brône (2008), Langacker's ground would be equivalent to Clark's layer 1. I claim that Langacker's CDS comprises all inputs which can potentially be elicited to produce or interpret meaning in a communicative situation. The CDS dynamically changes as usage events are being realised. Langacker (2001) considers CDS and ground to be cognitive domains that can be potentially tapped into for the meaning of linguistic elements. Domains can also be drawn upon to form mental spaces. Blends and construal operations involved in humorous communication therefore take place against the backdrop of CDS.

Humour would stem from sudden incongruous changes in the CDS achieved by means of construal mechanisms subverted for humorous purposes (Veale et al., 2006). Furthermore, I posit that both the semantic and pragmatic approaches to humour converge under this perspective, as conceptualisation—therefore dynamic meaning—is rooted in discourse, thus acknowledging both the meaning-making production and context-related interactional facets.

4.4 Cognitive linguistics and semiotics

Semiotics is “a system of thought which explicitly seeks to mediate between the natural environment and its perception in consciousness” (Daylight, 2012 p.1). Furthermore, if we consider Bezemer et al.'s (2012) definition of multimodal social semiotics (cf. Chapter 3), there is a clear link between sign production rooted in the organisation of social resources—which Bezemer et al. (ibid.) take as the object of study for social semiotics—and conceptualisation of our experience of the world as a meaning-making process grounded in discourse, as advocated by cognitive linguistics (Langacker, 2001). Social interaction is a

key element of discourse. The entrenchment of discursive patterns, stemming from recurring usage events, including social interaction patterns, form the basis of shared social resources. Let's consider the following example (Bezemer et al., 2012 p. 9):

[In an operating theatre], teacher and learner use a range of communicative resources to manage the positioning of retractors. In Example 1 the trainer uses hand gestures to signify to the trainee where and how to hold the retractor. Her verbal comment ("Just hold it *there*") is only meaningful in relation to those hand gestures. In Example 2 her use of speech ("just slide that one in laterally") becomes more precise, describing a movement and a direction, whilst a pointing gesture indicates which of the two retractors to use for making that movement. In Example 3 speech is not used at all. Here the trainee signals to the trainer, through self-initiated positioning of the retractor, that he feels he is capable of doing so. This is then followed by an evaluation of the trainer ("Yeah, brilliant"; "yeah perfect").

From a multimodal perspective, the interaction between gestures and speech in this communicative event is clear, with pointing and direction gestures complementing the meaning elicited by words. At the same time, the multimodal social semiotic analysis performed by Bezemer et al.'s views changes occurred during the usage event, where control gradually moves from trainer to trainee, as a sign of learning (p.10). Finally, from a cognitive perspective, there is a construal

of the situation in which the trainee adopts the trainer's viewpoint and which is framed by a teacher-learner domain governing the interaction between them.

4.5 Cognitive semiotics and multimodality

Brandt and Brandt (2005) argued that communication (interpersonal or intersubjective) is essentially semiotic, and that expressive blends should be understood as signs and fall under the study of cognitive semiotics, which is concerned with “the study of cognition in semiosis” (p. 216). In order to account for blends from a semiotic perspective, they argued that input spaces blend in a semiotic space, where the “very act of signifying” (ibid. p.225) occurs. They distanced themselves from Fauconnier's discourse base space, as they consider it to be ontological, rather than semiotic. They argued that it represents the speaker's view of reality. They likened their semiotic space to Langacker's ground (Langacker, 2001). Giora (2002), in turn, linked this semiotic space to Langacker's notion of current discourse space.

According to the description that followed in Brandt and Brandt (2005), though, their semiotic space could rather be equated to Langacker's usage event. They pictured the semiotic space as three concentric spheres, one pertaining to the expressive act itself, a larger sphere corresponding to the apprehension by participants of the circumstances of the communicative situation and the biggest sphere, comprising the other two, related to what they call the pheno-world—defined as “conditions that are universally given in the human phenomenological life-world”— (p. 226). I argue that Brandt and Brandt's spheres correspond to ground, context, and shared knowledge in Langacker's (2001) usage event. In my view, the construct of current discourse space is a large

notion: a fluctuating and evolving mental space elicited by the overarching discourse situation, of which usage events are specific instances at a given time. Brandt and Brandt's semiotic space, on the contrary, is not linked to the discourse event beyond the inner sphere of the speech event itself, which I see as analogous to Langacker's ground.

One obvious difference, though, is that while Langacker (2001) placed his usage event in a bigger current discourse space, shared by interlocutors, Brandt and Brandt took the semiotic space as inherent to any form of communication, which can involve an interlocutor or just interaction with oneself in the form of inner thought. To me this distinction is a matter of perspective, rather than of substance. Both accounts share the fundamental claim that meaning-making processes —Brandt and Brandt's (2005) signification and Langacker's (2001) conceptualisation— are rooted in discourse and in our experience of the world.

Brandt and Brandt represented blending networks as examples of semiosis in the Peircean tradition. They pointed to the following mental spaces (ibid.):

- a) Presentation Space, including the representamen (the signifying sign).
- b) Reference Space, including the object to be signified.
- c) Interpretant Space, showing how the sign must be interpreted.
- d) A blended space with the meaning resulting from combining the three previous inputs.

Their basis for rejecting Fauconnier and Turner's (1998, 2002, 2003) model is that, in their view, it does not account for inferences. Hence, they

claimed that more spaces are needed in the blending network. One is the semiotic space, linking the blending process to the context in which the act of signifying occurs. Another is the relevance space within the semiotic space allowing for the retrieval of inferences. They therefore claimed that their model of blending network is more comprehensive.

However, I reject this claim for two reasons: first, as they themselves acknowledge in their paper, not all sign relations are relations between mental spaces, and thus they cannot be represented in terms of presentation space, reference space, etc. Brandt and Brandt (2005) claimed that the overall structure of the network is nevertheless the same, but they offer no counterexample to illustrate it. If, as they claimed in the article, communication is essentially semiotic in nature, it necessarily involves signs, so a question remains to be answered: How those signs would be motivated if not by means of conceptualisation through mental spaces?

Secondly, and more fundamentally, in my view, they argued that this cognitive semiotic model of blending networks relies on intentional sign relations. They further posited the following as a major tenet of Cognitive Semiotics (p. 242):

The blending network is a cognitive remodelling of semiosis in general, as understood in the Peircean tradition, but only in the sense that a Presentation of a Reference is necessarily an intentional sign relation, linking a Representamen to an Object, and that the stabilization of such a relation through an Interpretant

corresponds strictly to the semiotic function of Relevance-making in the blending network.

In my view, to include only intentional communication implies that many instances of communication are excluded. The cognitive semiotics approach advocated by Brandt and Brandt proves to be too restrictive for a comprehensive analysis of communication. A broader view of cognitive semiotics, though, is advocated by Zlatev (2012), who defined it as “an emergent field dedicated to the transdisciplinary study of meaning” (p. 2), whose aim is to provide new insight into the process of human signification drawing from various disciplines and specially merging cognitive science and semiotics. Zlatev defended that the study of gestures is best analysed from this cognitive semiotics perspective, as he believes has been more or less explicitly done by Kendon (2002, 2004), McNeil (1985, 1992, 2000, 2014) and Müller and Cienki, 2009 (see also Cienki & Müller, 2008).

4.6 Conclusion

A brief introduction of the discipline is offered (Croft & Cruse, 2004; Evans & Green, 2006; Geraeerts, 2008), along with an overview of major assumptions within this approach, placing special emphasis on the experiential bases for knowledge advocated by cognitive linguists (Johnson, 1987; Lakoff, 1987; Talmy, 2000 a, 2000b; Langacker, 2001; etc.), as what I think is the obvious link between cognitive linguistics and multimodality. If our thoughts and construal of the world are shaped through our perception and bodily experience, then a cognitive-based multimodal study of spontaneous communication focusing on the interplay

between the verbal, prosodic and co-speech gestures channels (or modes) is warranted.

Paramount concepts in cognitive linguistics, such as domains (Langacker, 1987), construal operations (Croft & Cruse, 2004), image schemas (Lakoff, 1987; Johnson, 1987), mental spaces (Fauconnier, 1994), and blending (Fauconnier & Turner, 2002) are reviewed, as they will later be drawn upon to conduct a cognitive analysis of humorous instances identified in my sample. To that end, I will endorse Clausner and Croft's (1999) claim that image schemas are subsumed as a sub-type of domains, which in turn are overarching frameworks in which mental spaces arise.

Langacker's (2001) notions of current discourse space, along with that of linguistic units as abstractions of usage events, are advocated as the bases of my account of humour in conversation. In line with one of cognitive linguistics major tenets, i.e. that language is not based on an independent cognitive ability different from other cognitive functions, and that linguistic meaning is not separate from other forms of knowledge that we have (Geraeerts, 2008), I endorse Veale et al. (2006) view that humour does not need to be explained by means of humour-specific mechanisms. Instead, cognitive linguistics offers the tools to account for humorous communication as a highly complex phenomenon (Veale et al., 2015) which draws on a non-prototypical use of general construal operations (Brône & Feyaerts, 2003).

Then, the chapter delves into the link between cognitive linguistics and semiotics, and, more specifically, into a critical review of cognitive semiotics, as

defended by Brandt and Brandt (2005). Their model for explaining blending from a cognitive semiotic point of view is analysed and rejected, as in my view it does not offer better explanatory power than other tools such as blending (Fauconnier & Turner, 2002) or Langacker's (2001) current discourse space.

Chapter 5

Material and methods

5.1 Introduction

There is increasing recognition that in order to conduct a proper multimodal analysis of communication, it is better to study spontaneous instances of communication (Attardo, Pickering, & Baker, 2011; Attardo et al., 2013; Tabacaru, 2014). Regarding humour, many analyses focus on staged humour as a proxy for spontaneous communication (Urios-Aparisi & Wagner, 2011; Tabacaru, 2014). However, given that humour is based on familiarity, it could be argued that staged humour resorts to exaggerated or more ostentatious features in order to reach a wider audience (Flamson et al., 2011). Therefore, its validity as a proxy for everyday spontaneous communication could be questioned. Consequently, for the purposes of this dissertation, I have favoured the approach of studying non-scripted, non-rehearsed humorous communication.

Below I present the sample in detail, along with the annotation tools used for this analysis. This chapter ends with a brief overview of the quantitative results obtained for all the parameters studied: humour types, construal mechanisms, gestures, and prosodic features.

5.2 Sample

The sample analysed includes 14 interviews from *The Late Show with Stephen Colbert*. The choice of show has been determined by the fact that, typically, late night shows allow for humorous instances to occur more frequently

than other type of shows. Late-night talk shows belong to a genre which involves political satire, sketch comedy, monologues, and interviews to celebrities in a light and playful tone.

Interviews have been selected as a source of non-scripted speech in which humorous communication is likely to occur. In order to ensure that only non-scripted communication is accounted for, only utterances by interviewees have been analysed, avoiding mostly pre-scripted or rehearsed host's speech. Interviews were selected to have an equal number of men and women and a variety of ethnic and professional backgrounds.

Humorous instances found in interviews vary in length, type and number. Each interview was analysed in a different ELAN file, as explained more in detail in section 5.3.1. Prosodic features for each selected utterance were studied separately in Praat. All interviews are freely available on Youtube. The videos were downloaded with aTube Catcher, a free video software. Table 2 summarises the main features of the 14 interviews under study in this sample.

Interviewee	Gender	Age	Ethnic back.	Interview (s.)	Humorous instances	Humorous instances (s.)	Episode/season date
Alec Baldwin	Male	60	White	432,030	13	43,055	Episode 130 (S2) 18/4/2017
Alison Janney	Female	58	White	445,262	6	29,206	Episode 134 (S2) 24/04/2017
Amy Schumer	Female	36	White	514,597	12	38,495	Episode 140 (S2) 2/05/2017
Condola Rashad	Female	31	Black	389,859	4	17,430	Episode 106 (S2) 2/3/2017
Cristela Alonso	Female	39	Hispanic	364,549	12	44,325	Episode 82 (S2) 20/1/2017
Daniel Kaluuya	Male	29	Black	404,488	13	45,859	Episode 71 (S3) 16/01/2018
Elon Musk	Male	46	White	167,691	4	13,905	Episode 2 (S1) 9/09/2015
Michael Hayden	Male	73	White	448,257	5	14,050	Episode 109 (S2) 7/03/2017
Joseph Biden	Male	75	White	625,797	7	18,495	Episode 55 (S2) 6/12/2016
John McWhorter	Male	52	Black	412,383	7	31,120	Episode 94 (S3) 27/02/2017
Riz Ahmed	Male	35	Asian	543,923	13	39,420	Episode 198 (Season 1) 29/08/2016
Sheryl Crow	Female	56	White	290,501	3	9,535	Episode 132 (S2) 20/4/2017
Sigourney Weaver	Female	68	White	454,829	5	17,870	Episode 126 (S2) 5/04/2017
Susan Sarandon	Female	71	White	676,788	5	11,560	Episode 123 (S2)

Table 2. Interviews in the sample

The sample contains 103.83 minutes of interviews, out of which 109 humorous utterances were found, amounting to 6.24 minutes in total. For each humorous utterance, annotations on five parameters were made: a) transcription of the utterances selected, b) main construal mechanism underlying humour, c) type of humour involved, d) gestures made in the humorous utterances, e) prosodic analysis (pitch and intensity). Additional utterances were selected for

comparison of prosodic features between humorous and non-humorous utterances, as will be explained in more detail below. What follows is an overview of each interview with regards to types of gestures and humour identified.

ALEC BALDWIN

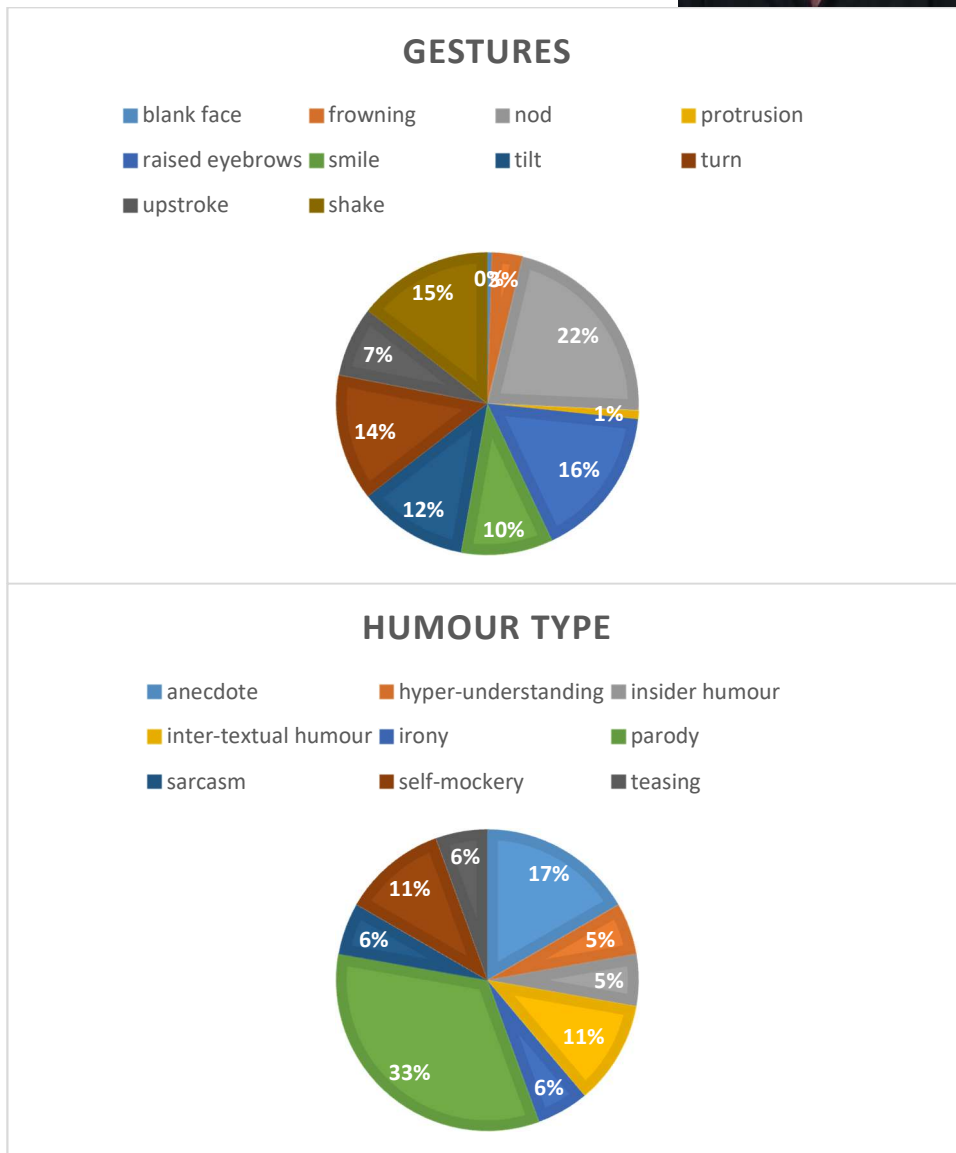


Figure 3. Overview of Alec Baldwin's interview

ALISON JANNEY

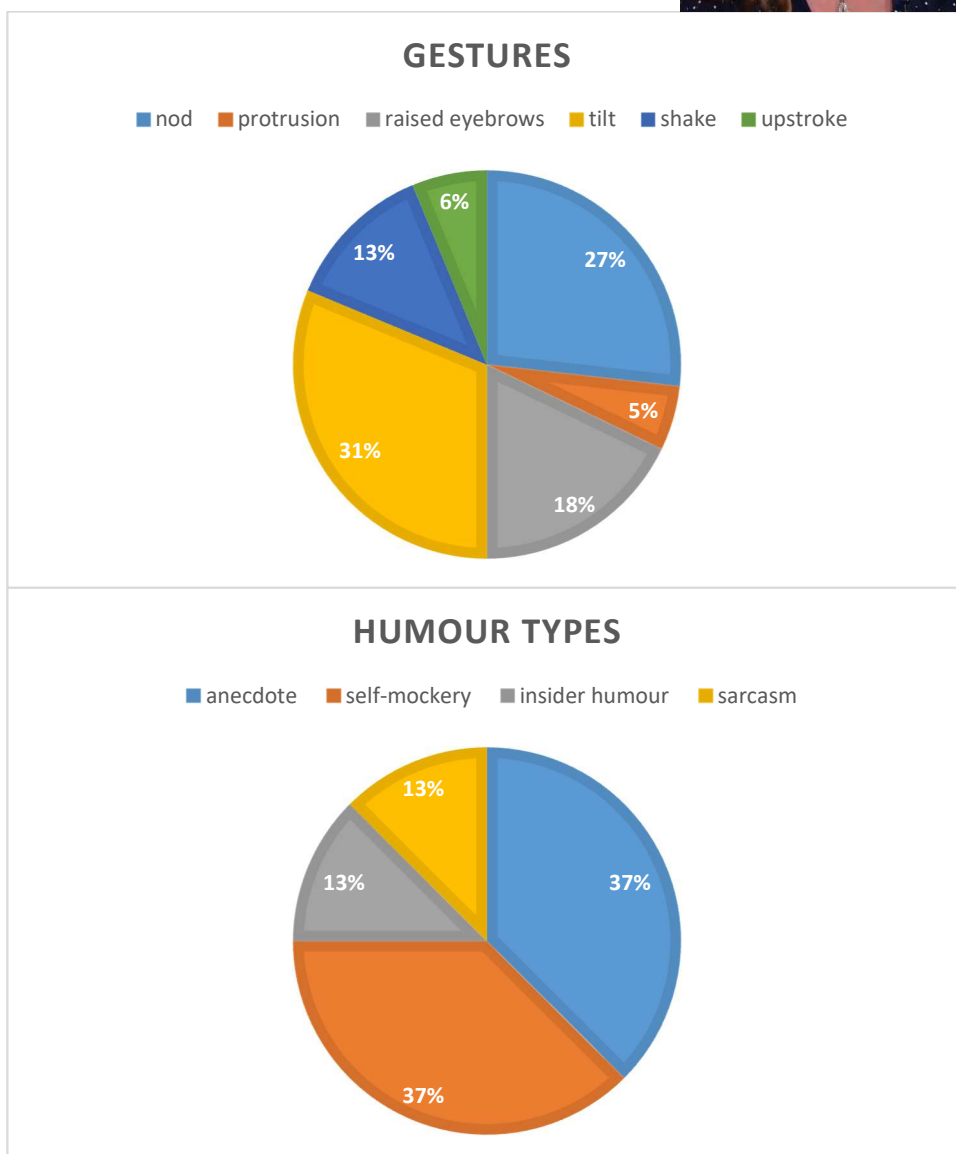


Figure 4. Overview of Alison Janney's interview.

AMY SCHUMER

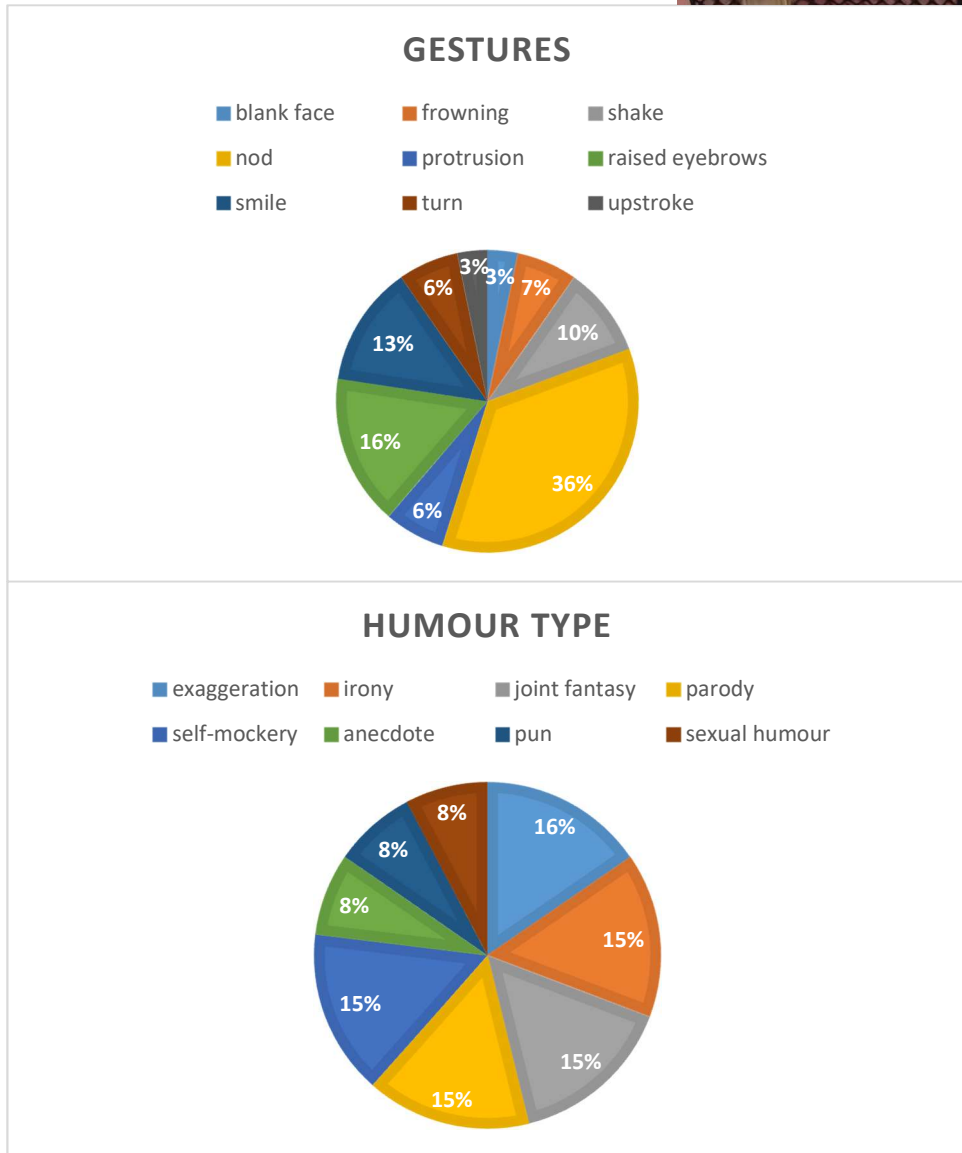


Figure 5. Overview of Amy Schumer's interview.

CONDOLA RASHAD

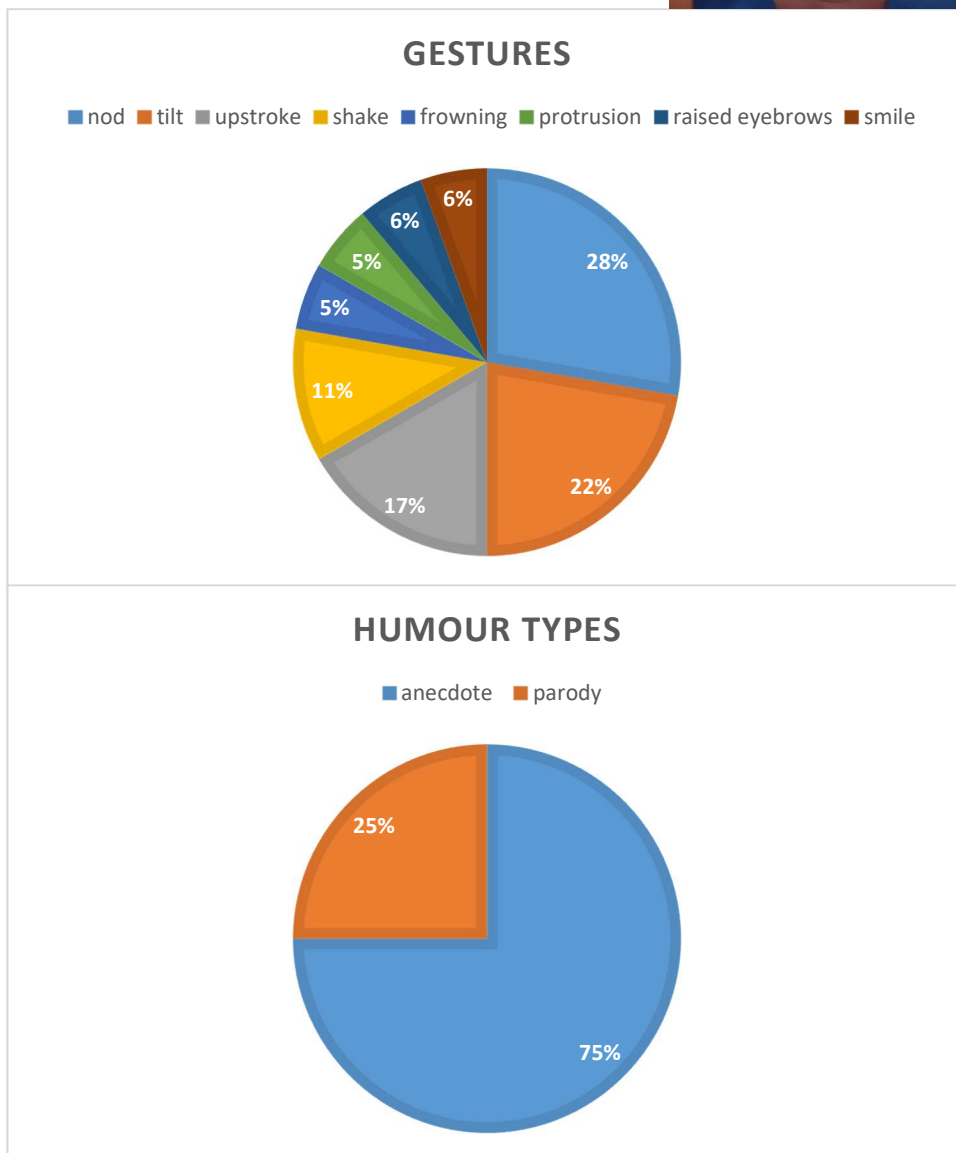


Figure 6. Overview of Condola Rashad's interview.

CRISTELA ALONZO

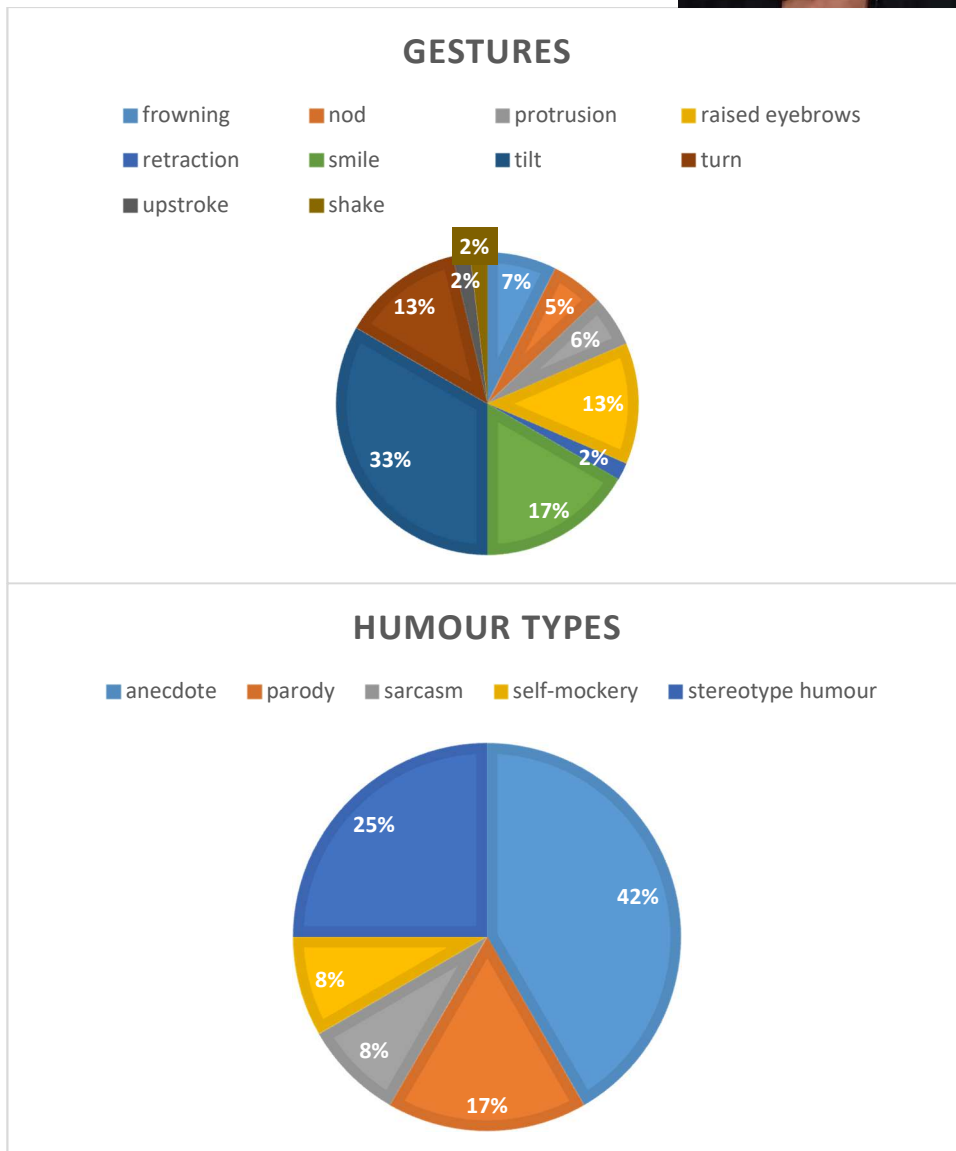


Figure 7. Overview of Cristela Alonzo's interview.

DANIEL KALUUYA

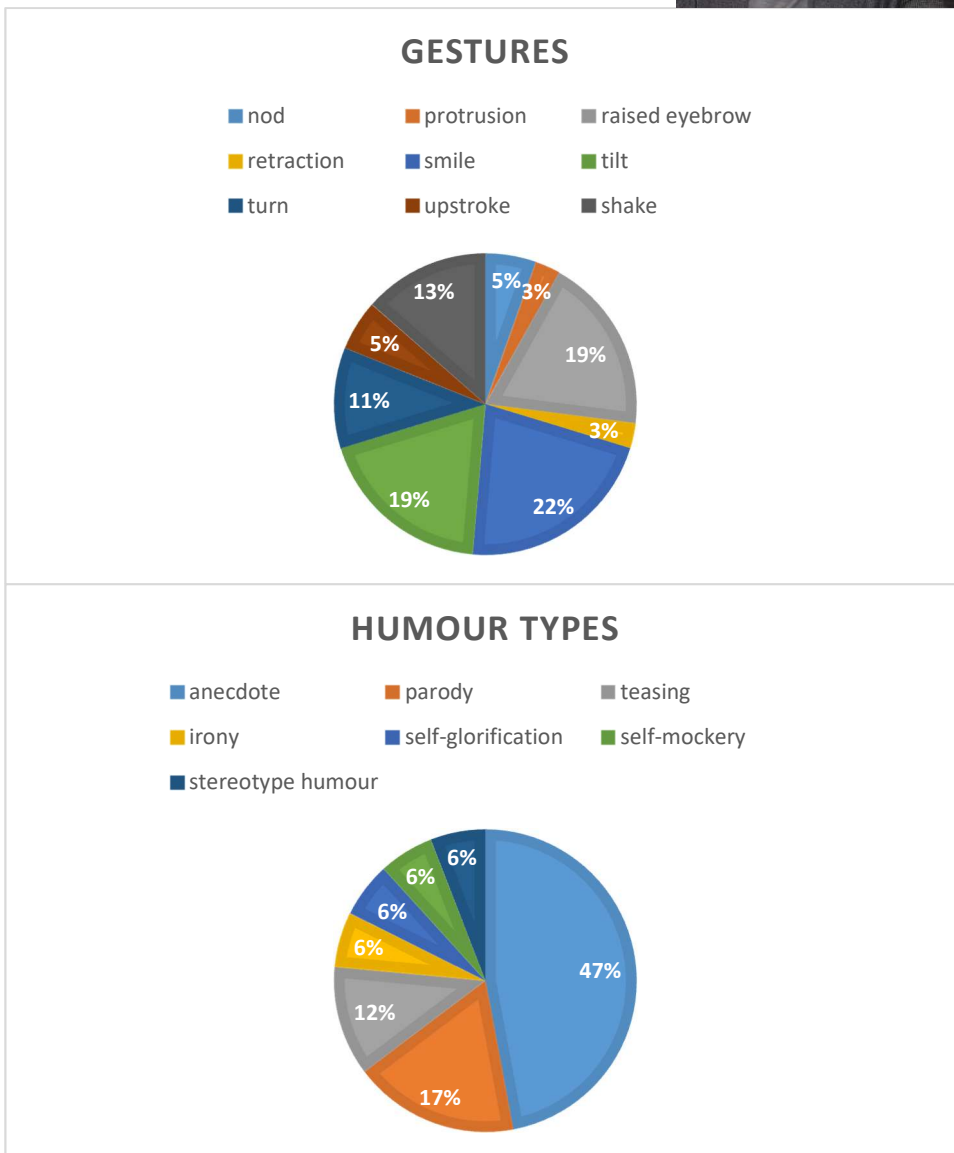


Figure 8. Overview of Daniel Kaluuya’s interview.

ELON MUSK

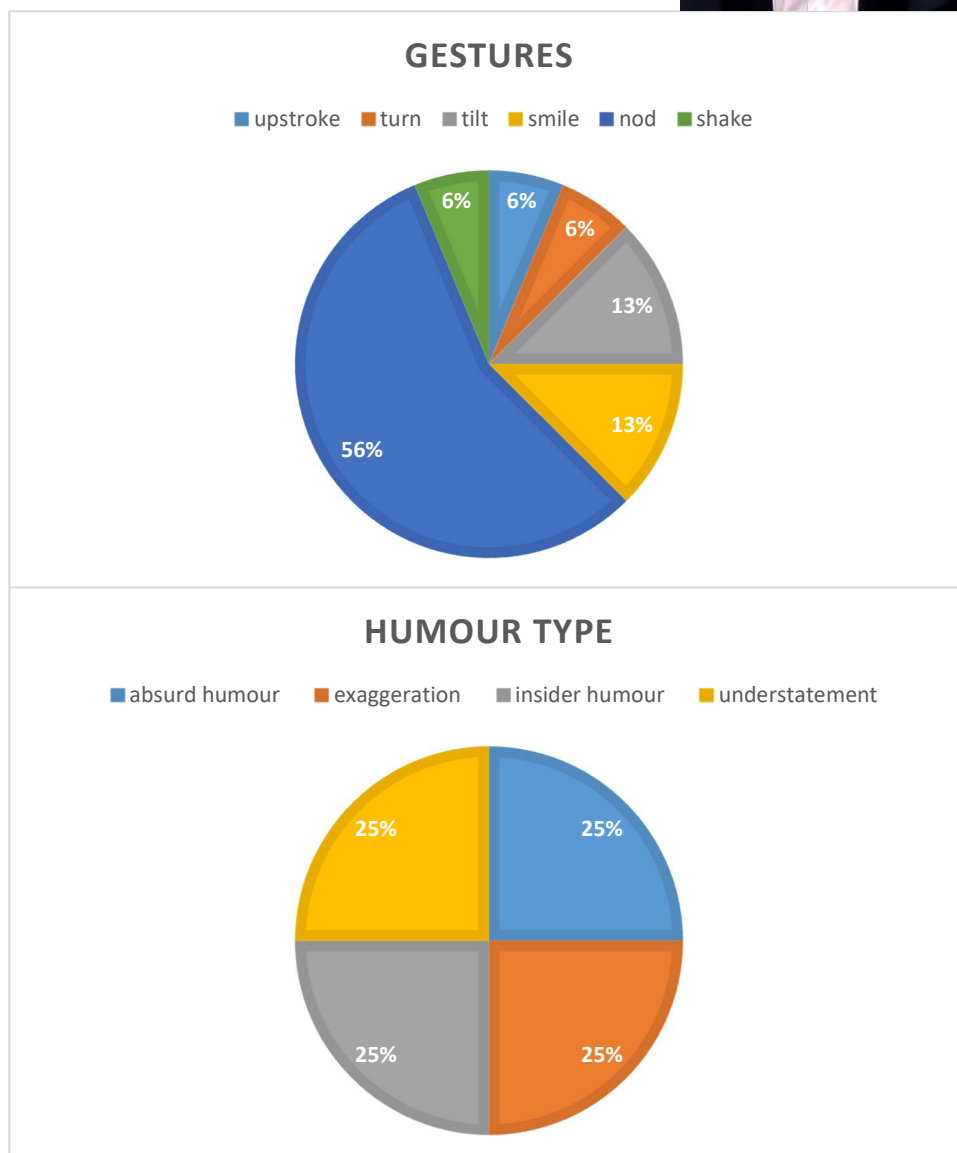


Figure 9. Overview of Elon Musk's interview.

JOSEPH BIDEN

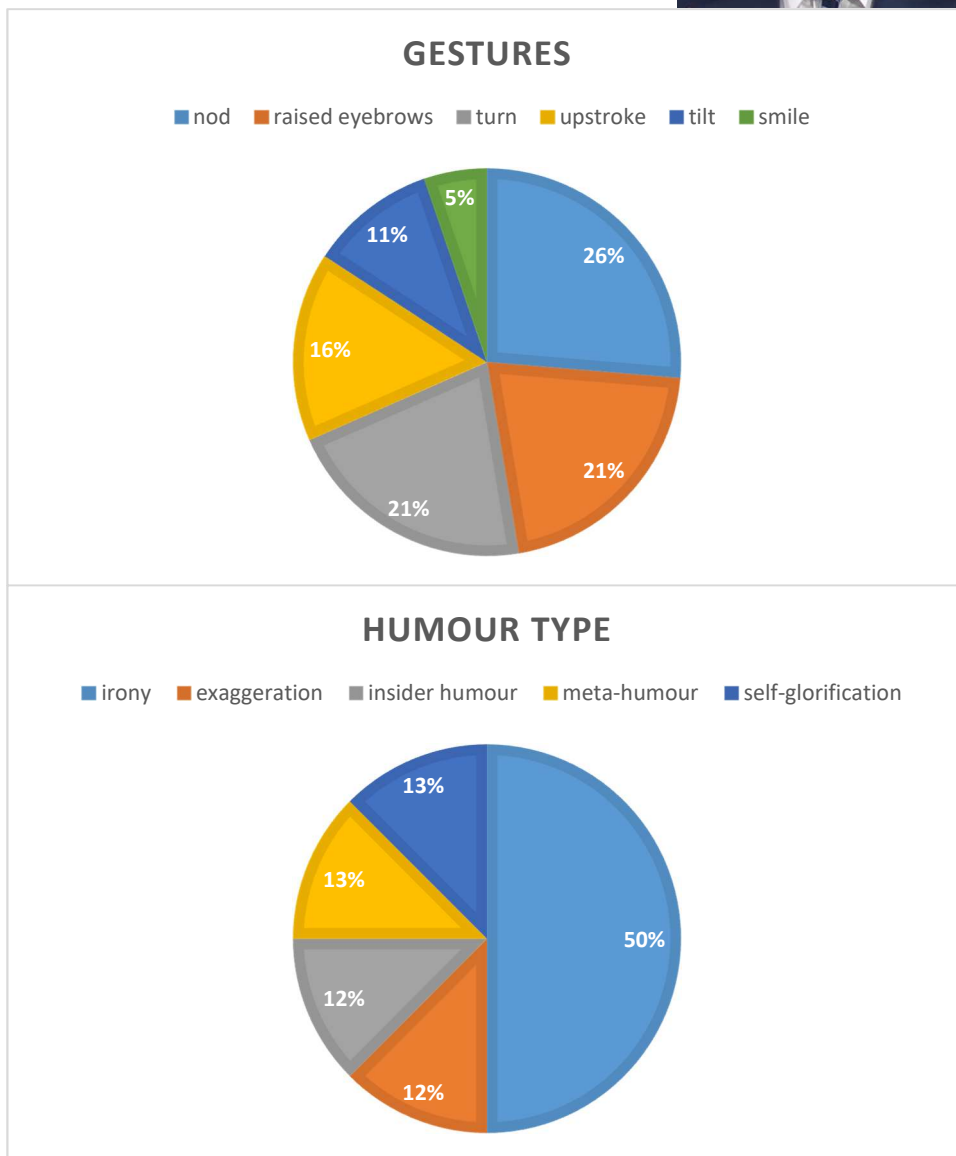


Figure 10. Overview of Joseph Biden’s interview.

JOHN MCWHORTER

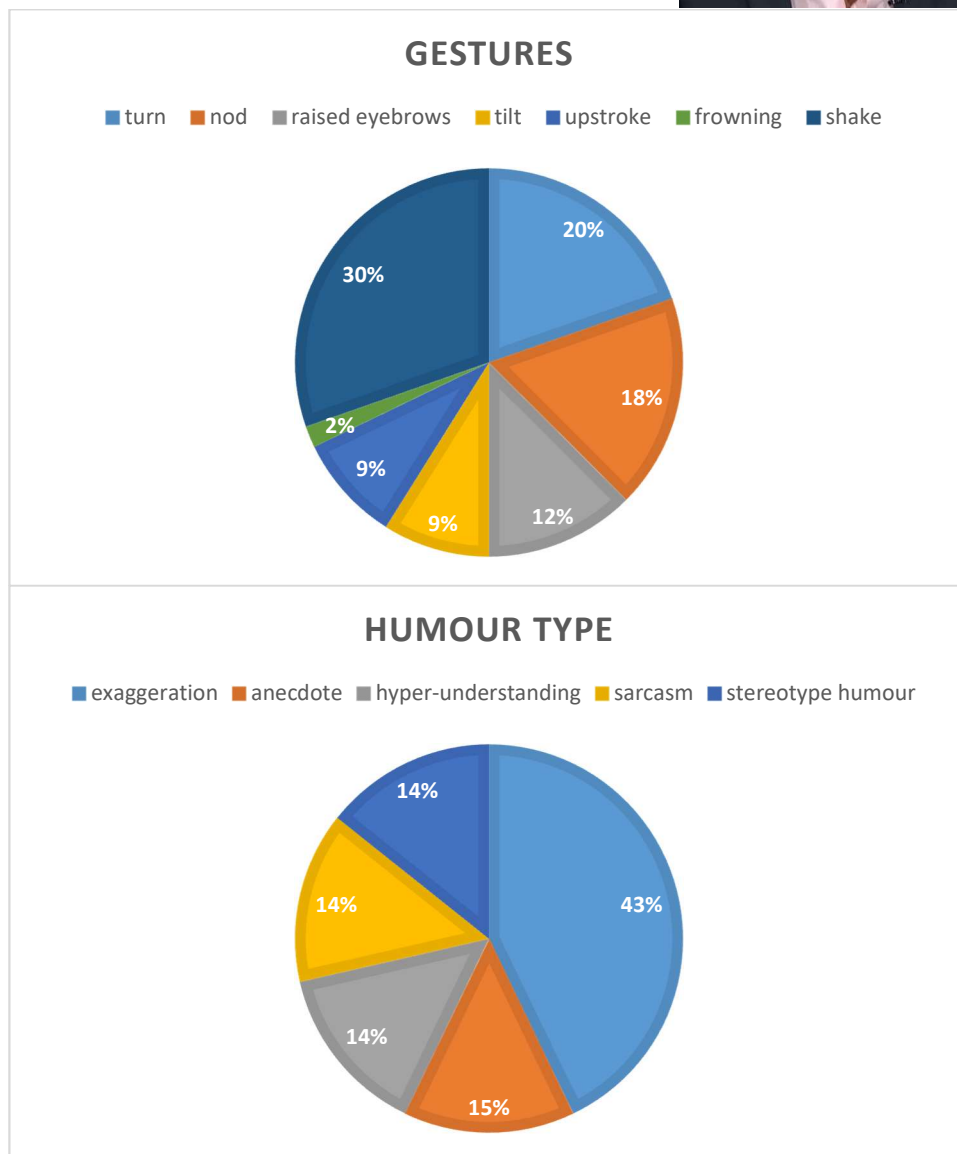


Figure 11. Overview of John McWhorter's interview.

GENERAL MICHAEL HAYDEN

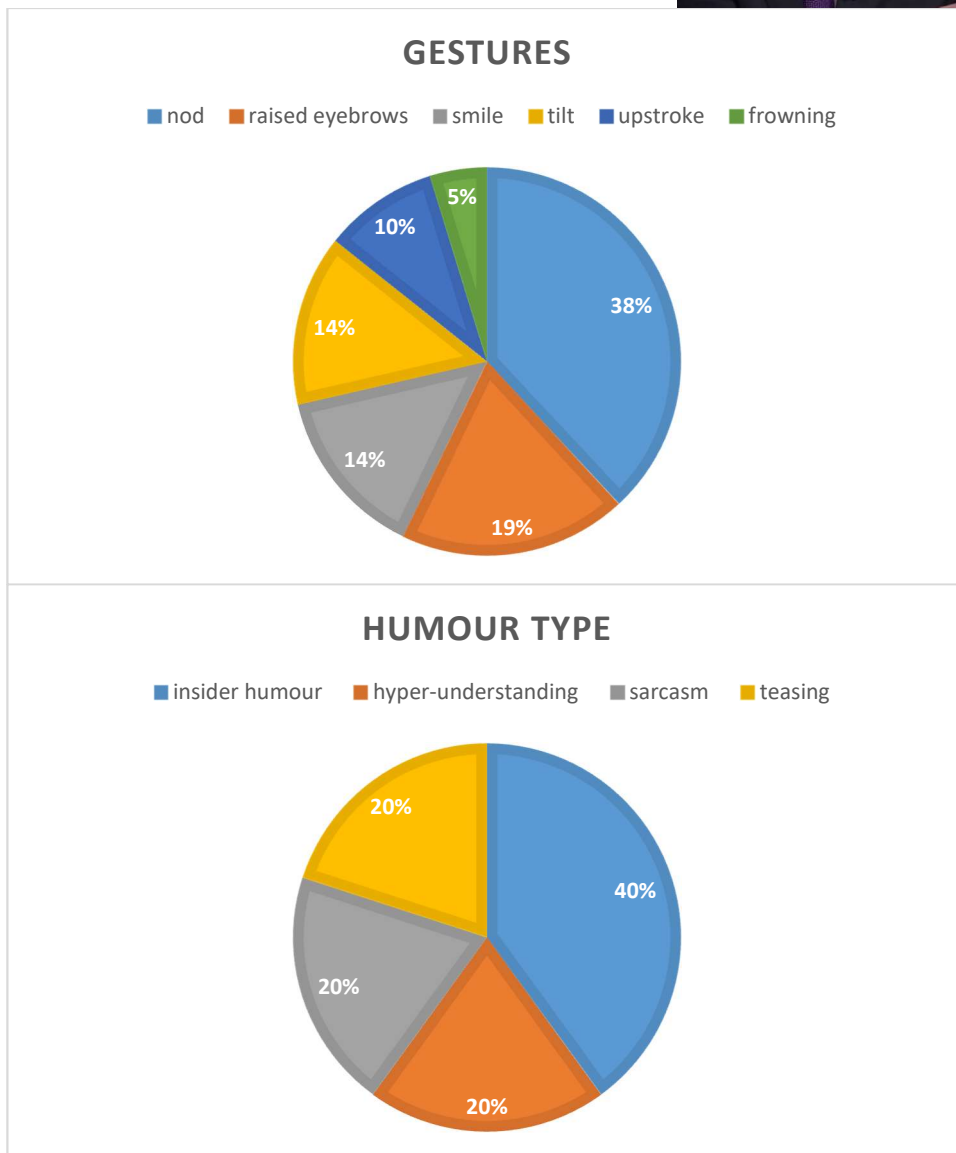


Figure 12. Overview of Michael Hayden interview.

RIZ AHMED

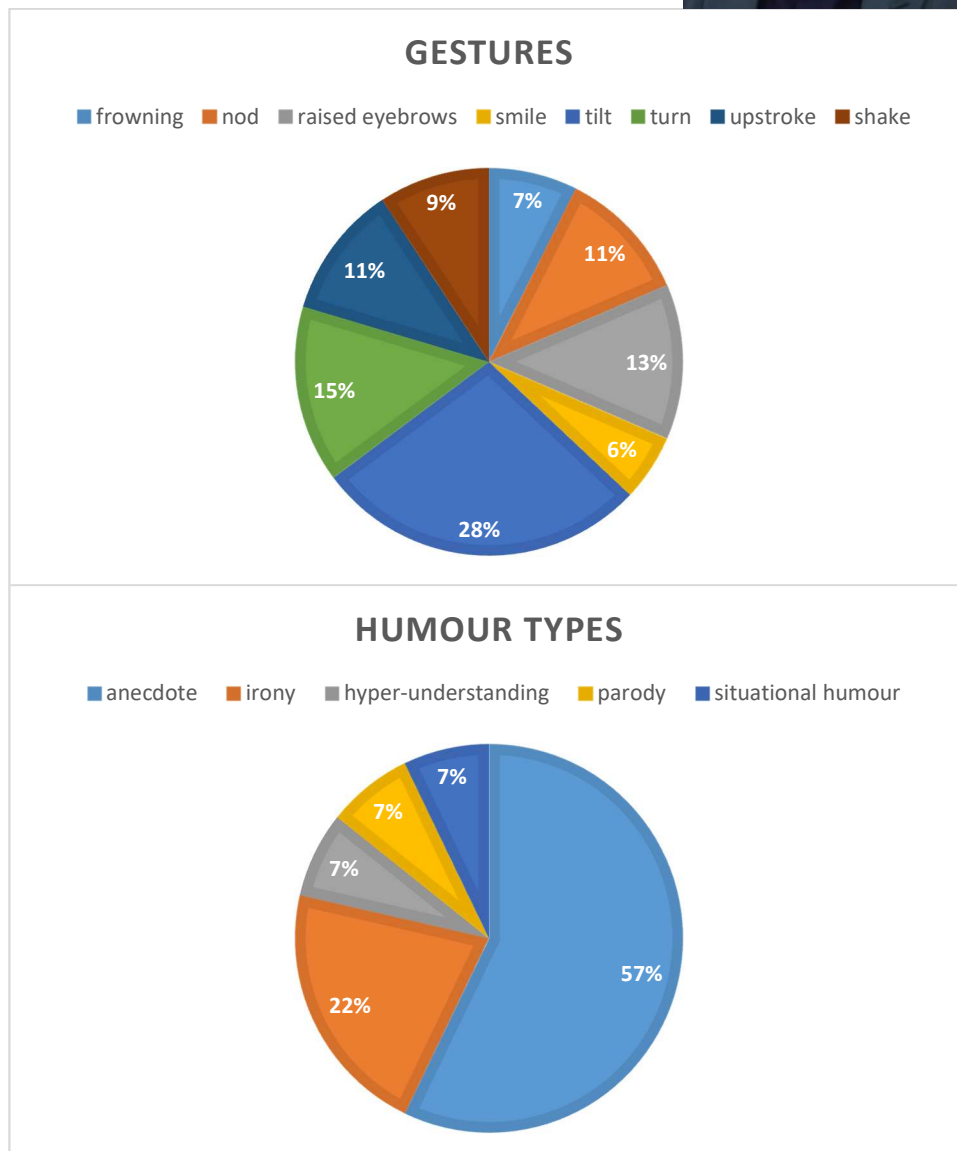


Figure 13. Overview of Riz Ahmed's interview.

SHERYL CROW

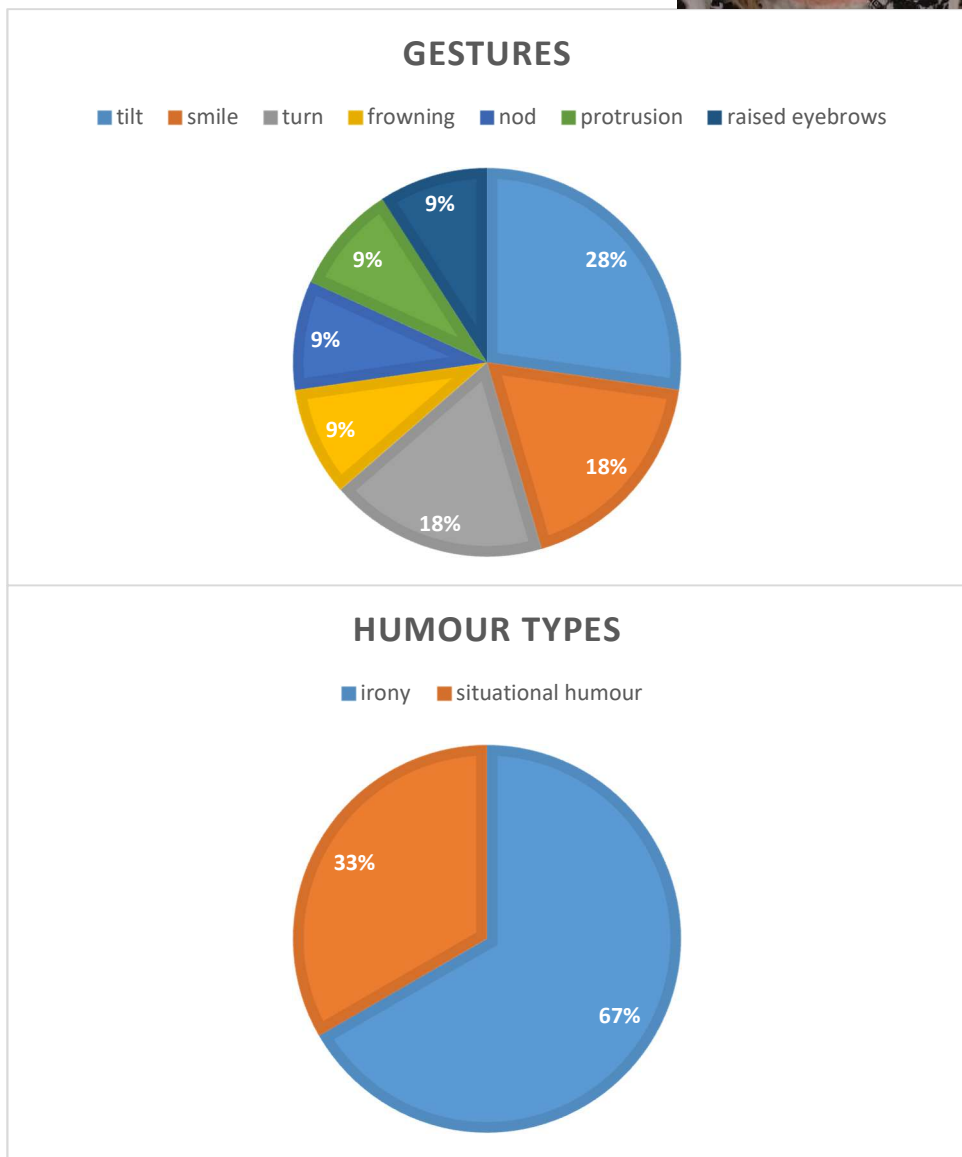


Figure 14. Overview of Sheryl Crow's interview.

SIGOURNEY WEAVER

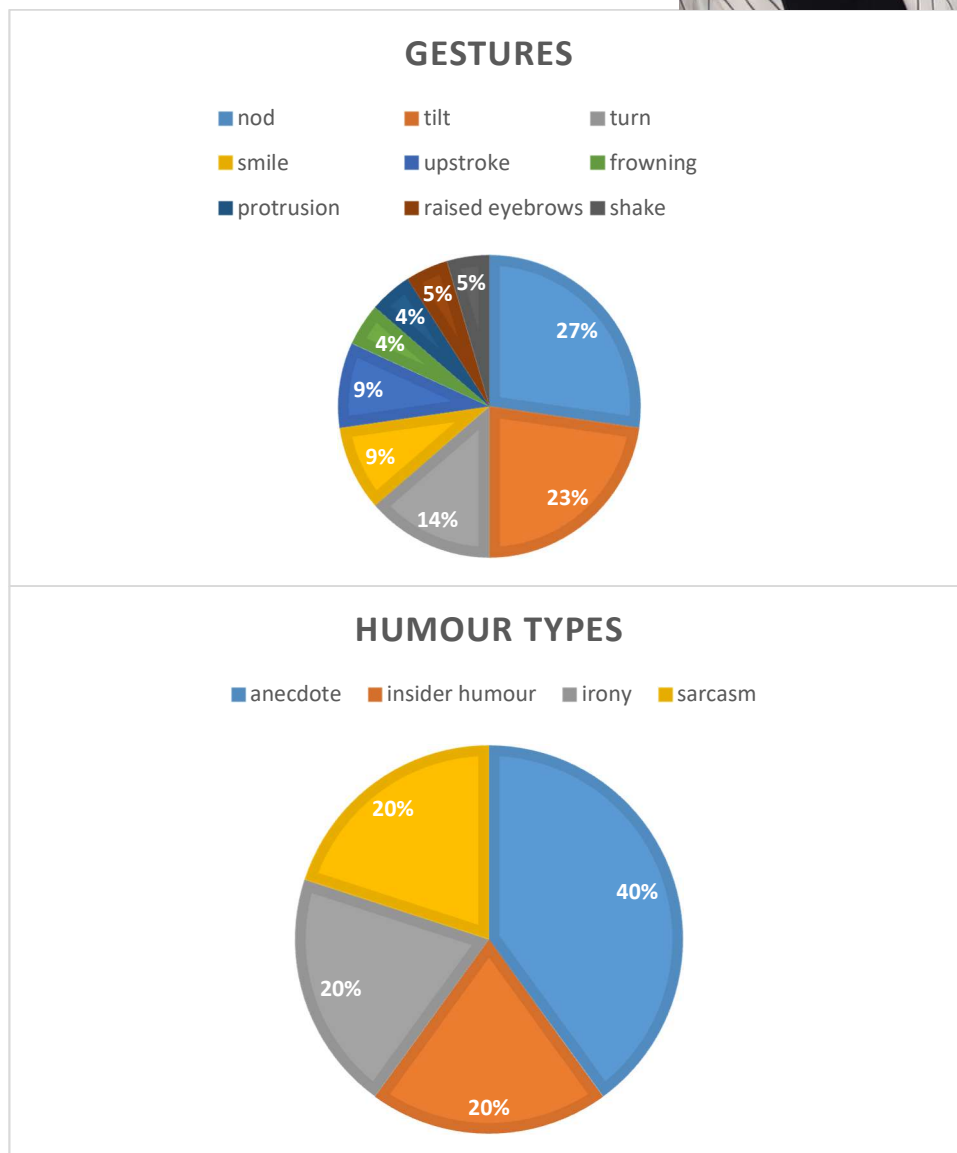


Figure 15. Overview of Sigourney Weaver's interview.

SUSAN SARANDON

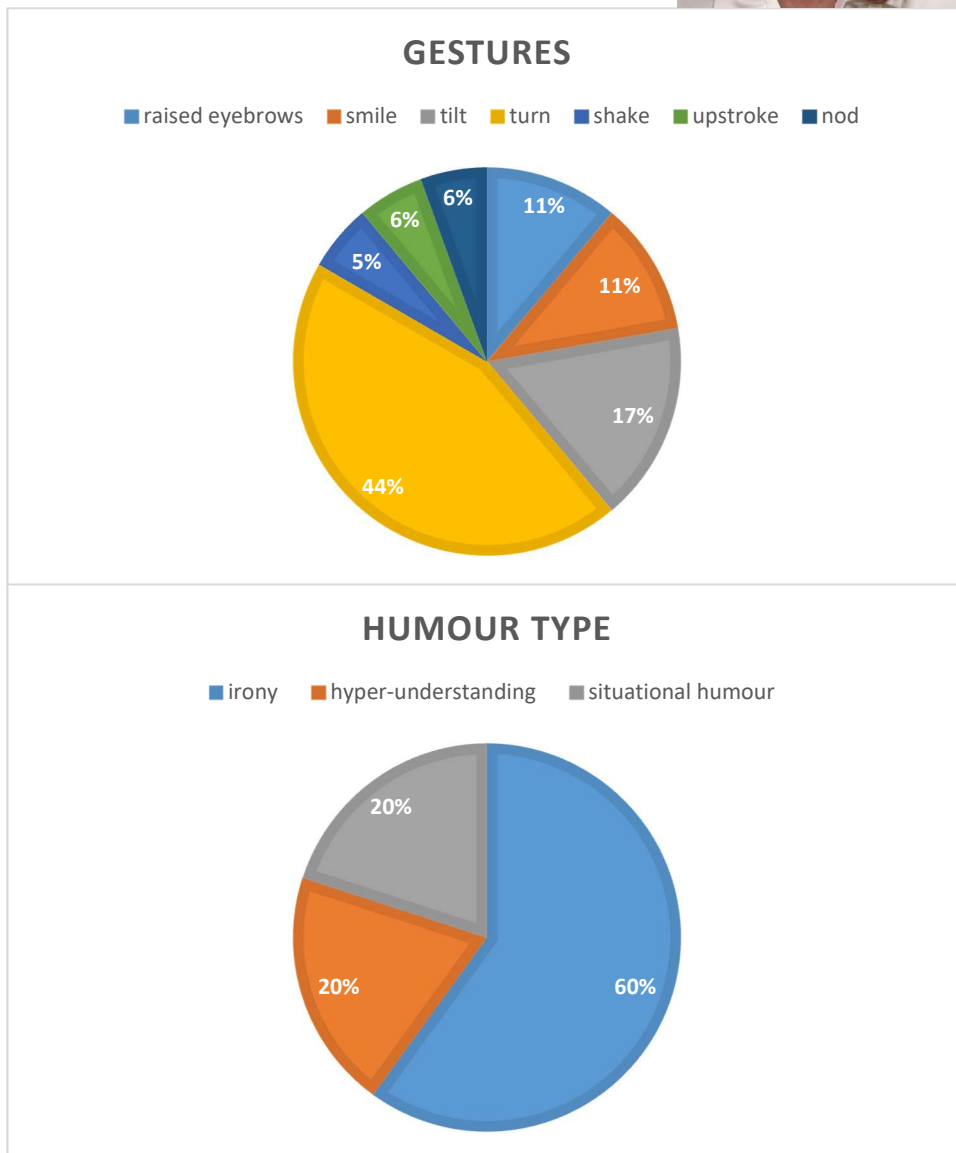


Figure 16. Overview of Susan Sarandon interview.

What follows is a breakdown of gestures and humour types per gender. No differences exist between male and female interviewees. The most frequent gestures and types of humour in both groups are in line with the overall results for the sample.

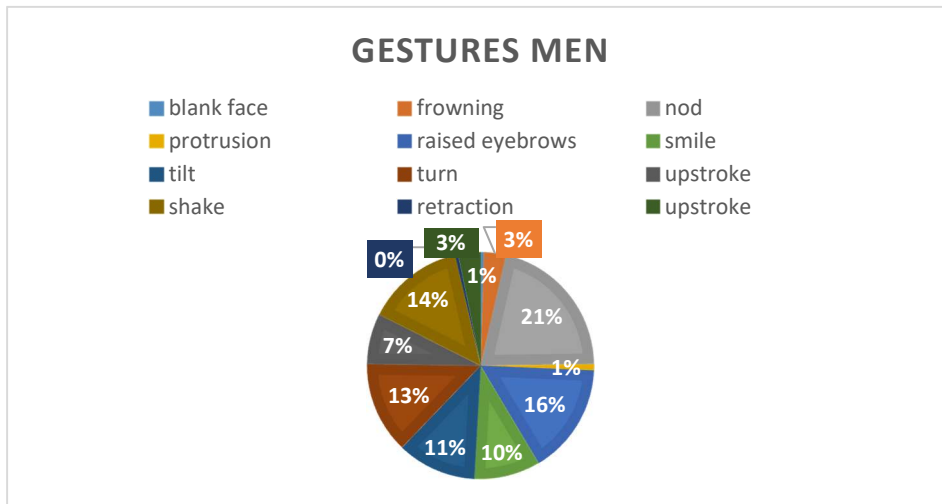


Figure 17. Gestures performed by men.

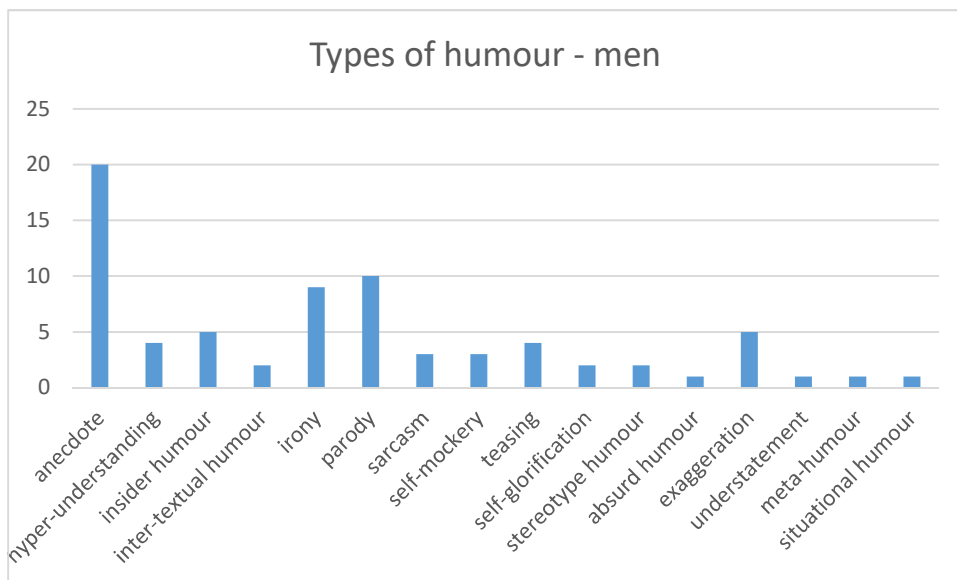


Figure 18. Types of humour (men).

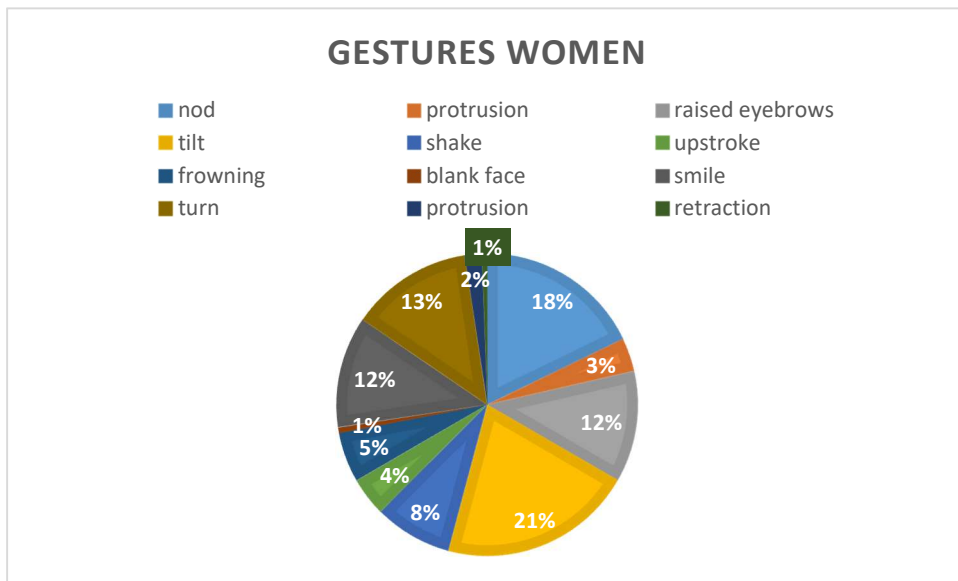


Figure 19. Gestures performed by women.

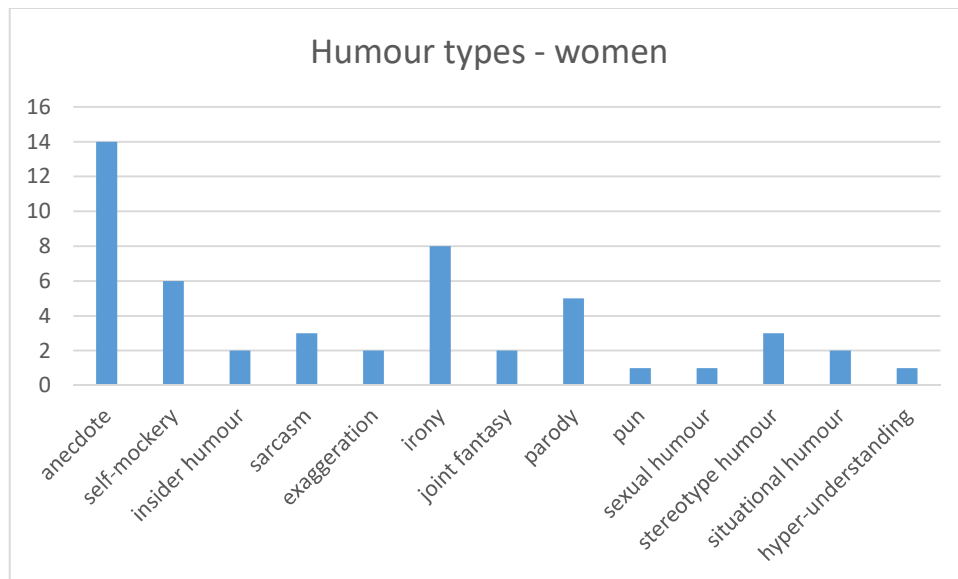


Figure 20. Humour types (women)

5.3 Annotation tools

Two different software applications were used to perform the multimodal analysis: ELAN, version 5.1 (ELAN, 2017) was used to upload and study videos, with annotations on transcription, construal mechanisms, gestures, and humour type. A fifth tier on prosody was added to report the results of the separate prosodic analysis conducted in Praat. Praat, version 6.0.32 (Boersma & Weenink, 2017)⁴ was used for the prosodic analysis, drawing audio files equivalent to the utterances selected in order to obtain measurements for pitch (fundamental frequency: F0), and intensity. Mean intensity and mean F0 values were then included in the prosodic tier in Elan for each utterance identified. Finally, statistical data was processed through SPSS (version 20).

5.3.1 ELAN

Elan is a video annotation tool, allowing for multiple complex annotation on different tiers. The videos were first captured and downloaded in mp4 format from *The Late Show with Stephen Colbert* YouTube channel using aTubeCatcher (version 3.8.9325 DsNet Corp., 2017)⁵. They were subsequently imported to ELAN, creating one file per interview.

ELAN offers multiple possibilities for viewing and manipulating videos. The control menu enables playing the video at different speed, thus facilitating

⁴ Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.32, retrieved 20 September 2017 from <http://www.praat.org/>

⁵ DsNet Corp. (2017). aTubeCatcher Version 3.8.9325, retrieved 1 July 2017 from <http://www.atube.me>

transcription (Fig.21). Precise gesture annotation is possible thanks to the function of moving through the video frame by frame (either backward or forward).

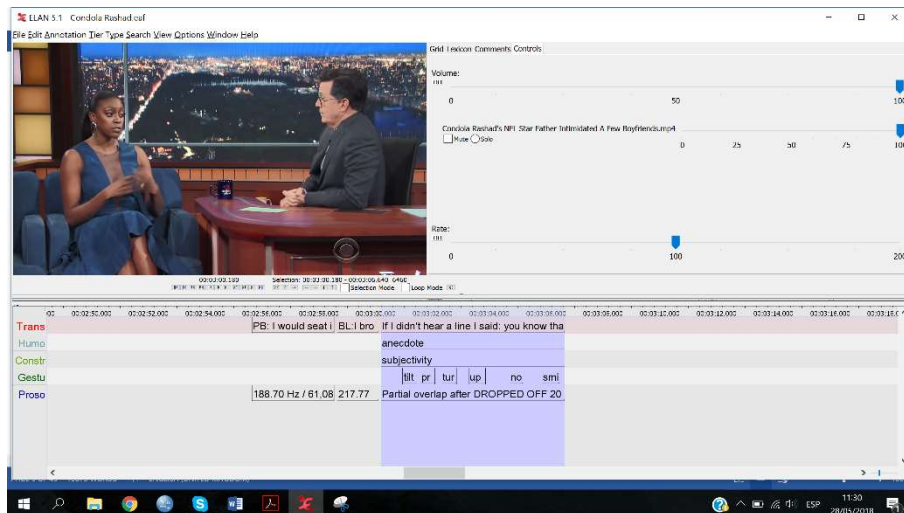


Figure 21. Control window in ELAN

Five tiers were created in ELAN to include the following annotations: transcript, type of humour, type of construal mechanism, gestures, and prosodic analysis Fig. 22). In total, 254 utterances were selected and annotated in ELAN: 109 humorous utterances, and 149 non-humorous utterances, divided into 66 pre-base utterances and 79 baseline utterances (see section 5.3.3 for definitions). Only humorous utterances were annotated on all 5 tiers. Both pre-base and baseline utterances were used for the purposes of prosodic analysis exclusively.

	00:03:18.000	00:03:20.000	00:03:22.000	00:03:24.000	00:03:26.000
Trans		PB: Uh...	BL: It's hard becau	Going to work	
Humo				anecdote	
Constr				viewpoint	
Gestu				til	tilt
Proso		257.44	261.40 Hz / 65.62	210.62 Hz / 61.	

Figure 22. Annotation tiers in Elan

Humorous utterances were selected first, on the basis of laughter elicited in the audience. The segmentation into baseline and pre-base utterances was done after careful selection of humorous utterances. The selection of baseline and/or pre-baseline utterance was not always possible, hence the smaller number of these types of utterances with regards to humorous ones. Once selected, utterances were transcribed on the first tier, before conducting the analysis of the subsequent levels.

ELAN allows for easy navigation through tiers and annotations, as it is possible to select annotations by type resorting to the grid window, where a list of all annotations made on a given tier can be selected to move through different instances in that tier (Fig. 23 and 24).

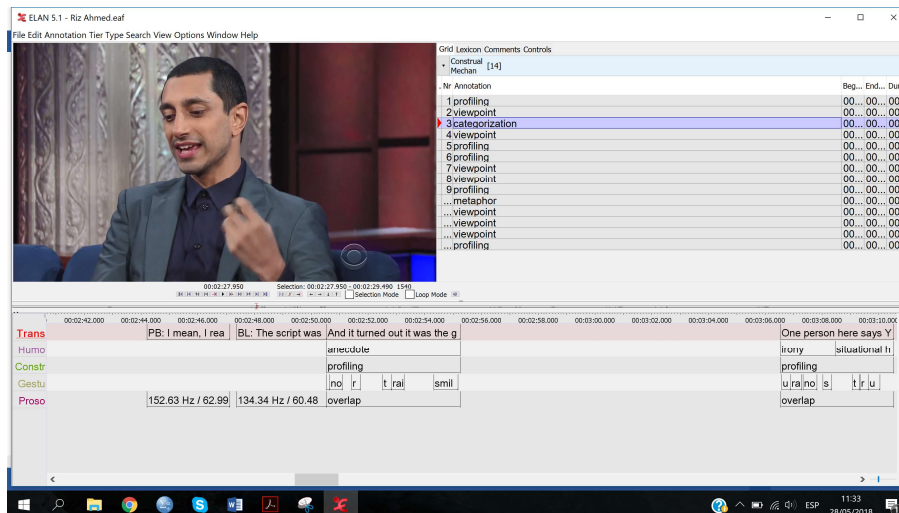


Figure 23. View of annotations by list of construal mechanisms

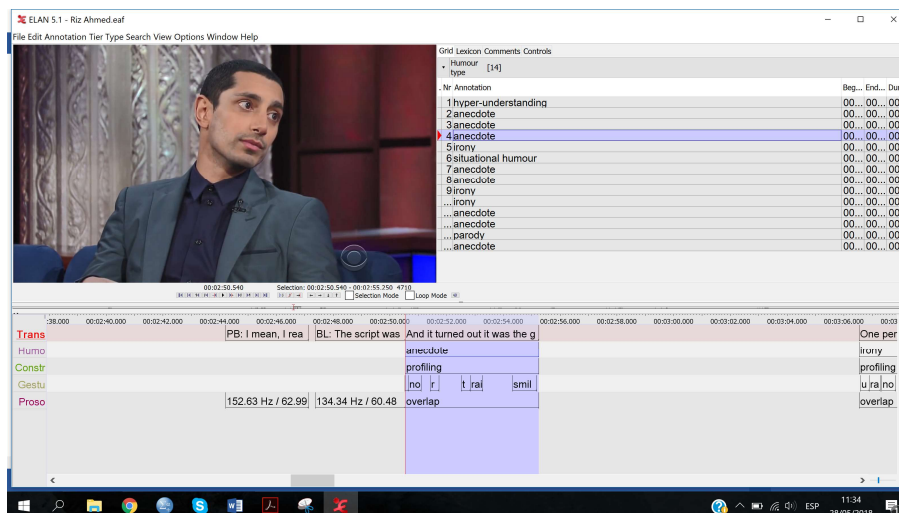


Figure 24. View of annotations by humour type

Upon selection of one annotation from the list displayed, the clip of video linked to that annotation is shown and can be played individually. In addition, the FASTsearch function facilitates searching occurrences of any given word either through one file or group of files, depending on the search domain previously

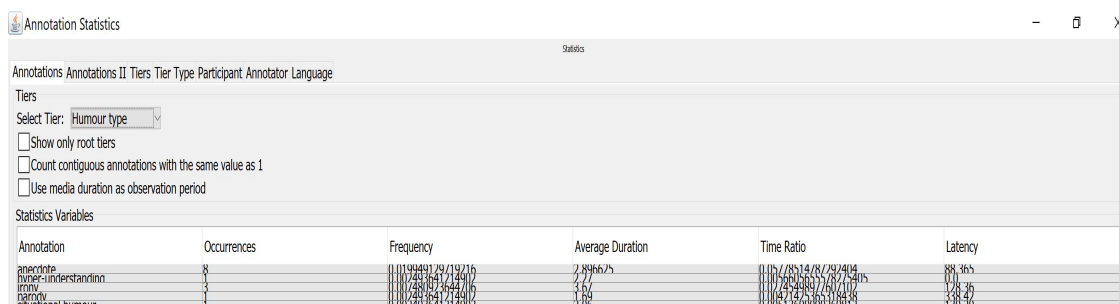
established. Opening the relevant video clip and file is as simple as just clicking on any of the results displayed on the list (Fig. 25).

The screenshot shows the FASTSearch application window. The search term 'frown' is entered in the search box. The search is complete, showing 5 occurrences in 5 annotations across 14 files in 0.039 seconds. The results are displayed in a table with columns for Nr, File, Tier, Before, ..., After, Parent, Child, Beg..., End..., and Dur....

Nr	File	Tier	Before	...	After	Parent	Child	Beg...	End...	Dur...
1	ALEC BALDWIN	Gesture	smile	frown	protrusion			00:02:39.565	00:02:39.630	00:00:00.065
2	ALEC BALDWIN	Gesture	raised eyebrows	frown	turn			00:03:01.184	00:03:01.654	00:00:00.470
3	Cristela Alonzo	Gesture	frowning	frown	smile			00:01:25.480	00:01:25.980	00:00:00.500
4	Riz Ahmed	Gesture		frown	turn			00:00:58.410	00:00:58.560	00:00:00.150
5	Riz Ahmed	Gesture	turn	frown	tilt			00:06:36.620	00:06:36.800	00:00:00.180

Figure 25. FASTsearch function

Furthermore, statistics on any given annotation can be obtained in ELAN through the Annotation Statistics option in the View menu. An annotation tier must be selected and the system shows different values, such as the number of occurrences and frequency of each annotation in the opened file (Fig. 26).



Annotation	Occurrences	Frequency	Average Duration	Time Ratio	Latency
anacronite	8	0.0044129/19276	2.89662s	0.00578514/8792904	88.365
humor-misunderstanding	3	0.0015617/64276	2.56	0.00293837/607107	0
irony	1	0.0005217/49107	3.08	0.00152516/3126819	3.3835
parody	1	0.0005217/49107	3.08	0.00152516/3126819	3.3835
situational humour	1	0.0005217/49107	3.08	0.00152516/3126819	3.3835

Figure 26. Annotation Statistics function

Finally, statistics can be saved as a text file, which can later be imported into SPSS for further analysis.

5.3.2 Praat

Praat was used to perform the prosodic analysis. Praat is a highly sophisticated piece of software enabling an extremely fine-grained study of prosodic features. For the purposes of this dissertation, only basic functions have been used, as the type of audio files under analysis do not meet the necessary quality conditions for further study, given that all sounds in the programme—interlocutors, audience, music and background sound—are included in just a single audio track, which cannot be broken down into different speaker or sound tracks.

In order to analyse the prosodic contrast between humorous and non-humorous utterances, the first thing to do was to extract the sound from each video file and import it into Praat. The soundtrack was extracted and downloaded in mp3 format using the convert/save option from VLC media player software⁶. Full interview sound files were then open in Praat to select and export smaller

⁶ VideoLan Organization. (n.d.). VLC media player (Version 2.2.6) [Software]. Available from <http://www.videolan.org/vlc/>

files with sound excerpts corresponding to the utterances selected in ELAN (Fig. 27).

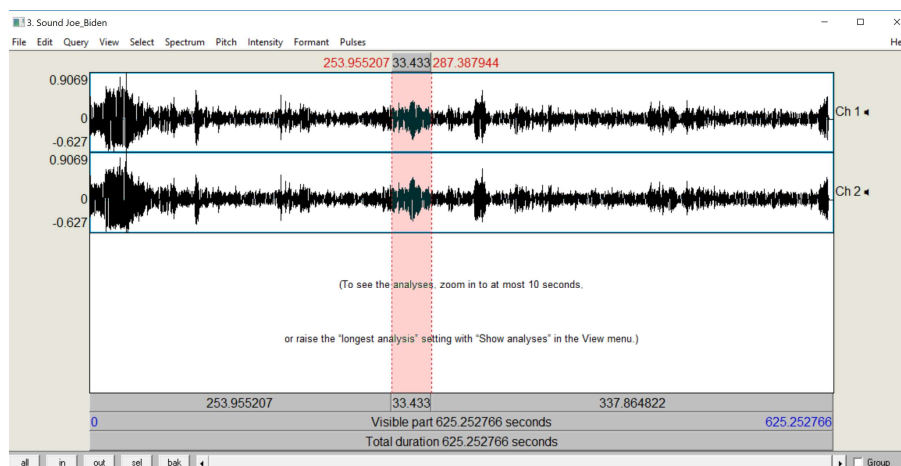


Figure 27. Full sound track of an interview opened in Praat and excerpt selected to be exported.

Once selected, sound files for each utterance were open in Praat to be annotated. Mean F0 and mean intensity values were obtained through the Pitch and Intensity menus respectively (Fig. 28). Both the file and the annotations were later saved as Praat binary files for further analysis and consultation. Mean F0 and intensity values were then copied into the ELAN tier for prosody. Finally, standard deviation of pitch and intensity in pre-base, baseline and humorous utterances was estimated with SPSS.

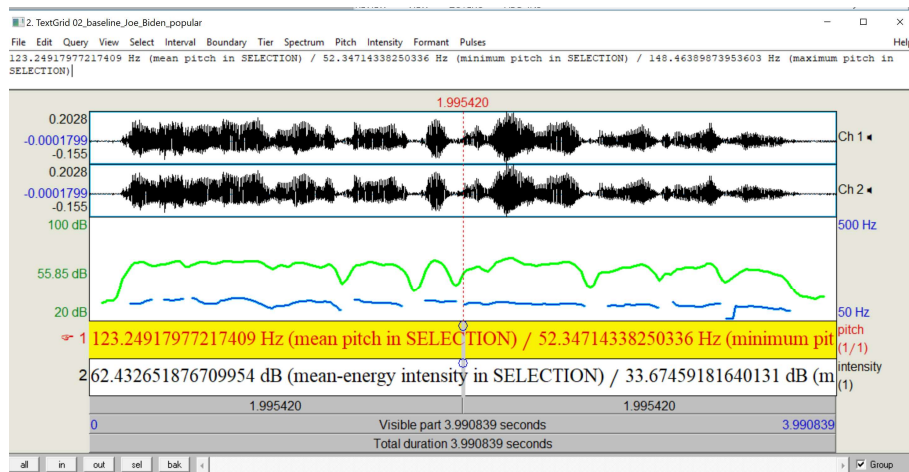


Figure 28. Sound excerpt open in annotation window with Praat.

5.3.3 Segmentation and types of utterance

Following Bryant (2010), three different kinds of utterances were identified with regards to the prosodic analysis: a) Humorous utterances: To identify humour can be a highly subjective task. For the sake of objectivity, utterances were considered humorous if they elicited laughter in the audience. b) Baseline and pre-base utterances were also selected to measure prosodic contrast between humorous and non-humorous instances. Baseline utterances were those said immediately before humorous utterances, whereas pre-base were those immediately preceding baseline utterances. A control analysis could thus be performed comparing pitch (F0) and intensity values not only between humorous and non-humorous (baseline) utterances, but also between non-humorous utterances (pre-base / baseline). Table 3 includes the total number and length of all types of utterances analysed.

Interview	Interview (s.)	Number of humorous utterances	Total length of humour instances	Number of pre-base utterances	Length of pre-base	Number of baseline utterances	Length of baseline
Alec Baldwin	432,030	13	43,055	8	21,524	8	20,114
Alison Janney	445,262	6	29,206	5	13,372	5	17,334
Amy Schumer	514,597	12	38,495	7	14,960	7	10,974
Condola Rashad	389,859	4	17,430	3	7,980	3	4,810
Cristela Alonzo	364,549	12	44,325	8	16,840	10	21,900
Daniel Kaluuya	404,488	13	45,859	9	20,276	13	22,845
Elon Musk	167,691	4	13,905	3	6,130	4	7,845
Michael Hayden	448,257	5	14,050	0	,000	3	4,410
Joe Biden	625,797	7	18,495	5	11,065	5	13,595
John McWhorter	412,383	7	31,120	5	13,210	5	11,910
Riz Ahmed	543,923	13	39,420	7	16,935	10	28,315
Sheryl Crow	290,501	3	9,535	1	1,160	1	,670
Sigourney Weaver	454,829	5	17,870	3	6,100	2	9,460
Susan Sarandon	676,788	5	11,560	2	3,260	3	5,560
Total	N 14	14	14	14	14	14	14
Sum	6170,954	109	374,325	66	152,812	79	179,742

Table 3. Utterances studied

5.4 Humour types

Different taxonomies of humour can be found in the literature, based on different perspectives and put forward by various authors. One widely accepted major distinction is between conversational, or situational, jokes and canned jokes. As defined by Attardo (1994, pp. 296, 297), a canned joke is a joke which has been used before the time of the utterance in a form similar to that used by the speaker, such as those which are found in books, collections of jokes, etc. A conversational joke, in turn, is improvised during a conversation and draws heavily on contextual information for its setup. Conversational jokes are therefore bound by the context, while canned jokes are context-independent. In fact, sometimes the difference between this two types of jokes is not so stark, but the discussion on the topic is beyond the scope of this thesis.

An often quoted classification of verbal humour is the one produced by Dynel (2009), based on a broad survey of previous literature. I agree with her claim that the categories established are not always clear-cut and can overlap at times; they are not mutually exclusive either.

For the analysis conducted in this thesis, I have followed the classification of humour types used in the Corinth corpus (Feyaerts et al., 2010; Feyaerts, 2013; Tabacaru, 2014), as it provides a comprehensive and fine-tuned framework of analysis.

This classification includes 23 types of humour as follows:

Situational humour, narrative joke, pun, irony, sarcasm, exaggeration, understatement, teasing, hyper-understanding,

misunderstanding, parody, register humour, anecdote, absurd humour, insider humour, joint fantasy, stereotype humour, meta-humour, self-mockery, self-glorification, gender humour, sexual humour, inter-textual humour.

5.4.1 Humour types

5.4.1.1 *Situational humour*

Situational humour refers to a type of non-verbal humour, which does not rely on words or verbal communication. It could be defined as based on an event incongruity, stemming from a breach of expectations. Situational humour is considered unintentional and is mainly perceived as humorous by actors non-involved in the humorous event itself (Elleström, 2002).

One example of situational humour in our corpus can be found in Sheryl Crow's interview (Fig. 29), when she closes a statement made on an imaginary presidential campaign in which she would run against Kid Rock by kicking in the air with her right foot. In this example, the act of kicking is clearly intentional, but humour arguably arises as such a reaction is perceived as incongruous by the audience, with no verbal communication involved.



Figure 29. Sheryl Crow kicking

5.4.1.2 Narrative joke

Narrative jokes may come in different formats (short story, riddle, dialogue, etc.). They are defined by the punchline at the end, i.e. a surprising ending leading to a shift in interpretation whereby the humorous effect is produced (Attardo & Raskin, 1991). Narrative jokes have been extensively studied in the literature (Raskin, 1985; Attardo & Raskin, 1991; Attardo, 1993, etc.). No narrative jokes have been found in the current sample.

5.4.1.3 Puns

Dynel (2009) defines puns as humorous instantiations based on the ambiguity and different possible interpretations of a word or set of words (idioms, collocations, etc.). Broadly speaking, puns are a play on words. Consider example (34) from our sample:

(34) Amy Schumer [on incident with PETA⁷ activists complaining about her wearing a coat with coyote fur]: I promised, I was like, PETA, I promised I would never wear fur again...visibly.

In this case, the pun lies in the word ‘visibly’, as it is highly incongruous with the situation explained, in which the expectation in order to comply with PETA demands would be for her to avoid wearing fur at all in the future, and not just visibly. The humorous effect is created by breaking such expectations through playing with the words,

⁷ PETA (People for the Ethical Treatment of Animals) is an American animal rights organisation: <https://www.peta.org/>

5.4.1.4 Irony

The discussion on irony as a type of humour is beyond the study at hand. Dynel (2009) states that not all ironic utterances are humorous. For the purposes of this dissertation, however, irony will be considered as a type of humour, as well as sarcasm.

Irony has traditionally been defined as the opposite of what is literally expressed. It lies at the difference between the literal and intended meaning of an utterance (Grice, 1989). Giora (1998) later defined irony as a pretence space where norms and expectations are violated.

Example (35) below presents an instance of irony in our sample.

(35) Sheryl Crow [on question whether she had got invited to the White House]: I didn't. I'm shocked that president Trump hasn't asked me to come by.

Sheryl Crow is a well-known democrat, who has spoken out to criticise Trump and his administration. It is therefore obvious that she would never be invited to the White House by Donald Trump. By stating that she is 'shocked' that she was not invited, she is producing irony, as there is a clear difference between her literal words and the fact that, as a Trump's critic, she cannot possibly expect to receive such an invitation.

5.4.1.5 Sarcasm

Sarcasm is considered to be a more aggressive form of irony, with a clear target criticised (Attardo, 2000b). In this light, the main difference between irony and sarcasm is that the latter implies an attack or mockery of the given target.

In example (36), General Hayden makes a sarcastic remark when he states that President Trump seemed to have forgotten that he was president of the U.S. and that, in that capacity, he had the authority to find out whether the wiretapping of Trump Tower he accused Obama of having ordered had actually happened. Obviously, it is implausible to think that someone can forget he or she holds such a high and powerful position, so General Hayden is making that remark to criticise Trump on his handling of this issue.

(36) General Hayden: He seemed to have forgotten that he was the president of the US.

5.4.1.6 Understatement

Tabacaru (2014) defines understatement as the opposite of exaggeration. Humour lies in a conspicuously played-down presentation of the facts. Consider example (37) below:

(37) Elon Musk [on question whether he is a superhero or a supervillain]:
I try to do useful things.

The host, Stephen Colbert, points to the fact that Elon Musk has to be the real Tony Stark (Iron Man), given that Elon Musk is a billionaire and that he is working on so many advanced technologies. So Stephen Colbert insists that Elon

Musk must be either a superhero or a supervillain, as he has all the main features of these characters. Elon Musk plays down the comparison and his work by stating that he just tries to do useful things. If we consider that Elon Musk's Space X and Tesla missions and ultimate goals include to enable people to live in other planets and to accelerate the advent of sustainable transport, it is obvious that those high ambitions exceed the mere desire to do useful things.

5.4.1.8 Teasing

Teasing is based on mock challenges, mimicry, etc. with the purpose of amusing both interlocutors (Dyner, 2009). Teasing relies on a common ground shared by interlocutors, who are aware that the other is not being serious (Veale et al., 2006). Let's take the following example:

(38) Alec Baldwin [Stephen Colbert complains he never gets invited to parties such as the ones organised by Obama at the White House]:

Why do you think that is?

When asking Alec Baldwin if he has ever been invited to White House parties, the host complains about himself never getting invited. Alec Baldwin does not answer the question; instead, he cunningly asks back why Stephen Colbert thinks he (the host) is never invited to such parties, clearly implying that he is not wanted as a guest at those parties. This is a pseudo-aggressive remark, said in a playful tone and clearly perceived as such by the host, who reacts by acknowledging the humour.

It could also be argued that other types of humour are at stake in this example. There is not always clear-cut boundaries among these humour categories, and various types can be at play in any given utterance. In (38), irony is also used, in the sense of implying the answer just by literally stating the question. Also, this example could be taken as an instance of hyper-understanding, explained in 5.4.1.9.

5.4.1.9 Hyper-understanding

In hyper-understanding humour relies on the ability of a speaker to detect and exploit weak points in the interlocutor's speech for their benefit. By doing so, humour stems from reversing or changing the interlocutor's intending meaning. As stated before, example (38) could also be described as an instance of hyper-understanding, as Alec Baldwin shifts the focus of the host's question to the host's himself.

Another example of hyper-understanding found in the sample is the following:

(39) Stephen Colbert: Do you know things that you can't tell me that I'd be fascinated to hear?

General Hayden: Yes.

Stephen Colbert: Dumb question. Can you give us even a hint?

General Hayden: Can you keep a secret?

Stephen Colbert: Turn off the cameras! Yes, I can keep a secret.

General Hayden: Me too.

In this example, by asking the host if he can keep a secret, General Hayden leads Stephen Colbert to believe that he is going to reveal some juicy, unknown piece of information. Instead, General Hayden answers his own question too by stating that he can also keep secrets and that he is therefore not going to unveil anything at all. General Hayden exploits the interpretation Stephen Colbert makes of his previous question to reverse and change the perceived meaning by making the real one explicit. As with example (38), (39) could also be considered an instance of teasing, as both types of humour —teasing and hyper-understanding— are clearly at play.

Both (38) and (39) could also fall under Dynel's (2009) definition of retort as a subtype of witticism in the form of a response to a previous utterance with which it forms an adjacency pair. It has also been given the name of trumping, as a kind of adversarial game (Veale et al., 2006).

5.4.1.10 Misunderstanding

Brône (2008) defined misunderstanding as a genuine misinterpretation by a participant due to an opposition between a salient and non-salient interpretation. As opposed to hyper-understanding, there is no intention on playing with the words or the situation to mock the other participant, but rather a true mismatch in the understanding of the event, which has a humorous effect. No occurrences of misunderstanding have been found in our sample.

5.4.1.11 Parody

Parody involves the humorous imitation of another person in a given situation, exaggerating certain traits in order to produce a humorous effect. For

example, in (40) Alec Baldwin explains how he achieves his highly acclaimed impersonation of Donald Trump:

(40) Alec Baldwin: Left eyebrow up. Right eyebrow down. Shove your face like you're trying to suck the chrome off the fender of your car.

In this case, the parody relies predominantly on the face he pulls to mimic Trump, rather than in Baldwin's words. Again, we are confronted to another humorous instance in which different humour types intersect, as the way he describes how to fake a "Trump face" is clearly sarcastic. Another example of parody is found in (41) below:

(41) Daniel Kaluuya: It's like...What would I say...? If I was white...



Figure 30. Daniel Kaluuya mimicking Stephen Colbert

In this example, Daniel Kaluuya is impersonating Stephen Colbert, pointing to the fact that the host awkwardly looked at his nails while discussing

racial issues and what Mr Kaluuya defined as the weird way that white people behaved with blacks. Again, parody is not the only type of humour present in this instance, as by mimicking the host, Daniel Kaluuya is also teasing him.

5.4.1.12 Register humour

Register humour involves a sudden change in language register—for instance, formal speech with interspersed slang—which produces a humorous effect. No examples of register humour have been found in the sample under analysis.

5.4.1.13 Anecdote

An anecdote is a story based on personal experience or on other people's lives (Dyner, 2009), which may produce a humorous effect (either intended or not). Perhaps not surprisingly, anecdote is the type of humour with the higher number of occurrences in this sample. Guests are in the show to talk about themselves, their work, etc., so it is understandable that anecdotes are so frequent. One such example is (42) below:

- (42) Cristela Alonzo [ending a story of how she went after and confronted a thief who had broken into her car and taken a Sephora bag from it]: (Speaking to a neighbour) I'm like: I've got my stuff back, don't worry about that. I look through my bag and I'm like, thank God, 'cause there was this lipstick I bought that was sold everywhere and I got it back.

In this case, the end of the anecdote has a humorous effect given the incongruity of facing such a potential danger—to chase and confront a criminal—just to recover a lipstick.

5.4.1.14 Absurd humour

Absurd humour lies on the violation of Grice's Maxim of Relevance (Attardo, 1993). Humour is attained by deliberately producing discourse in which any kind of logic with regards to the context is absent. No examples of absurd humour have been found in the sample.

5.4.1.15 Insider humour

Insider humour relies on references shared or made clear by participants in the event (Tabacaru, 2014). Common ground can be construed either prior to or during the speech event or conversation. The humour will only be perceived and understood by those sharing the same given references. Consider example (43):

- (43) Alison Janney [talking about Sean Spicer, Press Secretary for Trump's Administration from January to July 2017]: I mean, he is remarkable because he seems to be in a state of... fight and flight at the same time. You know what I mean?

In this example, only people familiar with Mr Spicer and with his press conferences as Trump's Press Secretary can understand the joke and partake in the humorous effect.

5.4.1.16 Joint fantasy

Joint fantasy implies multiple partners jointly creating a funny, imaginary situation or experience (Kotthoff, 2007). Let's take the following instance as an example:

(44) Amy Schumer [After explaining an incident with PETA over a coat with coyote fur she was wearing, in which she promised never to wear fur again visibly]:

At all times I'm wearing tortured ostrich underwear, and I want you all to know that. You can't see it.

Stephen Colbert: I've got a mental image.

Amy Schumer: I torture them in my apartment.

Here, Amy Schumer is building a fantasy into which she invites the audience by saying "I want you all to know that". Stephen Colbert, the host, contributes in turn to construing the fantasy by stating that he can picture it in his mind.

5.4.1.17 Stereotype humour

Stereotype humour revolves around received ideas and clichés. This type of humour exploits expectations about certain behaviours of individuals or communities which are widely known. For example, consider (45) and (46) below:

(45) Cristela Alonzo: We're Latino, we don't believe in depression.

(46) Cristela Alonzo: I'm Latina, I can't be in Law & Order.

In both examples, she's playing with stereotypes associated with Latinos, according to expectations on this group of people in the U.S. In (46), in particular, there is arguably a strong component of irony too, as she is implying the links between crime and the Latino community which prevails in part of the American population, notably highlighted by the current Trump Administration.

5.4.1.18 Meta-humour

Meta-humour arises when explicit reference to humour is made in an utterance, as in the following instance:

(47) VP Joe Biden: Now, you're all gonna laugh when you hear this.

In this example, Vice President Biden is making an explicit reference to the humorous effect he thinks his next words are going to have on the audience. Incidentally, this meta-humour reference turned out to elicit more laughter in the audience than the subsequent utterance, perhaps due to different expectations raised on what he would say next.

5.4.1.19 Self-mockery

Self-mockery occurs when speakers make fun of themselves and they are the target of their own joke. Consider example (48):

(48) Alec Baldwin: I'm sorry I didn't dress up for you.

Stephen Colbert: No, you're always dressed up. You look like you're about to do like Taekwondo or something.

Alec Baldwin: Or I'm about to clean your pool.

In (48), Alec Baldwin decides to mock himself and the clothes he's wearing for the interview by resorting to humorous self-denigration.

5.4.1.20 Self-glorification

Self-glorification is the opposite of self-mockery. In this case, the speaker achieves a humorous effect by praising themselves, as in (49) below:

(49) VP Joe Biden [On the possibility of running for President in the next election]: Hell, Donald Trump will be 74. I'll be 77. In better shape, I mean...

Once again, more than one type of humour are at play in this example. Apart from self-glorification, it could be argued that the statement is also sarcastic, as it may be intended more as an attack on D. Trump (whose fitness for office has often been questioned), rather than as a praise on Joe Biden himself.

5.4.1.21 Gender humour

To some extent, gender humour could be considered as a subtype of stereotype humour, hinging on gender-specific issues. No examples of gender humour have been found in our sample.

5.4.1.22 Sexual humour

Sexual humour relies on implicit or explicit reference to sexual issues. Raskin (1985) states that sexual humour arises when there is an opposition

between a sexual and non-sexual script or interpretation. Only one instance of sexual humour has been found in our sample (50):

(50) Amy Schumer [taking magazine with a provocative picture of her in the cover shown by the host]: Why are the pages stuck together?

There is a clear albeit implicit allusion to the act of male masturbation as a result of sexual arousal upon looking at her picture in the cover.

5.4.1.23 Inter-textual humour

Inter-textual humour involves references made to well-known sources shared by the participants in the event, from literature, media, cinema, TV, religion, etc. References need to be easy recognisable by the participants in the speech event for the humorous effect to occur. Consider example (51):

(51) Alec Baldwin [speaking about the casual clothes he had chosen to wear for the interview]: I call it the Springsteen look.

In this instance, Alec Baldwin refers to a highly popular and worldwide known American singer, Bruce Springsteen, usually dressed in casual style, to define how he was dressed for the interview.

As mentioned before, although a classification of humour types has been followed for the sake of analysis, boundaries between humour types are often blurred, and more than one humour type may concur in any given instance. Extensive humour research based on a large corpus would no doubt contribute to better define and detect categories (Tabacaru, 2014). In the current analysis,

in order to avoid subjective bias, a minimalist approach has been applied. That is, only one or two types of humour have been defined per utterance, choosing only the most salient and clear types. For example, in (52):

(52) Riz Ahmed [on question whether he had experienced problems entering the U.S. given his Pakistani origins]: Well, now I get searched by fans, so they can be really thorough, actually, really take the time with the frisking.

This instance has been annotated as an anecdote. It could be argued that there may be some sexual allusion when referring the thoroughness of searches conducted by fans at airports, but as I did not deem this sexual reference to be salient or clear, I decided not to annotate this utterance as an instance of sexual humour too. On the contrary, in (53):

(53) Daniel Kaluuya [explaining how he got to meet Oprah]: Even the story before is a gas story. So basically like: Denzel came out to me at the Globes...

To me, it is clear that two humour types are salient in this instance. On the one side, (53) is a clear example of an anecdote, in which Daniel Kaluuya is starting to tell a personal story. On the other, I believe it is also clear example of self-glorification, as he tells the anecdote calling Denzel Washington by his first name, and mentioning the Globes as if it were an event he was used to attending. All this gives the impression of him being a well-established celebrity, approached by famous actors, and being cool and at ease in this kind of event. However, as

he himself acknowledged in his subsequent utterance (“No; I look like a wanker now, don’t I”), he was actually nervous and overwhelmed by the situation, as he is a young actor who was at The Globes for the first time.

In this case, given that both types of humour are equally clear, I decided to annotate them both. As a result, the number of humour types annotated in the sample (123) is higher than the number of humorous utterances studied (109).

5.4.2 Overview of humour types in sample

As shown in Table 4 and Fig. 31, the highest number of occurrences in terms of humour type corresponds to anecdote. This is perhaps not surprising, as guest celebrities are typically interviewed in this type of show to talk about themselves and their work. In addition, the genre imposes a light tone with a touch of humour all through the show, so it is understandable that interviewees talk about themselves in a jocular manner.

Anecdote is followed by irony and parody as, respectively, the second and third most frequent type of humour found in the sample. Again, it could be argued that the playful tone set in the show leads to amusing interviews in which overt criticism or grave statements are avoided, resorting instead to ironic remarks when negative comments are to be made. Furthermore, parody is used in the sample as a useful resource to give emphasis to personal stories told and to make them livelier.

Humour type	Number of occurrences in sample	Relative frequency
anecdote	33	26.83%
irony	17	13.82%
parody	16	13.00%
self-mockery	9	7.32%
exaggeration	7	5.69%
insider humour	7	5.69%
sarcasm	6	4.88%
hyper-understanding	5	4.06%
stereotype humour	5	4.06%
teasing	4	3.25%
situational humour	3	2.44%
inter-textual humour	2	1.63%
joint fantasy	2	1.63%
self-glorification	2	1.63%
absurd humour	1	0.81%
meta-humour	1	0.81%
pun	1	0.81%
sexual humour	1	0.81%
understatement	1	0.81%

Table 4. Humour types found in sample

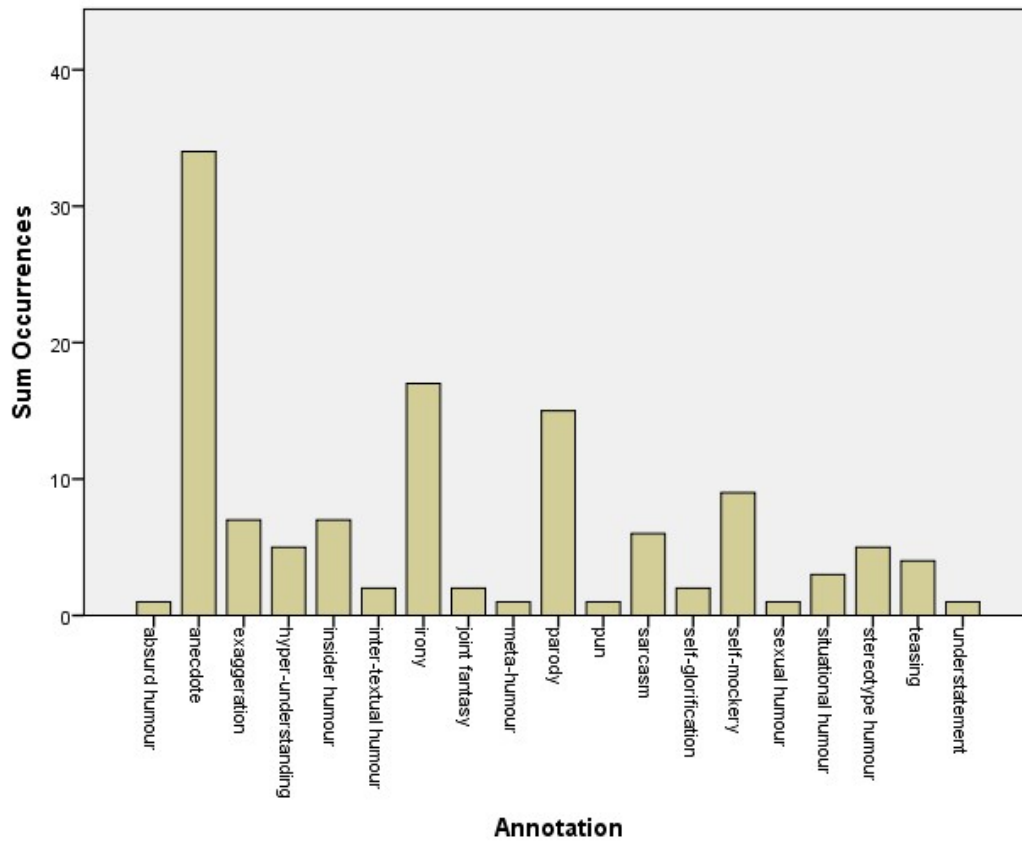


Figure 47. Humour types found in sample

5.5 Prosodic analysis

Prosody can be defined as a set of suprasegmental oral features stretching over one or more than one consecutive utterances (Cruttenden, 1986). In order to perform a thorough multimodal analysis of face-to-face communication, not only gestures, but prosody, must be taken into account to establish the role it plays in conveying a message. All suprasegmental phenomena that are constituted by the interplay of vocal features such as pitch, loudness, pause,

stress, tempo, duration, rhythm, and voice quality can be understood as prosodic (Wennerstrom, 2001).

The stream of speech is perceptibly broken up into units which can be interpreted (Brazil, 1997, p.5). Speech-prosody interaction can occur at different levels. Interactional, discrete, units: lexico-syntactic units for pragmatic actions co-occur with prosodic structures, i.e. turn-constructive units. Turn-constructive units begin with syntactic units and are marked by prosodic units as well. Pre-positioned items may occur (syntactically and prosodically marked). Pauses and turn-holding are possible, whereas projected clauses can be continued. Syntactic and prosodic structures may be expanded. Bauman (2006) states that prosody plays a role in information structure (Jannedy & Mendoza-Denton, 2005), making new information more salient than old information. The study of prosody needs to take into account the full spectrum of communicative functions: expressive, attitudinal, interactive, evaluative and stylistic in order to grasp its full communicative value (Kohler, 2006, p. 146).

Synchrony is not enough to show what modalities are linked, as signals flow in a stream: facial expressions or gestures may come earlier or later than the words they co-occur with. Enfield and Sidnell (2014) proposed that it is the recognizable construction, interpretation and common engaging in activities such as word searching or re-enactment of events what enables participants to relate various modalities to one another, seeing them as part of a larger system.

Different studies have been conducted on the prosody of humour, without conclusive results. Attardo, Pickering, and Baker (2011) found that punch lines,

puns, at the end of a conversational joke were delivered in a lower pitch but they concluded it was because punchlines occur at the end of a paratone (end of narrative), not because humour was prosodically marked. Punchlines did seem to be delivered at a slower pace, but the difference in speech rate was not statistically significant.

Bryant (2010) measured the prosodic contrast between ironic and non-ironic statements, measuring 5 acoustic dimensions: mean fundamental frequency (F0), F0 variability, mean amplitude —intensity—, amplitude variability, and speech rate, i.e. mean syllable duration. Only in speech rate were consistent changes found, with ironic utterances delivered more slowly. Archakis et al., (2010) found systematic differences in the occurrence of jab lines, which are one-liners occurring any time in conversation, not preceded by a narrative, as opposed to punch lines (Pickering, Corduas, Eisterhold, Seifried, Eggleston, & Attardo, 2009). In Archakis' study, jab lines were preceded by significant pauses and delivered at a different speech rate and intensity.

Therefore, there is a lack of consistency in results regarding the prosodic analysis of humorous utterances (Attardo et al., 2003; Pickering et al., 2009; Attardo, Pickering, & Baker, 2011). With regards to humour marking, some prosodic components might be explained by the need of “contrasting utterances and their meaning, and ultimately, to highlight the ‘mention-factor’ value of an utterance” (Urios-Aparisi & Wagner, 2011 p. 525).

In light of the above and partly following Bryant (2010) methodology, I decided to study if consistent prosodic contrast could be found between

humorous and non-humorous utterances to establish whether humorous utterances were prosodically marked in terms of intensity and pitch. To do so, I first chose humorous utterances as those utterances immediately preceding laughter in the audience, which was used as the criterion to determine whether a given utterance was humorous or not, in order to avoid subjective bias on the part of the annotator. Once humorous utterances had been selected, baseline utterances were delimited, defined as those immediately preceding humorous ones. Finally, as a control group, pre-baseline utterances were also selected, whenever possible, as those pronounced immediately before baseline utterances. As in Bryant (2010), that allows for comparisons between humorous and non-humorous utterances —baseline—, as well as between non-humorous utterances alone —baseline and pre-baseline utterances—.

Given that the focus of the present study is to analyse spontaneous humorous communication, speech was segmented into utterances. The segmentation was based on a usage-based approach, in which an utterance is equated to a usage event and defined as a unit-like instance of linguistic behaviour in which the speaker “attempts to achieve a particular interactional goal or set of goals using particular **linguistic** and **non-linguistic strategies**” (bold in original text) (Evans & Green, 2006 p. 130).

Utterances are unit like or “somewhat discrete” (ibid.) units to the extent that they convey a coherent idea, although it is highly difficult to provide a set of criteria to always precisely and unequivocally delimit utterance-units, due to the fact that utterance delimitation may simultaneously rely on grammar, semantics, phonological and pragmatic features which do not co-occur in fixed patterns. As

a result, segmentation was made trying to delimit utterances as stand-alone units, prosodically, semantically and pragmatically meaningful. Three types of utterances were identified:

a) Humorous utterances: the utterance immediately delivered before laughter in the audience. Laughter and humour are not always and necessarily linked, as one may well occur without the other (Attardo, 1994; Attardo, 2003). However, the idea of laughter as an outcome of humour is widely acknowledged in the literature (Attardo, 1994; Kotthoff, 2000; Hay, 2001; etc.) and has extensively been used as a parameter associated to humour in humour studies (Archakis & Tsakona, 2005; Flamson et al., 2011; Tabacaru, 2014; etc.).

b). Baseline utterances: the last words before the humorous utterances. Speaker's laughter or sounds were not included (Bryant, 2010).

c) Pre-baseline utterances: utterances immediately preceding baseline utterances.

Although prosodic contrast can be noticeable even between relatively distant utterances (e.g. over 10s.) (Bryant, 2010), for this analysis only utterances immediately preceding the target or baseline utterances (as baseline and pre-baseline utterances, respectively) were taken into account, so that we could safely consider that all utterances belonged to the same discursive and thematic unit.

As explained in section 5.3.3, utterances were first identified and segmented in ELAN. Then they were extracted from the Praat sound file of the given interview, to conduct single-utterance analysis as individual files.

Utterance-files were analysed and annotated in Praat, in order to obtain mean pitch (F0) and mean intensity. Those values were then annotated in ELAN. Statistical analysis of prosodic data was performed in SPSS (v. 20).

It was not always possible to pair a pre-base and baseline utterance with each humorous utterance identified. Also, the fact that there was only a single audio track including both the host's and guest's speech, as well as sounds coming from the audience, background noise and music, made it impossible to analyse the prosodic features of all utterances selected due to overlaps. When overlaps affected just a small part of the beginning or end of the utterance (overlaps under 10% of total utterance time), that part was excluded from the analysis and values on the remainder of the utterance were obtained. On the contrary, when overlaps were significantly longer (over 10% of total utterance time) or occurred in the middle of the utterance, the whole utterance was discarded for the prosodic analysis.

Table 3 summarised the number and length of utterances studied. Figure 32 shows the total length of each type of utterance studied:

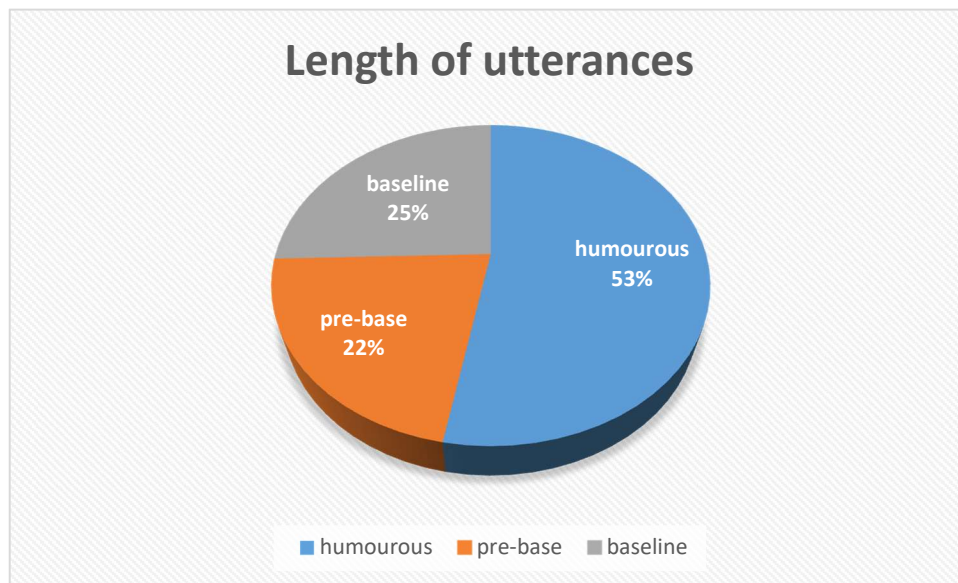


Figure 32. Total length of each type of utterance studied.

Mean pitch (F0 in Hz) and mean intensity (in dB) were obtained for each utterance. Then, all data was recorded in SPSS in order to estimate the standard deviation (SD) in mean intensity and mean pitch, for each type of utterance per interview, as a proxy measure of variability and prosodic contrast (Purandare & Litman, 2006; Bryant, 2010). SD values per type of utterance were compared within speakers through t-tests (independent variables) or non-parametric tests to determine whether there was a statistically significant difference in SD, which would lead to conclude that prosodic contrast in F0 and intensity was in turn significant. Kolmogoro-Smirnov normality tests were conducted prior to comparing SD values in order to determine which kind of test was suitable. An additional one-way ANOVA analysis of variance, along with a post-hoc LSD test, was conducted in order to corroborate the results obtained through independent-

variable t-tests. SD values in mean intensity and mean pitch were compared and analysed for the following pairs of utterances: humorous / baseline utterances and baseline / pre-baseline utterances within speakers (Table 5).

Interview	Type of utterance	of N -utterances)	Mean intensity (dB)	SD intensity	Mean pitch: F0 (Hz)	SD pitch
Alec Baldwin	pre-base	7	61,62	2,97	137,57	19,94
	baseline	8	64,19	1,81	140,21	19,70
	humour	10	66,46	4,43	185,97	70,30
Alison Janney	pre-base	4	62,50	5,27	165,56	37,04
	baseline	5	61,50	3,85	147,53	45,79
	humour	5	58,23	2,99	156,20	34,64
Amy Schumer	Pre-base	6	63,38	3,40	219,19	29,02
	baseline	6	67,97	9,04	225,54	32,33
	humour	11	62,04	2,45	196,42	26,40
Condola Rashad	Pre-base	3	63,57	2,27	232,68	38,19
	baseline	3	66,34	2,22	241,44	22,06
	humour	4	62,66	2,26	224,99	19,05
Cristela Alonso	pre-base	7	63,99	2,12	258,94	26,44
	baseline	9	64,29	2,89	260,02	43,63

	humour	12	62,25	2,82	260,07	58,89
Daniel Kaluuya	Pre-base	9	68,37	3,83	188,94	66,54
	baseline	11	67,94	3,46	170,05	49,67
	humour	12	65,33	3,93	162,33	48,20
Elon Musk	Pre-base	2	64,91	2,43	165,99	43,10
	baseline	3	64,92	2,67	179,17	75,22
	humour	4	65,48	3,81	144,43	5,19
Joseph Biden	Pre-base	5	60,04	1,59	146,21	18,35
	baseline	5	59,88	3,54	137,66	33,02
	humour	6	60,57	6,06	137,40	29,69
John McWorther	Pre-base	5	62,13	2,77	142,59	35,65
	baseline	5	51,66	25,26	134,14	27,36
	humour	6	67,10	6,83	151,27	53,27
Michael Haydn	baseline	3	64,01	1,79	182,21	6,87
	humour	5	62,16	2,21	174,10	11,49
Riz Ahmed	Pre-base	6	63,33	1,07	161,61	9,73
	baseline	9	64,13	2,56	167,72	21,54
	humour	8	62,69	2,31	158,59	11,97
Sheryl Crow	pre-base	1	61,53	n/a	205,99	n/a
	baseline	1	65,69	n/a	185,92	n/a
	humour	1*	61,53	n/a	192,53	n/a

Sigourney Weaver	Pre-base	2	62,20	1,45	177,14	11,87
	baseline	2	63,39	0,14	186,04	32,00
	humour	4	63,99	1,39	182,23	15,34
Susan Sarandon	Pre-base	1	64,33	n/a	177,63	n/a
	baseline	2	63,87	2,62	180,74	26,56
	humour	4	60,29	5,62	189,11	41,67

**There are two additional humorous utterances, not included in the prosodic analysis as one is speechless and for the other the speech cannot be isolated from overlapping sound.*

Table 5. Prosodic data per type of utterance in each interview.

5.5.1 Overview of results

The full results of the prosodic analysis conducted for each interview can be found in Annex 1. No statistically significant differences in SD values for F0 and intensity were found in the sample ($p = 0.05$), neither between baseline and humorous utterances, nor between baseline and pre-baseline utterances. The results in the present study bear out previous research on spontaneous humour, i.e. no prosodic contrast has been found between humorous and non-humorous utterances, when it comes to F0 and intensity SD values. Admittedly, the setting and casual tone of the programme, prone to humour, may not require humour be made particularly salient through prosodic cues.

5.6 Gestures

Communication is inherently multimodal: a multimodal analysis of discourse and human communication brings together information conveyed through different modalities, such as gesture, gaze, prosody, posture, etc. A comprehensive account of communication can only be achieved through careful exploration of how these modalities interplay.

Therefore, in order to try to fully comprehend and account for communication, analyses should focus on the interplay between gestures and speech, which Kendon (2004) defines as visible utterance action. That is, a multimodal utterance, comprising gestures, speech, and prosody.

Humour is arguably one of the most complex instances of communication, both in terms of production and comprehension. Research has taken an interest in studying how, or if, humour is multimodally marked (Attardo et al., 2003; Cheang & Pell, 2009; Attardo, Pickering, & Baker, 2011; Attardo, Wagner, & Urios-Aparisi, 2011, among others). My research focuses on how co-speech face movements and prosody interplay in spontaneous humorous utterances, in order to account for how humorous knowledge is multimodally constructed, represented and communicated through language, gestures and prosody.

Given the constraints of the sample, only face and head movements have been studied in this thesis. Due to camera shots and viewpoints, it was not warranted to get a full clear view of the upper body at all times, hence the focus on the head and face.

Whereas speech is used to convey complex situations and make reference to abstract points, gestures may be used to make the pragmatic value of verbal speech explicit (Muller & Cienki, 2009). It is not single cues that suggest particular interpretations, but rather their co-occurrence and density.

As explained in Chapter 3, gestures can perform the following functions: discursive, referential and interactional. Of the types of gestures along the Kendon's continuum (McNeill 1992), the ones relevant for the current analysis are gesticulations, i.e.: co-speech gestures made unwittingly by the speaker, mostly with discursive and interactional value. The gestures annotated in our sample are face and head movements, identified as follows.

5.6.1 Head movements

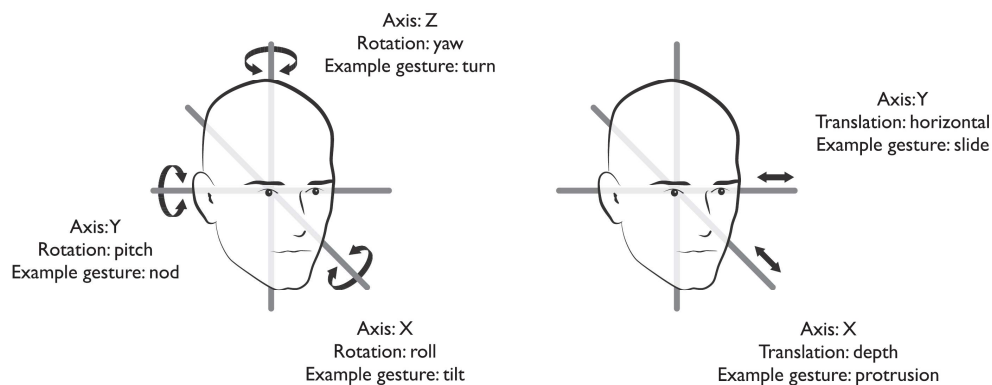


Figure 33. Head movements (Wagner, Malisz, & Kopp, 2014, p. 212)

Communicative head movements involve rotations around three axis, as shown in Fig. 33: axis Y, axis Z, axis X. In addition, two linear displacements are also common, along the Y and X axis respectively (Wagner et al., 2014).

It must be noticed that the actual realisation of these movements can be of great variety, in terms of extent, angles, etc. (Kousidis, Malisz, Wagner, & Schlangen, 2013). Head movements may overlap, such as a tilt and turn shown in Fig. 34:



Figure 34. Tilt and turn

The list of annotated head movements is the following:

- a) Tilt: rotation around axis X.



Figure 35. Tilt

b) Turn: rotation around axis Z.



Figure 36. Turn

c) Upstroke: upward rotation around axis Y.



Figure 37. Upstroke

d) Nod: downward rotation around axis Y.



Figure 38. Nod

e) Protrusion: forward slide along axis X.



Figure 39. Protrusion

f) Retraction: backward slide along axis X.



Figure 40. Retraction

Head gestures need to be interpreted in their multimodal context, taking into account other multimodal cues such as prosodic features, hand or face movements, etc. (Poggi, D'Errico, & Vincze, 2010). Head movements are varied and mainly pragmatic. Despite the polysemy entailed in head movements. Some of them can be associated with a semantic core, such as nods used by speakers to underscore the importance of what is being said (Poggi et al., 2010) or turns linked to the idea of negativity (Kendon, 2002).

5.6.2 Face gestures

Regarding facial expressions, Poggi and Pelachaud (1998), drawing on previous work by other authors (Ekman & Friesen, 1971; C. Goodwin, 1980; Ekman, 1982; Castelfranchi & Poggi, 1990), distinguished the following possible functions:

- a) Affective display: seven possible prototypical face expressions for emotions (happiness, sadness, anger, fear, disgust, surprise and embarrassment).
- b) Syntactic function: accents, emphasis, question marks, etc.
- c) Dialogic function: turn taking (eye and head movements).
- d) Referring function: to refer to an emotion is not currently being felt by the speaker. Included are deictic gaze, or facial emblems, even iconic eye movements (squinting as opposed to wide open).
- e) Attitude display: convey social attitude towards the addressee (anger, ordering, imploring, etc.).

In humour analysis literature, gestures are referred to as kinesic markers (Muecke, 1978) or gestural triggers (Tabacaru, 2014), helping to frame the utterance as humorous or stressing a shift in focus, clarifying by way of gestures what might be ambiguous by resorting only to speech, or just reinforcing the humorous message. All in all, gestures are considered to be multimodal cues intended to help listeners grasp the humorous intention. However, markers do not confer the humorous turn to speech, as this can be humorous even in the absence of such multimodal cues (Attardo et al., 2003).

With regards to face movements, the gestures annotated have been the following:

a) Smile



Figure 41. Smile

b) Frowning



Figure 42. Frowning

c) Raised eyebrows



Figure 42. Raised eyebrows

d) Blank face



Figure 43. Blank face

Both head and face movements may overlap, and they may be produced repeatedly. For the current analysis, a minimalist approach has been applied, annotating only the type of gesture produced, regardless of the angle, pace of repetition or extent. I have strived to annotate all repetitions of a given gesture separately. The reason to apply this minimalist approach is the fact that camera shots were not always taken from the same viewpoint or angle. As a result, despite a movement analysis conducted frame by frame, it could not possibly be ensured that every single movement was unmistakably and accurately accounted for. Therefore, I decided to opt for a more coarse-grain, almost-qualitative analysis, to detect co-speech gesture and prosody interplay patterns which could ultimately point to the direction that future and more fine-tuned research should take.

This minimalist approach enables understanding of multimodal communication. In order to provide a comprehensive and nuanced account of the

role played by any given gesture in an utterance, all gesture features should be taken into account, namely movement rate, amplitude, speed, repetitions, etc. (Bressemer, 2012; Wagner et al., 2014). Heylen (2008) suggest differences be made between repeated shakes, head sweeps, slow head movements, repositioning and other lateral movements to assign difference functions to each type of move. Having said that, the painstaking work required to annotate gestures and movements, not to mention features such as repetition, amplitude, etc., needs specific software, movement recognition devices and human resources that were not available for the research described in this thesis.

5.6.3 Overview of results

Table 6 and Fig. 45 summarise the results of the analysis conducted in terms of annotated gestures:

Gesture	Number of occurrences in sample	Relative frequency
nod	91	20.36%
tilt	79	16.67%
raised eyebrows	65	14.54%
turn	53	11.86%
shake	47	11.86%
smile	41	9.17%
upstroke	35	7.83%
frowning	17	3.80%
protrusion	12	2.68%
other ⁸	3	0.67%
retraction	2	0.45%
blank face	2	0.45%

Table 6. Gestures found in sample

⁸ The three instances classified as “other” include two instances of mimicking (therefore, not unintentional) in Amy Schumer and Alec Baldwin respectively. The third one refers to an utterance in Susan Sarandon’s interview in which she is looking to the ceiling and it is not possible to detect and annotate any head or face gesture.

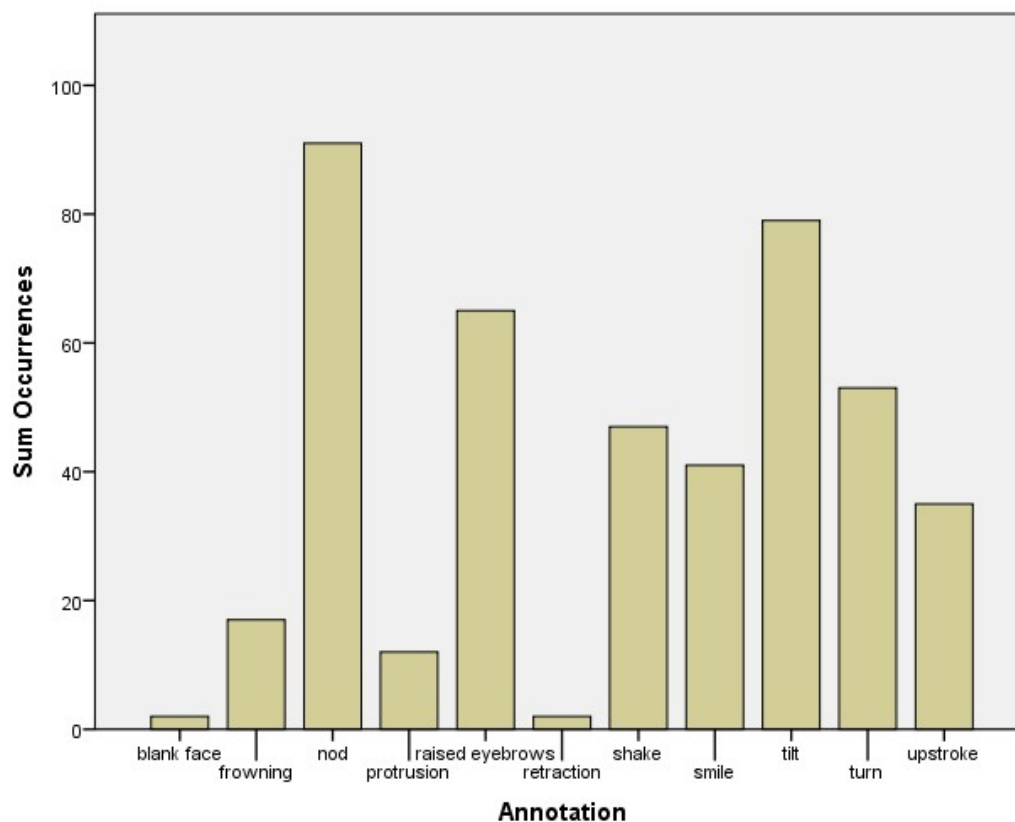


Figure 45. Gestures found in sample

5.7 Construal mechanisms

The purpose of this dissertation is to study how humour is multimodally expressed and cognitively motivated. To that end, an analysis of construal mechanisms (Croft & Cruse, 2004) underlying humorous utterances has been conducted. The classification of construal mechanisms used in this thesis is the one provided by Croft and Cruse (2004), which is summarised in Fig. 46 below. All types of construal mechanisms have been defined and illustrated with examples in Chapter 4.

Table 3.1 *Linguistic construal operations as instances of general cognitive processes*

I. Attention/salience
A. Selection
1. Profiling
2. Metonymy
B. Scope (dominion)
1. Scope of predication
2. Search domains
3. Accessibility
C. Scalar adjustment
1. Quantitative (abstraction)
2. Qualitative (schematization)
D. Dynamic
1. Fictive motion
2. Summary/sequential scanning
II. Judgement/comparison (including identity image schemas)
A. Categorization (framing)
B. Metaphor
C. Figure/ground
III. Perspective/situatedness
A. Viewpoint
1. Vantage point
2. Orientation
B. Deixis
1. Spatiotemporal (including spatial image schemas)
2. Epistemic (common ground)
3. Empathy
C. Subjectivity/objectivity
IV. Constitution/Gestalt (including most other image schemas)
A. Structural schematization
1. Individuation (boundedness, unity/multiplicity, etc.)
2. Topological/geometric schematization (container, etc.)
3. Scale
B. Force dynamics
C. Relationality (entity/interconnection)

Figure 46. Classification of construal mechanisms (Croft & Cruse, 2004, p. 46)

Figure 47 below shows the distribution of construal categories underlying humorous utterances found in the sample.

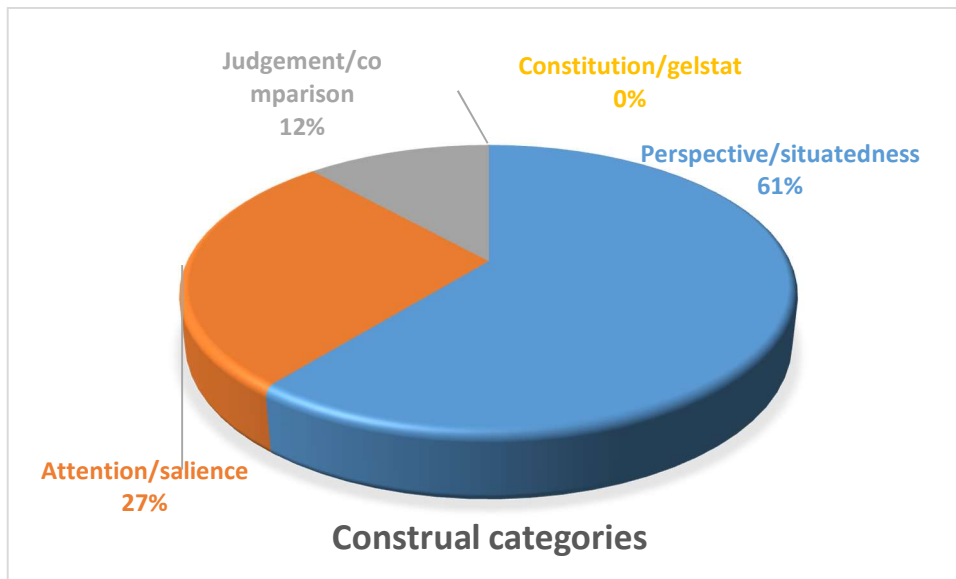


Figure 47. Distribution of construal categories found in sample

5.7.1 Overview of results

Table 7 summarises the number of occurrences per type of construal mechanism identified in the sample.

Annotation	Sum	Relative frequency
viewpoint	46	38.01%
profiling	23	19.00%
subjectivity	14	16.94%
dynamic attention (sequential scanning)	8	9.68%
deixis spatiotemporal	8	9.68%
categorization	7	5.78%
metaphor	7	5.78%
deixis (empathy)	4	3.30%
deixis epistemic	1	0.83%
metonymy	1	0.83%
objectivity	1	0.83%
scope of predication	1	0.83%

Table 7. Type and number of occurrences of construal mechanisms identified in the sample.

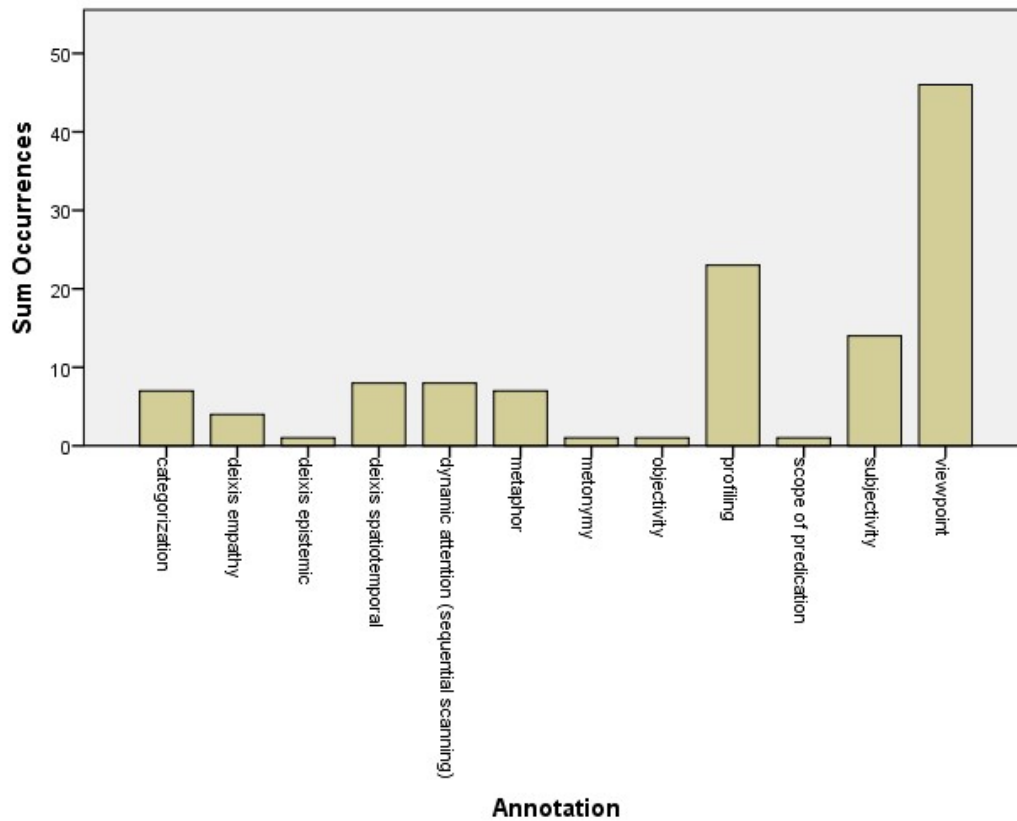


Figure 48. Construal mechanisms found in sample.

5.8 Conclusion

In this chapter, a full account of the sample studied and the methodology followed has been provided. First, a brief explanation was offered as to why I have favoured the study of spontaneous, non-staged humorous communication to avoid the necessarily more ostentatious multimodal features in staged humour, as it is intended to reach a very wide and unknown audience. The different elements employed to conduct the multimodal and cognitive analysis of humorous utterances have subsequently been presented in different sections of the chapter. Section 5.2 focuses on the collection of the interviews included in the sample. It offers a quick overview of the number of interviews selected, as

well as of the number and length of humorous and non-humorous utterances studied. Section 5.3 explains the various annotation tools used for the analysis, notable ELAN, as a video annotation software, and Praat, with which the prosodic study was undertaken.

Section 5.4 presents the classification of humour types used in this study, with definitions of each type of humour. Whenever possible, examples from the sample have been included to expand on the definition of the humour types found. The multimodal part of the analysis is described in sections 5.5 and 5.6. In 5.5 an overview of the prosodic analysis and types of utterances studied is provided. Section 5.6 includes a list of gestures annotated in the sample. Only face and head movements have been studied, given that a clear view of the full upper body was not always available, and therefore it was not possible to systematically include other types of gestures, such as hand movements. Images from the sample analysed are included for each gesture listed. Finally, as in previous sections, the chapter offers overview of the results obtained with regards to construal mechanisms.

Chapter 6

A multimodal analysis of humour

6.1 Introduction

In this chapter, a multimodal analysis of some of the humorous instances found in the sample is presented, taking the quantitative results obtained in terms of frequency of gesture type as a starting point. As a result, I zoom in on the following gestures, with the highest number of occurrences in the corpus, as shown in Table 6 in Chapter 5: head nods, head tilts, raised eyebrows, head turns, head shakes, and smiles.

The gestural movements of the hands and arms are probably the most studied co-speech gestures (Wagner et al., 2014), with less attention devoted to head or face expressions (Hadar et al., 1985; Poggi & Pelachaud, 1998; McClave, 2000; Kendon, 2002; Lee & Marsella, 2010; Kousidis et al., 2014; Ishi, Ishiguro, & Hagita, 2014; Tabacaru, 2014; etc.). Head movements and raised eyebrows have been found to serve as beats (Hadar et al., 1983; Pelachaud, Badler, & Steedman, 1996; Krahmer & Swerts, 2007; Guaiatella et al., 2009; Flecha-García, 2010; Tabacaru, 2014), that is, non-representational gestures used to punctuate speech (Kendon, 1980; McNeill, 1992). Head nods are considered to generally signal agreement (Lee & Marsella, 2010), whereas head shakes are associated with explicit or implicit negative statements (Kendon, 2002).

Given the sample constrains as to what parts were visible at any given moment, as well as image definition as a result of the angle of the camera shot during the interviews, I focused on annotating head movements, i.e., nods, upstrokes, shakes, turns, tilts, protrusion, retraction, and certain face gestures, namely raised eyebrows, frowning, blank face and smile/laughter. Therefore, my analysis is a contribution to the previously mentioned studies, and will draw on their results to confront my own analysis with them. My research is novel insofar as it deals with humour in spontaneous communication and joins the multimodal analysis to a study of the cognitive mechanisms underlying humour.

Face-to-face interaction is inherently multimodal. Modalities belong to different systems of representation, i.e., language, gestures, and posture, and they translate into actions which mediate between the concrete and the abstract. Modalities are considered as complex cultural tools, to the extent that they are socially acquired or learnt and used in interaction (Norris, 2013). Gestures and body movements help to organise turn-taking (Mondada, 2013). Gestures mainly have discursive value and are essentially pragmatic (Cienki, 2013a).

What follows is an in-depth analysis of the pragmatic and discursive value of the head and face gestures most frequently found in the present sample. The first section is devoted to facial displays, specifically raised eyebrows and smile/laughter, followed by a section on head movements, with special emphasis on tilts and nods. Cross-reference data on the occurrence of each type of gestures with humour types and construal mechanisms is also included, as shown in Tables 8, and 9.

	nod	tilt	raised eyebrows	smile	turn	shake
irony	19	11	9	6	15	2
inter-textual	2				2	
self-mockery	2	6	5	1		5
teasing	1	1	4	4		1
hyper-understanding	2	2	2	4	1	1
parody	5	11	8	7	6	10
anecdote	15	31	15	9	19	11
insider humour	17	2	1	1	2	1
sarcasm	3	5	2	1	2	3
exaggeration	8	3	5	2	3	9
pun	6					
joint fantasy	3		2	1		1
stereotype humour	3	4	7	2	2	2
self-glorification		1	2			
understatement	1	1				1
absurd humour	2	1		1		
meta-humour	1		1			
Situational humour	1		2	2	1	

Table 8. Head and face gestures per type of humour

	nod	tilt	raised eyebrows	smile	turn	shake
profiling	4	1		1	1	2
categorization	5			1	3	2
metaphor	8	7	5	4	2	6
profiling	17	8	13	7	9	7
viewpoint	24	46	23	14	17	17
spatiotemporal deixis	4	4	4	2	5	1
deixis empathy	2		3			1
subjectivity	4	8	5	6	2	5
dynamic attention	7	5	5	3	2	1
scope of predication			1			1
objectivity	1					1
deixis epistemic			1	2		1
metonymy	4				1	

Table 9. Face and head gestures per type of construal

6.2 Facial expressions

Face gestures have been assigned various communicative functions in the literature (Poggi & Pelachaud, 1998): (1) affective display, with seven universal (Ekman & Friesen, 1971) prototypical faces showing happiness, sadness, fear, anger, disgust, surprise (Ekman, 1982), and embarrassment (Castelfranchi & Poggi, 1990); (2) syntactic function, when facial expressions punctuate

questions, emphasis, intonational accents, pauses, etc. (Poggi & Pelachaud, 1998); (3) dialogic function, when face signals are produced in interaction to manage turn taking (C. Goodwin, 1980); (4) referring function, when speakers use face movements to refer to an emotion that they are not feeling at the moment of speaking (Ekman, 1979); (5) attitude display, when face gestures express the speaker's attitude towards the interlocutor, e.g. to express anger through frowning, or agreement through a smile (Poggi & Pelachaud, 1998).

Ekman (1979) argued that there are instinctive face movements which do not manifest as people suppress them under certain circumstances, e.g. "in the presence of an authoritative figure" (ibid. p. 180), whereas certain learnt behaviours may be so embedded that they occur without the speaker being aware of it. Ekman, therefore, concluded that the distinction between voluntary/involuntary or intended/unintended gestures is not relevant as it does not reflect reality.

6.2.1 Raised eyebrows

6.2.1.2 The meaning of raised eyebrows

Eyebrow movement studies go back to Ekman seminal work on the issue (1979). He advanced that facial signals could be used either to express emotions or as conversational tools. He distinguished 7 distinct eyebrows actions, called action units: 5 involved in the expression of emotions, such as anger, fear, surprise, disgust, etc., and 2 used as signals in conversation, mainly to punctuate speech or to underline certain bits of information (Fig. 49). Ekman claimed that both functions are essential to grasp the full complexity of communication.

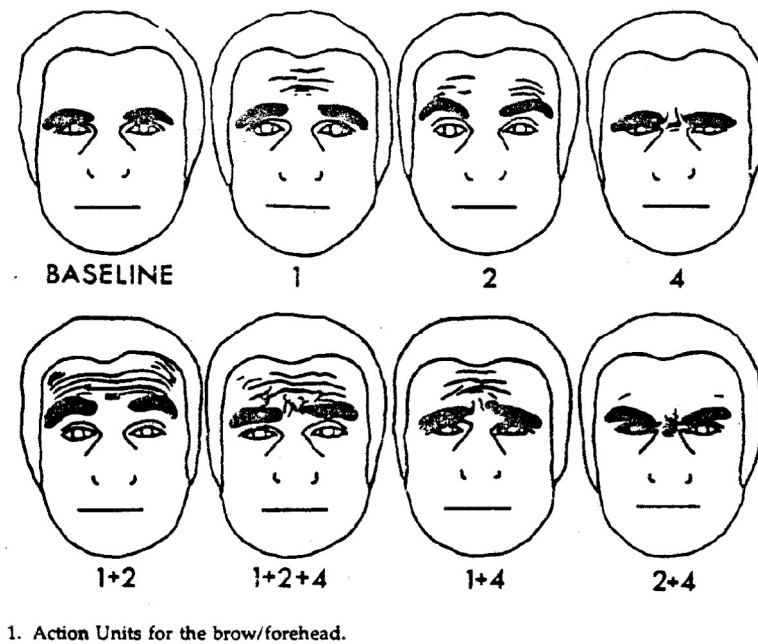


Figure 49. Action Units for the brow/forehead (Ekman, 1979, p. 174)

Citing Darwin (1872), Guaiatella et al. (2009) referred to a primitive function of raised eyebrows to enlarge the visual field and thus enhance visibility. This could be associated with the traditional link between raised eyebrows and the expression of surprise or astonishment, as acknowledging that something deserves increased attention. They drew a distinction between slow and rapid eyebrow movements and claimed that rapid raised eyebrows in conversation are used as an attention-getting device by interlocutors and are related to turn-taking. This is consistent with Kim, Cvejic, and Davis's (2014) claim that eyebrow movements may be used to alert the hearer about an important upcoming bit of information. They analysed the interplay between brow and head movements, and concluded that the latter might serve to underline content over bigger chunks

of speech. However, they posited that differences in gesture timing and realisation may be simply due to physical properties, e.g., muscles that need activating, etc.

Flecha-García (2010) defended that raised eyebrows may serve to guide the coherent structure of discourse, as participants tended to raise eyebrows more at the beginning of segments in which they introduced a new topic. In addition, they were used to underline certain bits of information. Tabacaru (2014) delved into the multimodality of humour, and provided an in-depth analysis of the relation between raised eyebrows and frowning with the humorous instances she identifies in her corpus. She found that raised eyebrows are consistently associated to sarcasm, and concluded that they can serve as gestural triggers (Tabacaru & Lemmens, 2014), which co-occur with the “core humorous part of an utterance” (Tabacaru, 2014, p. 217) to signal that the utterance has to be interpreted as humorous.

6.2.1.2 Eyebrows and prosody

Recent studies have zoomed into the meaning of eyebrow movements and their interplay with prosody. Flecha-García (2010) remarked that raised eyebrows and pitch accents were aligned in English, the former slightly preceding the latter, pointing to the use of raised eyebrows to underline certain parts of the utterance. She also noticed that raised eyebrows occurred more frequently at the beginning of large discourse segments. She further claimed that regardless of whether raised eyebrows are intended or unintended gestures, that is, made consciously or unconsciously by the speaker, they can both be considered as signs, as both

nevertheless carry meaning which can be interpreted by the interlocutor. She hypothesised that raised eyebrows may be intended signals when used for emphasis and unintended gestures, stemming from the speaker's speech production process, when functioning as markers of discourse structure.

The issue of the connexion between raised eyebrows and prosody was also briefly touched upon by Tabacaru (2014), who claimed that raised eyebrows are used as gesture-triggers for humour, more specifically for sarcasm. The distinction posited by Flecha-García (2010) is in line with Cavé et al.'s (1996) classification of gestures into rhythmic, i.e. expressive gestures, associated with rises in fundamental frequency, and dynamogenic, i.e. discursive gestures, linked to flat F0 curves. Guaiatella et al. (2009) also focused on eyebrow movement correlation with prosodic features, namely, with changes in F0. They concluded that eyebrow rises are used as tools to gain the interlocutor's attention, either to express the will to take the turn in conversation or to signal the importance of a particular portion of speech. Their study focused on rapid eyebrow movements, but they argued that both rapid and slow movements in general can have semiotic value and have the potential to be used as rhythmic or dynamogenic gestures.

6.2.1.3 Raised eyebrows in sample

Regarding facial movements, raised eyebrows appear with the highest number of occurrences (65) in the present sample. As seen in Table 9, the occurrences cluster around the most frequent construal mechanisms, i.e., profiling and viewpoint, and are more or less evenly distributed among less frequent construals. As for humour types (Table 8), the distribution pattern is

consistent with what can be statistically expected, given the high number of occurrences both in raised eyebrows along with anecdote, irony, and parody, which are the most frequent humour types found. Interestingly, no strong correlation appears between raised eyebrows and sarcasm, as opposed to results obtained by Tabacaru (2014), although, arguably, the number of instances of sarcasm in the present sample (6) is too low to draw any significant conclusion. Let's now analyse examples of the most frequent instances, as per the data summarised in the tables.

Example (54) shows an instance of raised eyebrows in an anecdote. Sigourney Weaver is speaking about her recent trip to Cuba, and how she was so pleasantly surprised to see the classical education afforded to all artists. At one point, she is describing the Art School building, likening it to a womb, as if artists would emerge from it re-born after their artistic education period. Arguably, the instance elicits humour in the audience due to the unlikely comparison of the building to a womb, and the specific description of the entrance to the building as having the form of a vagina.

(54) Sigourney Weaver: It's in an art school built on the golf course. They got rid of the golf course after the revolution and they built this womb-like... It's literally...You walk into a vagina-like opening...



It's....



literally...



You walk into a...

Figure 50. Raised eyebrows in example (54)

Sigourney Weaver raises her eyebrows at the beginning of the sentence “It’s literally”. Her gesture could be interpreted as a way of signalling the non-conventional nature of what she is going to say next. I don’t believe she intends the utterance to be humorous, although it is perceived as such by the audience. Rather, she is simply describing the building and conveying her thoughts and emotions about it in the most effective way possible. Therefore, raising her eyebrows may be a signal of acknowledgment of the unexpected nature of her description and choice of words, or as a way to alert the hearer, as if to say, “I know this is going to be surprising to you”.

Arguably, though, this particular instance of raised-eyebrows cannot be linked to a discursive function, as no change in turn takes place either before or after the statement, and given that there is no new topic introduced: the building as womb metaphor had already been presented in the previous remark. The utterance is accompanied by a variety of gestures, among which head tilts, nods, frowning, and smiles, in addition to representational hand gestures depicting the entrance to the building. Furthermore, therefore, in this particular case no link can be drawn between the raised eyebrows and the humorous nature of the utterance, as, in my view, the speaker does not intend to produce humour.

The following example (55) shows an instance of irony, which is produced intentionally by the speaker. Sheryl Crow is discussing her latest album, which includes a song about collusion between Russia and a fictional American President, far earlier than suspicions arose about possible Russian interference in 2016 American elections.

(55) Sheryl Crow: I was sitting with my buddy, going, “Wow, what’s the craziest thing you can imagine... like Russia, like Russia hacking. That’d be crazy. And like, uh...the President being involved with Russia. That’d be so fucked up!”



The President...

Figure 51. Raised eyebrows in example (55)

Sheryl Crow raises her eyebrows upon uttering “The President...”, so it can be claimed that she underlines the bit of information which for her is the most important and where the highest incongruity lies, as if conveying that the idea of having a President colluding with a foreign power is so serious that it is beyond imagination. In this case, “The President” is not only highlighted by the raised eyebrows, but also by protruding the head. In addition, the prosodic accent falls on the word ‘president’. Hence, various multimodal cues, i.e., accent, raised eyebrows, and protrusion of the head align to give clear prominence to that part of the sentence, which is arguably the most relevant for the speaker, and, in Tabacaru’s (2014) terms, where the core of the irony, in this particular example, is placed.

Irony arises as she is portraying a situation in which she, along with her colleague, came up with what they thought was a crazy idea, and the product of their wild imagination. However, it turned out to be the object of a federal investigation within the U.S. administration, as probable cause for Russian meddling and, possibly, collusion with the then candidate Donald Trump, had been acknowledged. Clearly, she is underlining the implausibility of her idea at the time to highlight the opposition between the situation as they construed it then, in which no grounds to suspect anything of the like were possible, and the current circumstances, in which the legitimacy of the President himself is questioned.

Finally, in example (56) we find an instance of raised eyebrows and parody. Stephen Colbert is interviewing Alec Baldwin, and he brings up a letter that Alec Baldwin received from President Nixon after Baldwin had lost the election for president of his school at George Washington University. Alec Baldwin then takes the letter to read what he considers to be the best part of it.

(56) Alec Baldwin: You know what the greatest part of this thing is? It's that he writes: "From our mutual friend Mark Weinberg I've learnt of the disappointing results, as far as you are concerned".



From our mutual friend...

...as far as you are concerned.

Figure 52. Raised eyebrows in example (56)

In this example, Alec Baldwin resorts to parody to delimit the part of the letter that he finds most interesting, as conveying the lack of tact by Nixon, or simply the fact that he did not feel sorry for Alec Baldwin's defeat. In order to do so, the actor mimics precisely those Nixon's words, as opposed to the first excerpt from the letter that he reads normally. The parody is shown by a change in voice quality, significant lower pitch, a palm-up gesture, head shake, and raised eyebrows. On this occasion, eyebrows are raised over the entire remark "as far as you are concerned". As I see it, in this particular instance the use of the raised eyebrows could be twofold. On the one hand, to frame the chunk of letter that Alec Baldwin considers more significant, arguably because it is the most telling part about Nixon's attitude towards his defeat, or because he feels it showed lack of empathy. On the other, the raised eyebrows could also be associated to the expression of surprise felt by Alec Baldwin on reading that part of the letter. Perhaps, the raised eyebrows in this case can also be an indication by Alec Baldwin to the hearer of the unexpected nature of such words.

Let's now turn to cases of raised eyebrows in instances in which the underlying mechanisms are viewpoint and profiling, as there is where the occurrence of raised eyebrows clusters with regards to the type of construal operations. As said before, this is to be expected because those are the two most frequent construal operations found in the sample. An instance of raised eyebrows against the backdrop of viewpoint as a construal mechanism can be found in John McWhorter's interview. To the question of whether there is a word in English whose meaning has not changed through time, the lexicographer answers that the most frequently used and dullest words are the ones that change the least, and gives the conjunction 'and' as an example, which elicits laughter in the audience and has therefore been annotated as a humorous utterance:

(57) John McWhorter: For example, 'and'. I have nothing remotely interesting to say about 'and'.



Figure 53. Raised eyebrows in example (57)

This is an interesting example, as the speaker keeps his eyebrows raised throughout the whole statement, but he raises them a bit higher upon uttering

'and', 'nothing', and 'interesting', which, in this case, are the more meaningful words. Again, raised eyebrows are aligned with the most prominent words, so I argue that they serve to reinforce the information structure of the utterance by underlining the most significant bits of information provided.

As for a co-occurrence of raised eyebrows and profiling, an example can be extracted from General Michael Hayden's interview, in a humorous instance already discussed in Chapter 2 as a token of hyper-understanding. General Hayden tricks the host into thinking that he is going to give away a juicy piece of information by asking a question usually put when people are going to say something important. However, he subsequently profiles just the literal meaning of the question, by answering it himself too in order to stress that he is bound to keeping secrecy.

(58) General Hayden: Can you keep a secret?

Stephen Colbert: Turn off the cameras...I can keep a secret.

General Hayden: Me too.



Me...



...too.

Figure 54. Raised eyebrows in example (58)

General Hayden raises his eyebrows upon the pronunciation of “Me too”. Admittedly, raised eyebrows in this example serve to highlight the unexpected turn in the conversation, as under normal circumstances, speakers do not ask this question in order to answer it themselves. It can also be associated to the idea of the surprise that is intended to cause in the hearer. Furthermore, by underlining the incongruity underlying his reply, General Hayden’s clear humorous intent with this remark is reinforced.

The results of the analysis above are consistent with the claim that raised eyebrows can be used as underliners (Ekman, 1979; Tabacaru, 2014) to highlight parts of the speech that the speaker considers that deserve more attention. They may therefore be used to direct the hearer’s attention to the most meaningful parts of the speech to favour a correct interpretation, as intended by the speaker. Furthermore, raised eyebrows tend to align with prosody to make salient elements in the utterance clearer to the interlocutor; to that end, the upper part of the face, the area around the eyes, encompasses the strongest cues to prominence (Swerts & Kraemer, 2008).

6.2.2 Smile and laughter

Laughter and humour do not consistently go hand-in-hand, as the former can occur without the latter and vice-versa (Morreal, 1983). In fact, some argue that laughter is not even an intrinsically social phenomenon, as there can be solitary laughter (Weeks, 2016). Laughter is considered meaningful in itself and can be used to refer forward or backward in the conversation (Jefferson et al., 1987). Norrick (1993) argued that laughter is the natural response to humour by

the hearer, and that its absence points to failed humour. Provine (2000) and Hay (2001), nevertheless, dissociated humour and laughter as a necessary and inevitable response. Having said that, laughter remains as a valid indicator of humour in the literature, given how frequently they co-occur (Holt & Glenn, 2013; Gironzetti, 2017).

With regards to smiles, they have traditionally been associated with the expression of enjoyment (Frank, Ekman, & Friesen, 1983). Only recently has research taken an interest into exploring from a multimodal perspective how smiling and humour are related (Harris & Alvarado, 2005; Attardo et al., 2013; Ikeda & Bysouth, 2013; Gironzetti & Menjo, 2014; Gironzetti et al., 2015). Ikeda and Bysouth (2013) associated smile and laughter, along with other multimodal cues, e.g. gaze, as features that, combined, may have communicative value to signal, for example, appreciation. Attardo et al. (2013), in turn, found in their study that smile and laughter were the only multimodal cues frequently co-occurring with humour. They concluded, though, that neither smiling nor laughter can be taken as markers of humour, as they are not consistently associated with it.

For them, laughter and smile are not linked to the hearer's reaction to humour, but to the way speakers signal their humorous intention, as part of a negotiation between interlocutors to determine when a given turn is to be comprehended as humorous (Attardo, 1994; Eisterhold, Attardo, & Boxer, 2006). This hypothesis was borne out by other studies, showing that humorous turns were consistently associated with higher-intensity smiles and smiling synchrony among interlocutors (Gironzetti & Menjo, 2014; Gironzetti et al., 2015). On a different note, some authors consider that laughter promotes cooperative

behaviour and it is not intentional (Smoski & Bachorowski, 2003). Interestingly, listeners can identify whether a person is smiling just thanks to acoustic cues (Drahota, Costall, & Reddy, 2007).

Smile annotations in the sample include some instances of laughter, as the current analysis does not aim at disentangling the nature of both phenomena, but rather at looking into their association with humour and their possible function to frame humorous instances (Attardo et al., 2013). Smiling amounts only to a relative low frequency of gestures found in the sample. This may be due to the nature of the gestures themselves, as raised eyebrows or head movements tend to be shorter and repeated, whereas smiles can be maintained over a longer segment of speech, and therefore be annotated just once, despite covering longer stretches of utterances. A better measure of the co-occurrence of smiles and humour is to look into the percentage of humorous utterances in the sample in which speakers smile.

Smile has been annotated in 41 out of 109 humorous utterances, which means that it is present in the production of 37.61% of the humorous instances identified. Interestingly, there is great variety in the use of smiling across speakers. This frequency reveals that smile or laughter are by no means necessary to frame a given utterance as humorous nor to signal the humorous intent of the speaker. Perhaps, the context itself is prone to humour, given the genre of the programme and the interviews themselves, and it might be a reason why smiling is not be essential, then, to frame any turn as humorous, as this is taken for granted in such a setting.

As for the co-occurrence of smiling with types of humour or construal mechanisms, it is more evenly distributed across all types than raised eyebrows. Still, the highest number of smiles are found in combination with viewpoint, with regards to construal operations, and with anecdote, parody, and irony, in terms of humour types, which again is consistent with the higher presence of these humour forms and construals in the sample.

Cristela Alonzo is the speaker in the sample who most often resorts to smiling upon the production of humorous instances, as in example (59) below, in which she is discussing with the host differences in how they were taught certain aspects of religion as children.

- (59) Cristela Alonzo: You're gonna have a crown in Heaven?
- Stephen Colbert: You get to Heaven, you get a crown. That's what I was told.
- Cristela Alonzo: My mum would say: "Mija, in Heaven, you're gonna get electricity, running water..."



Figure 55. Smile in example (59)

She smiles at the end of the humorous utterance, corresponding to her last statement in the previous dialogue. I believe that smiling in this case is an indication of the humorous nature of Cristela Alonzo's utterance, and of her intent for the remark to be taken as such, despite pointing to a difficult childhood as the daughter of an immigrant Mexican mother with very few financial means. Another example of smile and viewpoint can be found in the same interview, but this time co-occurring with frowning, which, in my view, assigns a different value to the smile. Again, Cristela Alonzo is referring to her childhood, this time to explain how she became passionate about comedy to the point of wanting to become a comedian herself. She explained that she watched a lot of comedy on TV, as she spent a lot of time alone in her house and TV was her main companion.

(60) Stephen Colbert: How did you get to comedy from there? Were there clubs in your town? How did you...?

Cristela Alonzo: No clubs whatsoever. I loved TV. I was a latchkey kid...

Cristela Alonzo: [on —inaudible— reaction by someone in the audience]: Ok... Ooh, times were tough.



Ooh! Times were tough!

Figure 56. Smile and frowning, in example (60)

In this case, Cristela Alonzo's smile, which includes a short bout of laughter, does not co-occur with speech but substitute for it. In my view, after expressing her astonishment at the reaction by a member of the audience with the first 'Ok', she clearly conveys her disapproval through frowning, which has traditionally been taken as a sign for disagreement (Tabacaru, 2014). Also, the position of the hands seem to be signalling disbelief or lack of understanding, as if asking 'What is going on?' or 'What was that?'

Her smile in this case may be taken as an attempt to attenuate the rejection of the audience's reaction so as not to come across as outright confrontational, which would be out of place, and possibly even detrimental for her career, in the context of the programme. In addition, it could be a way of trying to conceal her negative emotions and to save face in order to keep the light and humorous tone expected at the interview, an attitude which is further reinforced by her response,

after which she seems to dismiss the whole episode with a shrug, before turning to the host and resume her speech where she'd left it before this interruption, by repeating "I was a latchkey kid". Beyond illustrating different ways in which smiling can be used, apart from expressing appreciation, encouraging cooperation or signalling humorous intent, I find this particular example to be a showpiece of how different multimodal channels interplay to convey much more than what is simply said.

Turning now to instances of smile co-occurring with the most frequent humour types, we can consider example (61) of smile and anecdote, in which Condola Rashad is telling that she used to go with her mother, a famous actress, to theatres where her mother was playing and act as her assistant, and even take notes of her performance:

(61) Condola Rashad: I would sit in the audience and I took notes. I brought them to her a few times. If I didn't hear a line I said: "You know? That line, at the end of it you just kind of dropped off and I need more".



Figure 57. Smile in example (61)

Once more, we find the smile occurring at the end of the statement, which is often the case in all instances of smiles found in the sample. I claim that Condola Rashad's smiling in this case is not so much as to signal humour as such, but an acknowledgement of the implausibility of the situation she is depicting, with her, a young girl, lecturing her mother, a seasoned and renowned actress, on how she could improve her performance.

Something similar can be said of the following example (62) extracted from Daniel Kaluuya's interview. The British actor is describing his experience at The Globes, as a newcomer in a world of glamour and big film stars. He is specifically telling how he met Denzel Washington, and how overwhelming the whole experience was for him.

(62) Daniel Kaluuya: And Denzel spoke to me, and I'm like, he was like...I was like: I can't believe he knows who I am, I can't believe he knows who I am. And then...[mimicking Denzel's words to his wife] "He can't believe I know who he is".



Figure 58. Smile in example (62)

I posit that this smile, which is produced right after the remark, is not used either to frame the turn as humorous or to signal a humorous intent on the part of the speaker. He is vividly describing the situation from his point of view, in very honest terms; that is, a young, unknown British actor in one of the most important gatherings of the film industry for the first time, getting acquainted with the elite of Hollywood stars.

In my view, his smile is a reflection of his own positive feelings of bewilderment and overwhelm at such a dream come true for him. It is, therefore, an inward-looking smile, which is highlighted by the fact that the smile is not directed at the hearer, and no eye-contact is made by Daniel Kaluuya, which, to me, points that no agreement is sought on the part of the interlocutor in this particular utterance.

Finally, example (63) below illustrates an instance of smiling and irony, from Vice President Biden's interview. It conveys Joseph Biden's answer to the

question of what he felt on the 2016 election night, in which fellow democrat Hillary Clinton lost to Donald Trump.

(63) Vice President Biden: Well, I've been in the closet since then. I haven't come out...No.



Figure 59. Smile in example (63)

Vice President Biden smiles right after pronouncing 'no', which, in turn, negates the meaning of the preceding utterance, intended as ironical. In this case, I see the smile as clearly pointing to the humorous ironical nature of his remark, reinforced by the verbal negation of what he had just stated, as if to dispel any possible doubt about it. Of the last three examples analysed, this is the only smile that can undoubtedly be tied to humour, as intended to frame what has been said as such. On another note, the type of humour or of construal mechanism involved in a given utterance does not seem to bear any strong link with the smile produced.

According to the analysis conducted, smiling is not consistently associated with humour. Even when it does appear in humorous utterances, it is not necessarily to frame humour or as a signal of the humorous intent. Smiling, therefore, is just but one of the various multimodal cues that can be elicited at any given time by speakers to reinforce whichever message they want to convey, whether humorous or not, but it can also be just an expression of attitude or affective display, not necessarily tied to an intended communicative value by the speaker (Ekman, 1979).

In any case, the outcome of my analysis is rather qualitative, as no differences have been drawn between smiles in terms of intensity (Gironzetti and Menjo, 2014; Gironzetti et al., 2015), and as the focus has been placed solely on the production of humorous instances, with no account of smiles and laughter that speakers produce outside those utterances or of smiles elicited in the interlocutor. Further and more fine-tuned research is needed to fully comprehend the nature of the relationship between humour and smile/laughter.

6.3 Head movements

Head posture changes constantly during speech (Hadar et al., 1983). Their exact instantiation is very varied in terms of angles and extent, with different gestures possibly overlapping (Wagner et al., 2014). Head movements can have semantic, discursive, and communicative functions. Head shifts can regulate interaction, yielding or asking for turn (Hadar et al., 1984), or requesting backchanneling (Maynard, 1987; McClave, 2000; Altorfer et al., 2000). Furthermore, head movements by listeners tend to synchronise with those made by speakers (Hadar et al., 1984).

In addition, head movements, as hand gestures (McNeill, 1992), can have deictic functions and locate referents in abstract space; that is, a given space will be signalled by the hands and/or the head to correspond to a certain element of the speech, e.g. different people or places the speaker is referring to, and that space-element correlation will be kept during speech. Furthermore, switches in narrator while speaking, i.e. when the speaker refers someone else's words, can also be indicated by head movements (McClave, 2000).

Nods are typically associated with agreements, while shakes are linked to negation (Kendon, 2002). We usually look away and move the head upwards when thinking about something (Lee & Marsella, 2006). Head movements can also be recruited to signal the humorous nature of what is being said (Lee & Marsella, 2010) or to give prominence to the elements of the utterance which carry the humorous meaning (Tabacaru, 2014). McClave (2000) considered that head gestures can also be the result of cognitive processes, frequently accompanying lexical repairs.

Finally, the relationship between interlocutors has been found to have an impact on the frequency of head motion events, e.g. speakers nod more in interactions with less familiar listeners, than when speaking with relatives or close friends (Ishi et al., 2014). Head movements must be interpreted in the multimodal context in which they occur, as their interplay with other modalities, e.g., face expressions, can have an impact on their function (Poggi et al., 2010; Wagner et al., 2014; Heylen, 2008). For example, head tilts are typically associated with disbelief or lack of understanding, but can yield a different interpretation of surprise or interest when joined by certain facial displays (Heylen, 2008).

6.3.1 Head tilts

Head tilts and head nods are the most frequent gestures found during humorous utterances in the present sample. Head tilts have been linked to disbelief, lack of understanding (Heylen, 2008), interjections expressing denial, word searching, (Lee & Marsella, 2006), and in segments or speech with weak boundaries, such as unfinished utterances or stretches in which speakers pause to think (El Kalioubi & Robinson, 2004; Ishi et al., 2014). Head tilts loosely align with prosody (Cvejic et al., 2010), although they do not synchronise with certain parts of utterances as strongly as nods or shakes (Ishi et al., 2014). In addition, Tabacaru (2014) claimed that head tilts also serve as gestural triggers, underlining those parts in the utterance which are more significant to enable the humorous interpretation.

In our sample, head tilts have been annotated mainly with anecdote, followed by a large margin by parody and irony. This seems natural, as those are the three most frequent humour types identified in the interviews. As for construal mechanisms, head tilts overwhelmingly cluster around viewpoint.

Example (64) has been extracted from General Hayden's interview. Stephen Colbert had brought the subject of President Trump accusing President Obama of having wiretapped Trump Tower for campaign surveillance purposes. General Hayden explains that to obtain a court order for wiretapping is extremely difficult, and that they are only issued when probable cause of serious criminal activity is proved. The subsequent exchange follows:

(64) Stephen Colbert [after mentioning that President Trump had called for an investigation on the issue]: Can a President just find out by himself that this happened?

General Hayden: Yeah. That's why I wondered what happened on Saturday morning. He seemed...He seemed to have forgotten that he was the President of the United States.



Figure 60. Tilt in example (64)

General Hayden tilts his head upon pronouncing the first 'seemed', and keeps it slightly tilted throughout the utterance, except for a further prominent tilt upon saying 'forgotten', which is additionally punctuated by a nod. The onset of both tilts are clearly aligned with accented words: 'seemed', and 'forgotten'. It could be argued that the first tilt is associated to the unfinished utterance, although there does not seem to be a pause for thinking or retrieving the word. Instead, I believe that General Hayden is marking the beginning of a humorous

sarcastic turn, as if requiring closer attention by the hearer. As said before, the second tilt falls upon 'forgotten'. I posit that both 'seemed' and 'forgotten' are the most relevant elements in the utterance to facilitate its sarcastic interpretation. Hence, in this case, tilts fall within Tabacaru's (2014) gestural triggers, as well as play a part in information structure (McNeill & Levy, 1993; Jannedy & Mendoza-Denton, 2005), highlighting the most relevant pieces of new information.

The following example (65), however, of an occurrence of a head tilt in an instance of parody, with viewpoint as the main underlying construal, does not seem to have the same function. In this case, Riz Ahmed is referring to how he switches to the required accent for the character he has to play in a given series, film, etc. the moment he sets foot in the place where that is going to be filmed, and keeps the character's accent even when he is not shooting. He explains that this leads to embarrassing situations in which, once the shooting is finished and he goes to thank the team, already in his natural British accent, people react badly as they feel betrayed.

(65) Riz Ahmed: They can feel quite betrayed, really. Because we wrap the whole shoot and I go to people: "Thanks a lot, mate. Nice one; I really appreciate it". [Reporting the team's reaction] "I don't even know you, you know?"



Figure 61. Tilt in example (65)

In this instance, there is a short tilt that falls upon ‘know’, while he is reporting the reaction he gets from his team when they realise he talks differently from what they are used to. According to the literature survey reported above on the functions of tilts, this example could be associated to the expression of disbelief, and, in my view, it also has a clear use in terms of information structure, aligning with prosody to underline the most prominent word in that utterance. On another note, it is interesting to note that he does not express the transition to indirect speech verbally. Instead, he turns his head and gaze to a different space allocated to the referent, i.e. the team he has worked with, in a distinct location to identify them. So the switch in narrator is marked by this gaze and head movements, to set up a new space, along with a clear frowning to convey the disapproval he gets from the team members, and by the semantic content of the sentence itself, which can only fit with his previous statements as conveying the reaction of her fellow work colleagues.

Example (66), from Alison Janney's interview, shows an example in which a tilt is combined with a head shake, in an anecdote with viewpoint and profiling as main construal operations. She is explaining how she became interesting in acting while in college, especially after realising she could not major in Psychology, as she was unable to conduct rat experiments in the lab.

(66) Alison Janney: I tried to major in Psychology and I had to deal with the rats, and I was like...I cannot do with the rat-lab part of Psych...Whatever, I don't know.



Figure 62. Tilt and shake in example (66)

Upon uttering 'whatever', she tilts her head to the right, while also performing a shake. The tilt could be associated to the unfinished utterance, i.e., weak boundaries (Ishi et al., 2014), or it might reinforce the expression of rejection and denial, clearly shown by the head shake and the position of the hands, as if pushing away the idea of having to work with rats in the lab. In any case, no clear connection seems to exist between the tilt and the humorous effect

of the utterance, nor to the profiling of the rat experiments as part of how she explains her decision to study theatre at college.

An interesting example from the same interview occurs in a later statement, in which she is telling that she was cast by Paul Newman to act in a university play he himself directed, as a former student of the same college. When asked why she thinks she was chosen by Paul Newman to play the leading role, she explains that, in the casting, she came up with a monologue on fast driving, as she knew Paul Newman was very fond of cars, and she hints to the possibility that her canning and seducing pretence to be a car-fan woman enticed Paul Newman into giving her the role. In the sample, this instance has been annotated as an anecdote, with viewpoint as the main underlying construal mechanism.

(67) Alison Janney:

And it probably had nothing, no reason, to do why I got the part. But I like to think that I...You know... I catfished him into... cast me.



...into...

cast

me



Figure 63. Head tilts in example (67)

The utterance identified as humorous corresponds to the latest part of this statement, i.e. “I catfished him into cast me”. This utterance is punctuated by repeated head tilts. At first, short tilts accompany “I catfished him into...”, whereas the last part, “...cast me”, is framed by slightly more sweeping head tilts. It gives the impression that she is a bit unsure about what she is going to say at the beginning, as the seduction innuendo is clear, but then decides to go for it, but making clear through this head motion of repeated tilts that the remark has to be taken as humorous, and not at face value. Repeated tilts are not common in Western societies, as opposed to in India, where they are widely used for backchanneling (Wagner et al., 2014). The beginning of each head tilt cycle in this instance is clearly aligned with the accented part of the corresponding segment of speech, i.e., “I CATfished him into...CAST me”. The non-serious nature of the utterance is further reinforced by a smile at the end.

In my view, this latter example is proof of the versatility of gestures at large, and head movements in particular, which can be recruited for different functions and that always need to be interpreted in the context in which they are being performed.

6.3.2 Head nods

Nods are typically associated with the semantic expression of agreement or acceptance (Poggi et al., 2010), but they can have different functions (Wagner et al., 2014), such as confirmation, approval, appreciation, emphasis, etc. (Poggi et al., 2010). They can also be used for turn taking or giving (Ishi et al., 2014) and strongly correlate with prosody (Swerts & Krahmer, 2010). Head nods produced

by speakers can function as backchannel requests, to which listeners have been found to be extremely sensitive (McClave, 2000).

Most nods annotated in the sample co-occur with viewpoint and profiling, as far as construal mechanisms are concerned, and with anecdote, parody, and irony, with regards to humour type, thus following the expected statistical pattern, given the highest frequency of these construal operations and humour types, respectively.

The following example (68) illustrates an instance of a single nod in an utterance identified as belonging to hyper-understanding, with profiling as the main underlying mechanism. Stephen Colbert welcomes lexicographer John McWhorter and tells him that he listens to his podcast all the time, to which McWhorter replies “Keep listening”. Arguably, humour arises from the unexpected nature of the remark, as a simple “thanks” what would have typically been expected. By profiling the action of listening to McWhorter’s podcast, instead of taking the remark just as a polite onset of the interview, the lexicographer reverses to a certain extent the host’s statement in order to exploit it for his own benefit.

(68) Stephen Colbert: Thanks for being here. I listen to your podcast
all of the time...

John McWhorter: Keep listening.



Figure 64. Nod in example (68)

John McWhorter produces a single nod, starting upon the pronunciation of ‘keep’, which is also the most prominent word in the utterance, and ending in the baseline position at the end of the word ‘listening’, so both the head movement and prosody strongly correlate. Although it could be claimed that the nod in this example signals appreciation or approval, I believe that in this particular instance it is used to place emphasis on the utterance, as a way of highlighting a non-conventional remark which, therefore, requires more attention and effort to be processed on the part of the hearers.

A different function can be interpreted in example (69), where several nods are performed in the turn, which is Vice President’s Biden reply to the question of what he expects from the Trump administration:

- (69) Vice President Biden: Now, you’re all gonna laugh when I say this [burst of laughter in the audience]. But the honest-to-God truth is, I don’t know. But I’m being deadly honest to you. I don’t think the President himself knows for certain.



You're all gonna

laugh when I say this.

Figure 65. Nod in example (69)

Two utterances were identified as humorous here: “You’re all gonna laugh when I say this”, as an example of meta-humour with empathy deixis as the main construal mechanism, and “I don’t think the President himself knows for certain”, as an instance of insider humour and viewpoint. The nod onset co-occurs with ‘laugh’, which is the accented word in the utterance, and the head keeps down in this single nod until the end of the utterance. I posit that this nod is a backchannel request and that it serves to engage the audience and to make sure the speaker has their full attention.

A further example of what I believe to be a nod to request backchanneling from listeners is found in the subsequent humorous remark. In this case, it correlates with the word ‘certain’, at the end of the statement. In my view, this nod is again intended to capture the hearers’ acknowledgement of what is being said, as if to make sure that they fully grasp its meaning. A different nod is produced in this very same utterance, coinciding with the word ‘President’, which happens to be the most prominent word in that stretch of speech. I claim that this particular nod serves to underline the term ‘President’, as the most relevant word

introducing new information, in which the agent switches from Joseph Biden to the President.

An illustration of a nod signalling confirmation and agreement can be found in example (70). Given the Pakistani background of actor Riz Ahmed, Stephen Colbert has asked him if he has experienced troubles travelling to the US due to his physical appearance. Riz Ahmed replies that he frequently gets “randomly” selected for searches at security control, to which the subsequent exchange follows:

- (70) Stephen Colbert: A secondary search?
- Riz Ahmed: Well, it’s a random search, really.
- Stephen Colbert: Ok, I thought maybe it was a slightly deeper search.
- Riz Ahmed: Well, now I get searched by fans, so they can be really thorough, actually. Yeah.



So they...



can be really...



...thorough...

Figure 66. Nods in example (70)

Riz Ahmed performs a nod co-occurring with ‘they’, and a double nod upon ‘thorough’. Clearly, with his remark, the actor picks up on Stephen Colbert’s tendentious question, agrees with the host’s suggestion, and confirms it himself with his words. The nods, therefore, play a role to signal such confirmation. In addition, they loosely align with the two most prominent words in the last sentence, thus contributing, along with prosody, to information structure, as those are the two key words to comprehend the message.

The last example (71) analysed shows an instance of repeated nodding. Stephen Colbert asks Sigourney Weaver if they have starting shooting the second part of the film *Avatar*, to which she awkwardly replies that they haven’t and that if she had said the opposite in an interview the day before, she was wrong. It so appears that they had indeed starting shooting the film but that the information was probably under embargo and no one was allowed to speak about it. The hosts then said that he understood and further asked confirmation that they would start shooting at some point, to which the actress replied as follows:

(71) Sigourney Weaver: I'm sure we will start at some point.



I'm sure we will



...start...



... at some point.



Figure 67. Nods in example (71)

The whole utterance is punctuated by nods, clearly signalling confirmation. Furthermore, there is strong correlation with prosody, as the word 'start' is made more salient in the statement by means of prosodic accent and also underlined by a longer nod, in which the head is kept down until Sigourney Weaver starts uttering "at some point", when her head shifts to baseline position. A last nod is performed right after the utterance is finished, which, along with gaze directed at the listener, could be interpreted as a signal to yield the turn.

As seen through these examples, nodding in humour performs the same functions as in non-humorous utterances, identified in previous studies (McClave, 2000; Poggi et al., 2010; Swerts & Kraemer, 2010; Ishi et al., 2014). No correlation between nodding and a particular type of humour or construal mechanisms has been found. This analysis also illustrates the polysemy of gestures, even within the same instantiation, in which they can be interpreted as having different functions, e.g. information structuring and confirmation, as in example (70).

6.3.3 Head shakes and head turns

Head shakes and head turns also frequently appear in the sample, although in fewer numbers than nods and tilts. Both are performed along the same axis, with shakes involving repeated sweeps, and turns a single movement (Wagner et al., 2014). In addition, shakes differ from turns in terms of amplitude and the angle of the movement performed (Kendon, 2002). Kendon (2002) argues that head shakes share a core meaning of negation, although their different instantiations can correlate with inclusivity, e.g. with words like ‘everyone’ or ‘everything’, intensification or superlative, e.g., with expressions such as ‘very’, ‘a lot’, ‘exactly’, etc. (C. Goodwin, 1980; Schegloff, 1987), uncertainty, and lexical repairs (McClave, 2000; Kendon, 2002). The underlying implicit notion of negation in head shakes accompanying expressions of intensification or inclusivity is justified as no exceptions are envisaged to what is being conveyed by those expressions (Kendon, 2002). In addition, shakes can co-occur, precede or be performed after speech. With regards to head turns, they are typically associated to discourse management, in terms of turn-taking and

turn-yielding, as well as signals structuring abstract spaces where to locate referents in speech (McClave, 2000).

As for the correlation with types of humour and construals in the sample, it follows an expected pattern, with the highest number of shakes and nods happening along with the most frequent types of construal operations and humour, namely, viewpoint and profiling, on the one side, and anecdote, parody, and irony, on the other. Interestingly, around a third of shakes occur in instances of exaggeration as a type of humour, which is consistent with the association made in the literature between shakes and the expression of intensity or superlative.

One such example is the following, in which Amy Schumer is telling about an incident she had with activists from PETA, an animal welfare organisation, who had criticised her for wearing a coat with a hood lined with coyote fur.

(72) Amy Schumer: If I made a list of the animals I care about more
 than coyotes [burst of laughter in the audience].
 It would be a list of every animal. I don't care.

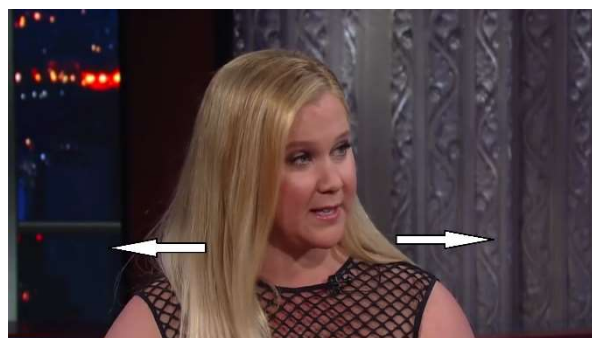


Figure 68. Head shake in example (72)

Here, Amy Schumer performs short repeated head shakes on several occasions. First, upon uttering the word ‘coyotes’. This can be interpreted as the attitude of the speaker about coyotes, which basically is rejecting the idea that they should be protected. Later, while pronouncing the second part of the statement, she keeps shaking her head in short, quick, and low-amplitude movements. In this example, several of the interpretations for headshakes identified in the literature, as explained above, coincide, namely: an expression of inclusivity with ‘every animal’, and negation conveyed in ‘I don’t care’.

Another example from the same interview illustrates a different use of a turn. In this exchange, Amy Schumer is talking about the comments made by some journalists about her body, deemed to be too plump for current Hollywood beauty standards. This instance has been annotated as an exaggeration, regarding humour, and based on viewpoint as the main underlying construal mechanism.

(73) Amy Schumer: They talk to me as if I need to be buttered to fit through a door.



They talk to me like I need to be...



...battered to fit through...



...a door.

Figure 69. Turns in example (73)

Amy Schumer performs two turns during her remark. First, upon uttering 'battered', she turns to make eye contact with the host. I interpret this turn as seeking confirmation of the hearer's attention, as well as underlining the word 'battered' as the one of the most relevant elements in the utterance. Then, she turns her head again upon the pronunciation of 'door'. Once again, there is a correlation with prosody in terms of what words are most prominent, which, at the same time, I believe is linked to information structure in the sentence, highlighting the most meaningful words, i.e., 'battered', and 'door'. Furthermore, taking the latter turn along with gaze and hand gestures, I posit that it is pointing to the abstract space where the door is. As a matter of fact, she refers to the door shortly afterwards, in an utterance not taken as humorous, and she turns again to that very same space when pronouncing the word 'door'.

Let's now consider example (74), in which different gestures co-occur, among which repeated shakes and a turn. Stephen Colbert has said to Elon Musk that given that he's trying to do great things, he is a billionaire, and he is involved in the development of many state-of-the art technologies, he has to be either a

superhero or a supervillain, so that he has to choose one or the other. Elon Musk replies:

(74) Elon Musk: I try to do useful things.



I try to do...



...useful things.

Figure 70. Shake and turn in example (74)

He precedes his statement with repeated head shakes that are maintained throughout, which could be interpreted as negating the idea that he is a superhero or a supervillain. Indeed, this utterance has been categorised in the sample as an example of understatement, as Elon Musk is trying to downplay the importance of what he does with regards to the idea of superhero saving the world, as portrayed by the host. In addition, Elon Musk performs a turn at the end of the statement, making eye contact with his interlocutor, while simultaneously

instantiating a nod and a tilt. I believe those movements have the function of signalling the end of his remark and yielding the turn in the conversation. The nod could also be seeking approval from the interlocutor, while the tilt could also be associated to a further expression of denial or disbelief with regards to the previous host's words.

I posit that these example further demonstrate the complexity surrounding the performance of head movements, whose analysis can certainly yield different interpretations as to what functions can be assigned and what role they are playing in any given instance of communication. A comprehensive study of such roles must therefore also necessarily include the point of view of interlocutors in the communicative event. Table 10 summarises the functions of gestures identified in the sample:

Gesture	Function identified in sample
nod	Agreement, underliner, backchannel request, confirmation
tilt	Gestural triggers in sarcasm and anecdote, expression of disbelief, information structure, rejection
raised eyebrows	Information structure, underliner, gestural trigger in irony and hyper-understanding
turn	Change in narrator, change in viewpoint, turn yielding
shake	Rejection, inclusivity
smile	Frame turn as humorous or ironical, face-saving, affective display

Table 10. Functions of face and head movements identified in sample

6.4 Multimodal density: a prospective test

In light of the analysis above, it seems clear that none of the gestures co-occurring with humour that have been identified in the sample can be considered markers of humour, as they do not consistently co-occur with humorous statements and do not help to predict humour. Furthermore, as seen in Chapter 5 (cf. Section 5.5), humorous instances in the sample are not marked by prosodic contrast either. Therefore, if we admit that humour is a marked form of communication, requiring more cognitive effort by the hearer to be processed and understood, the question remains as to whether humour is signalled multimodally at all, in order to facilitate its apprehension.

After failing to find head movements or face expressions that could be considered as markers, and having ruled out prosodic contrast as a way to underline humour, one may wonder if the key perhaps lies in the amount and frequency of gestures produced during humorous utterances, as opposed to non-humorous ones. In the same vein of studies that have shown that the context in which communication occurs, and the participants in the communicative event exert an influence in the amount of multimodal cues we produce, e.g. nodding more with less familiar interlocutors, or repressing certain face gestures when confronted with authoritative figures (Ishi et al., 2014), maybe the presence of humour in a statement will lead to more gestures accompanying and punctuating that utterance.

Going back to the notion of current discourse space (CDS), let's remember that the viewing frame encompasses the current scope of conceptualisation in any instance of communication, i.e., a usage event. Within the viewing frame we

find the focus, that is, the profiled element, as well as the channels corresponding to the two poles entailed in any utterance: conceptualisation, and vocalisation. Langacker (2001), therefore, placed the multimodal part of the utterance in that vocalisation pole, in which prosody, intonation, and the verbal message interplay on the basis of how the event is being conceptualised. The strongest link between conceptualisation and the multimodal expression of a given event lies in the window frame. Therefore, it might be argued that the biggest impact on the way an utterance is constructed multimodally is precisely how the event is being construed. This is in line with the notion of growth point (McNeill, 2013), in which the conceptualisation processed itself is twofold, with an analytic (language) and an audiovisual (multimodal) component.

The second biggest influence on the multimodal production of an utterance would be the ground within the CDS, including interlocutors, the interaction, and the immediate circumstances, followed by the larger context of speech, which includes more distant social and cultural circumstances, and the shared knowledge, as a set that is easily accessible but not explicit in the usage event.

Arguably, in the interviews included in the sample the shared knowledge, context and ground remain stable, as the interlocutors and the audience is the same, as well as the assumptions presumably shared on the genre of the programme, expectations of amusement, knowledge about the interviewees, etc. As a result, I posit that higher multimodal density in humorous utterances in a given interview could be claimed to arise from the humorous or marked nature of the utterance itself. In order to test this hypothesis, I decided to conduct a prospective comparative test between humorous and non-humorous utterances,

using the interviews with the highest and the lowest multimodal density in humorous utterances respectively. Multimodal density in the present study –not to be mistaken with modal density (Norris, 2014)– will be measured by estimating the number of occurrences of head and face movements annotated in the relevant utterances, i.e. humorous or non-humorous, and divided by the number of utterances of the particular type involved.

A prospective brief study was conducted to test this hypothesis, namely, that humorous utterances will be characterised by higher multimodal density. To do so, two interviews were selected, with the highest and lowest multimodal density respectively. John McWhorter was the person producing the highest number of gestures per humorous utterance, with an average of 5.85 gestures. On the opposite side, Susan Sarandon only produced 2.75 gestures per humorous utterance identified. Previously selected baseline and pre-baseline utterances (cf. Chapter 5, Section 5.5) as non-humorous utterances were analysed, with head movements and face gestures performed during those utterances annotated. The result for John McWhorter yielded an average of 5.1 gestures per non-humorous utterance, while Susan Sarandon performed 2 gestures per non-humorous utterance. The amount of gestures produced, then, did not vary significantly, and we cannot conclude that humorous utterances are produced with higher multimodal density; therefore, the starting hypothesis is rejected.

Having said that, many caveats can be raised with regards to the approach taken in this study. To begin with, the size of the sample as a whole, and of the two interviews in which the hypothesis was tested in particular, is not large

enough to be able to extrapolate and generalise the findings. Furthermore, a major challenge is linked to the analysis of the sample itself. First, the segmentation of utterances was made applying a usage-based approach, but it is clear that different criteria can be implemented to decide how to delimit utterances. A different segmentation may well yield different results.

Second, the study is rather qualitative, insofar as the technical means to achieve more accuracy in annotating gestures are not available, e.g. eye tracking, full view of the speaker at all times, etc. The decision to favour the study of spontaneous communication involves that no control over the camera shot exists, which results in available images varying in size, angle, etc. This, in turn, poses a big challenge as far as the annotation of gestures is concerned.

A further non-negligible issue in this regard has to do with the difficulty in always clearly and unequivocally identifying gestures, not only due to the different shots available, but also because movements very often overlap. Hence, it is easy to mistake certain movements such as tilts, nods and turns or shakes, when they overlap, or in cases of short, quick shifts. In addition, no account of the amplitude or rate in gestures has been taken, whereas differences may lie in the various ways in which a movement such as head shakes, for example, can be instantiated (Wagner et al., 2014).

Yet another element which may bear a huge influence in the interpretation of gestures is that of the interplay of all modes recruited in a single instance, i.e., mode density (Norris, 2004a; 2014). In this regard, our study is partial, as modalities such as hand-gestures, body-posture shifts, gaze, etc. have been left

out due to sample constraints —the hands were not always visible, for example— and the lack of technical means to provide more accuracy, as with eye-tracking technology to study gaze.

Finally, a more fundamental question remains, as to where the humorous nature of an utterance lies. Our study has only be concerned with speakers, and, following standard practice in the literature, humour was considered to have occurred when an utterance lead to an outburst of laughter in the audience. But certain remarks that elicited laughter did not seem to have been uttered with a humorous intent by the speaker —although, admittedly, this is a subjective interpretation which could only be confirmed by the speaker—, as Susan Sarandon’s comment on the ceiling (cf. example 28, Ch. 4). As a result, any study aiming at gaining an insight into how humour is multimodally framed has to consider the humorous intent in the speaker, which adds a further layer of complexity in the analysis.

At the same time, the fact that a certain utterance may be perceived as humorous even in the absence of such an intention on the part of the speaker has to be also acknowledged. Consequently, my research has led me to realise that to fully understand what makes an utterance humorous, the hearer’s perspective must be taken into account too. But the question of where humour lies is still open. My claim is that it can be tied either to the speaker’s intention or to the hearer’s perception, which may not always overlap, as in the case of failed humour (Hay, 2001), or when humour is unwittingly produced by speakers.

In spite of the objections that I have just listed, I believe that this study also has clear merits. It confirms previous studies on the meaning of gestures, which seem to carry the same pragmatic and semantic functions in humorous and non-humorous communication. It reaffirms the importance of conducting studies on spontaneous communication, as the production of multimodal cues is heavily influenced by the participants and context of communication, and, as a result, it may differ considerably in staged settings trying to reach out to a wide audience. It bears out previous findings pointing to the non-existence of multimodal markers of humour. Furthermore, it seems to refute the hypothesis that higher multimodal density is a feature of humorous communication —as far as the sample is concerned—. In sum, no multimodal behaviour specific to humour that sets it apart from non-humorous communication has been found. Instead, speakers just recruit and elicit any modality at their disposal, fully exploiting related affordances, to try to convey their message in the most effective manner, regardless of whether that message happens to be humorous or not.

6.5 Conclusion

In this chapter, I have outlined the results of my study of face expressions and head movements in humorous utterances, as recurrent multimodal cues in every interaction, gaining an insight into the functions that can be assigned to raised eyebrows, smiles, nods, tilts, shakes, and turns as facilitators of communication. The results of the analysis conducted bear out the outcome of previous studies on these gestures in the literature (Poggi & Pelachaud, 1998; McClave, 2000; Kendon, 2002; Lee & Marsella, 2010; Attardo et al., 2013; Tabacaru, 2014; etc.). These gestures are not specific to humorous

communication; they are rather used as multimodal tools by speakers and hearers in communication at large.

In addition, no consistent pattern of association between certain types of humour or construal mechanisms and any of the particular gestures analysed has been found, with the exception of the highest number of shakes clustering around instances of exaggeration, and most tilts concentrated around occurrences of viewpoint. Further research will be necessary to determine whether these are just casual links or not. Hence, given that these gestures cannot be considered as humorous multimodal cues *per se*, the question arises as to whether there exist any multimodal tool that is consistently used to facilitate the interpretation of humorous utterances.

As seen before, context plays a major role in determining the way we communicate, with a clear impact on what gestures we perform, their frequency, and intensity (Ishi et al., 2014; Wagner et al., 2014). For example, the less common ground and familiarity shared by interlocutors the higher need to mark messages multimodally (Ishi et al., 2014). In addition, when only aural communication occurs, prosodic cues are more salient (Cvejic, Kim, & Davis, 2012). In other words, the current discourse space (Langacker, 2001) and all the elements therein will have an impact on the multimodal cues produced in interaction. Furthermore, modal density, defined as “the intricate interplay of various modes of communication or the intensity of a certain mode that a social actor employs” (Norris, 2004b, p. 102), is higher in communicative actions pertaining the foreground in a given context, i.e., where speakers place their focus of attention and have greater awareness of the actions performed (ibid.).

Norris emphasises that it is not just about counting how many modes interplay, but to disentangle the extent to which they intertwine. The intensity and frequency of gestures is higher when co-occurring with prominent parts of the speech (Ambrazaitis & Bruce, 2009; Ferré, 2010). The analysis conducted in this chapter proves that not only the interplay between different modalities, but also the overlap of co-occurring gestures, even from the same modality, will have an impact on how an utterance may be interpreted.

In light of the above, it is clear that gestures cannot be just taken as elements merely punctuating speech, but as part of a “*multimodal* construction in which the different modalities of expression available are *deployed* by the speaker in the course of building a unit of expression according to the rhetorical needs of the interactive moment” [emphasis in the original] (Kendon, 2002, p.147).

Chapter 7

A cognitive analysis of humour

7.1 Introduction

This chapter will offer an analysis of humour from a cognitive linguistic perspective, on the basis of the instances found in the sample. An in-depth study of the most frequent construal mechanisms and types of humour identified is included. In addition, examples from the types of humour with the highest number of occurrences in the sample are accounted for drawing on widespread cognitive linguistic notions, such as current discourse space (CDS) (Langacker, 2001; Antonopoulou, Nikiforidou, & Tsakona, 2015), as well as blending (Fauconnier & Turner, 2002; 2003; Coulson, 2005a; 2005b; Coulson & Oakley, 2005), conceptual integration network (Fauconnier & Turner, 1998), and mental spaces (Fauconnier, 1994; 1997).

Cognitive linguistics offers the best framework for a comprehensive study of humour (Brône & Feyaerts, 2003; Brône, Feyaerts, & Veale, 2006; Veale et al., 2015; Dynel, 2018). It does away with artificially drawn boundaries between syntax, semantics, and pragmatics in order to account for conceptualisation at all levels of language structure (Brône et al., 2006), acknowledging the key role of discourse and embodiment in meaning construction (Langacker, 2001). It is an umbrella paradigm in which different approaches to humour can converge and be reconciled. Hence, my analysis draws on basic notions in cognitive linguistics and will subsequently bring together other approaches (Clark, 1996; Attardo, 2001; Coulson, 2005b; 2005c, 2015; Brône, 2008; Veale, 2015; Giora, 1998;

McGraw & Warren, 2010; Giora, Givoni, & Fein, 2015; Giora, Jaffe, Becker, & Fein, 2018; Gibbs, Samermit, & Karzmark, 2018), with a more or less evident cognitive stance, into that basic account. I also assess compatibilities among different terms advocated by those various authors to try to look for overlaps and convergence that may ultimately show the way towards a possible unified approach.

Furthermore, I review the notion of incongruity in relation to humour, and go through the vast array of related, albeit different, interpretations posited by authors as one of the most fundamental elements of humour (Clark, 1996; Giora, 1991, 1998; Attardo, 2001; Brône & Feyaerts, 2003; Coulson, 2005b; 2005c, 2015; Veale, 2015, etc.). Finally, I delve into various explanations as to what makes something funny (Morreal, 1983, 2009a, 2009b; Forabosco, 2008; McGraw & Warren, 2010; Veale, 2015; Giora et al., 2017; Gibbs et al., 2018, etc.).

7.2 Construal mechanisms identified in sample

A major tenet in cognitive semantics is that concepts are not isolated units in the mind. They can only be understood and processed in a context of presupposed, background knowledge structures, known as domains (Clausner & Croft, 1999). Concepts, therefore, are grounded or anchored in domains, and both are conceived through our experience, as well as our bodily and sensorial perception of the world, which can be construed in different ways. The mental processes whereby our experience of the world is thus conceptualised are called construal mechanisms (Croft & Cruse, 2004).

7.2.1 Construal mechanisms and humour

Croft and Cruse (2004) compiled a list of construal mechanisms, bringing together previous classifications (Langacker, 1987, 1991; Talmy, 2000a, 2000b) and unifying terminology. They established four broad categories of construal mechanisms: a) attention/salience, b) judgment/comparison, c) perspective/situatedness, d) constitution/gelstat (cf. section 4.2.4 in Chapter 4). Brône and Feyaerts (2003) singled out some of these categories and certain construals therein as the most relevant to humour: profiling and metonymy as selection mechanisms under the category of attention/salience; metaphor and figure/ground under judgement/comparison, and viewpoint.

Selection is the ability to focus only on what is considered to be the most relevant part of the linguistic input in a given instance of communication. In doing so, certain elements of a domain, frame or mental space may be highlighted, i.e. profiled, or a salient point can be used to refer to a different element within the same domain (Brône & Feyaerts, 2003). Certain contextual factors can determine the degree of saliency of elements, or, in the case of jokes, of different interpretations of humour (Giora, 1991, 1998; Giora, Givoni, & Fein, 2015; Giora et al., 2018).

Judgment/comparison operations are pervasive and constitute one of the most basic categories of construal mechanisms (Langacker, 1987, pp. 103-105). The notion of figure-ground reversal has been widely exploited to account humour (Attardo & Raskin, 1991; Attardo, 2001; Veale et al., 2006; Brône, 2008; Tabacaru & Lemmens, 2014). Furthermore, the notion of mapping, central in

conceptual metaphor (Lakoff & Johnson, 1980; Lakoff, 1993; Ruiz de Mendoza, 2014; Forceville, 2015) is a central piece of incongruity resolution, as it is what allows to establish a similarity between opposed scripts (Raskin, 1985; Attardo, 2001; Attardo, 2015).

Regarding viewpoint, a different perspective on a given communicative event automatically yields a different conceptualisation of such event (Langacker, 1987; Ritchie, 2006; Feyaerts et al., 2010). Viewpoint and mapping are two central elements of conceptual integration (Fauconnier & Turner, 1998, 2002, 2003) and mental space theory (Fauconnier 1994; 1997), hence the suitability to apply such an analysis to humorous instances (Coulson, 2005a, 2005b; Dore, 2015). Furthermore, blending can account for creativity as upon merging input spaces, the resulting blended space may well have new emergent structure not present in any of the inputs (Fauconnier & Turner, 2002).

The results of the present study bear out Brône and Feyaerts's (2003) analysis, as the most frequent construal mechanisms identified underlying humour instances in the sample are viewpoint, profiling, and subjectivity. Both subjectivity and viewpoint fall under the perspective category and are clearly related. Croft and Cruse (2004) defined subjectivity as the conceptualisation of an event in which speakers includes themselves (cf. section 4.2.4 in Chapter 4), i.e. they position themselves at the centre of the speech event. Both viewpoint and subjectivity, as all construals based on perspective, are built upon the particular way we are in the world, both spatiotemporally and conceptually. Our standpoint will determine our scope of attention, i.e. the viewing frame within a current discourse space (Langacker, 2001). It is within that viewing frame that we

will focus our attention on a salient element, and that we can direct our interlocutor's attention to the particular element we want to profile, i.e. to make more salient.

If we think of the characteristics of the sample under analysis, which consists entirely of interviews where only the interviewee's speech has been studied, it is understandable that perspective, all construal operations therein included, is the most frequent category of construal mechanisms found in the sample. In the end, guests are there to talk about themselves, and it is only natural that they base their construal of the communicative event on their own personal perspective.

The purpose of this thesis is to analyse humour, the reason why only humorous utterances have been analysed. Nevertheless, it would be interesting to further study a set of non-humorous utterances within the sample to determine whether underlying construal mechanisms also fall mainly under the perspective category. I do not believe there are specific construal mechanisms associated with humour (Veale et al., 2015). I rather claim that the pragmatic and discourse context of the communicative event will ultimately determine the nature of the conceptualisation operations at hand (Attardo, 2001; Yus, 2003; Coulson & Oakley, 2005).

Furthermore, leaving aside canned jokes, i.e., jokes which have been "used before the time of the utterance in a form similar to that used by the speaker" (Attardo, 1994, p. 295), the context in which a conversation occurs plays a major role in the perception of a given utterance as humorous.

Let's consider example (75) below:

(75) Daniel Kaluuya: White people say very weird stuff to you.

This utterance was categorised as humorous in the sample as the audience laughed. It was part of the answer to a question put by Stephen Colbert on whether black people in America and Britain can be said to have similar experiences with regards to their racial background. I find that it is a clear example in which the humorous nature of the statement cannot be said to lie in the decontextualised semantic content of the utterance itself. One could easily imagine a more formal context, such as a conference on racial discrimination, for example, in which this very same utterance would not have been perceived as funny. However, the same underlying construal mechanism would be in place. So humour, in this case, does not hinge upon either the semantic content, or the conceptualisation process. It could be argued that the humorous nature of the utterance resides instead in the context in which it has been said, namely, a casual late-night show interview in which a light tone is expected and in which participants —audience, host and guests alike— are prone to humour and laughter.

7.3 Cognitive analysis of humorous instances

7.3.1 Current discourse space

Cognitive linguistics considers that all linguistic units are abstractions from actual instances of language use, i.e. usage events (Langacker, 2001). Usage events entail comprehensive conceptualisation of the event, articulated around two poles: conceptualisation itself, with full contextual apprehension by

participants, plus a vocalisation pole. Interlocutors' complete understanding of the circumstances surrounding their interaction, i.e. the discourse they are engaged in (Langacker, 2001, p.144), is part of their conceptualisation of the usage event. As seen in Chapter 4, Langacker defined the CDS as the “mental space comprising those elements and relations construed as being shared by the speaker and hearer as a basis for communication at a given moment in the flow of discourse” (2001, p. 144). The ground, in turn, comprises the speech event, the interlocutors, their interaction and the immediate circumstances, mainly time and place of speech (Fig. 71).

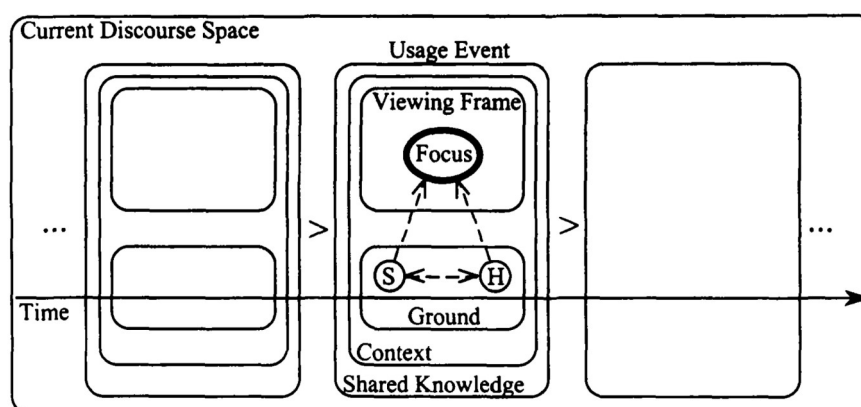


Figure 71. Current Discourse Space (Langacker, 2001, p. 145)

Regardless of the active or reactive role of speaker (S) and hearer (H), both have to process utterances on the basis of the double conceptualisation-vocalisation pole. As they communicate, they will direct their focus of attention within a given viewing frame or scope of attention, determined by the CDS as a whole and delimited by “how much we can conceptualise or hold in mind at any

given instant” (ibid. p. 144). In successful communication, both S and H will coordinate their actions to focus on the same conceived entity, which is the profiled element within the viewing frame. Both the viewing frame and the CDS, however, are dynamic and flexible. The viewing frame can be directed anywhere, to real or invented situations. The focus of attention can also change within the viewing frame.

Consecutive usage events conform the CDS, drawing on previous utterances and setting expectations for subsequent ones. Crucially, each usage event can entail changes in the CDS, to elicit a different context, bringing a backgrounded element to the fore of the viewing frame or drawing on shared knowledge to establish a new viewing frame. Notice that the ground in CDS refers to the immediate circumstances surrounding the interaction that is taking place, whereas the context of speech includes physical, mental, social and cultural elements. Finally, the shared knowledge is “the body of knowledge presumed to be shared” by both interlocutors and “reasonably accessible” (ibid., p. 145).

As explained before, the viewing frame is determined by our conceptualisation field, or scope of attention at any given time. Fig. 71 clearly outlines the conceptual and vocalisation pole of every utterance.

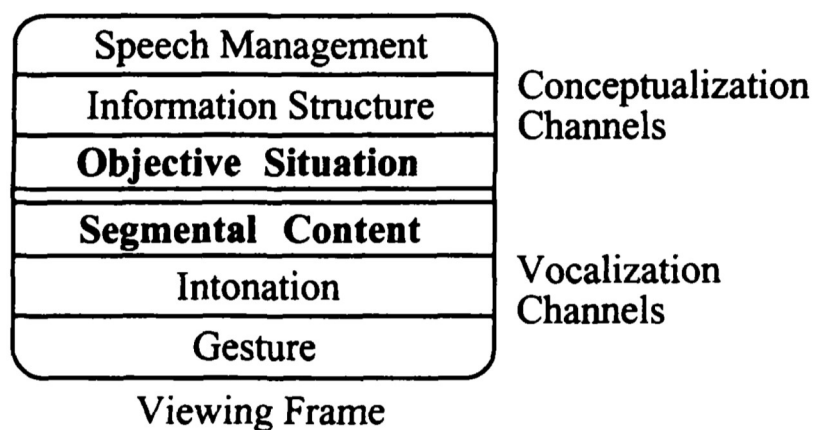


Figure 72. Viewing frame (Langacker, 2001, p. 146)

The conceptual end includes three channels, namely speech management, related to turn taking; the information structure channel, in which old and new information are processed and weighted, and the objective situation, which is the main object of conceptualisation. This is the most salient part of the conceptualisation pole, and belongs to the focus of attention. This focus of attention also comprises a salient part of the vocalisation pole, which is the segmental content, i.e., what we are actually saying. The other two channels of this vocalisation pole are intonation and gestures, two additional modes in face-to-face communication.

I have chosen Langacker's (2001) CDS as the basic framework for analysing humorous instances as it elegantly and simply account for all elements relevant to communication. By claiming that the CDS is the basis for conceptualising any linguistic unit, Langacker underlines the roots of language in discourse. Language units, therefore, regardless of their structure level, (semantic, lexical, construction, utterance, multimodal utterance, etc.) share a

grounding process not only in embodiment, understood as our perceptual and interactive experience of the world, but also in discourse and human interaction, as the most basic form of communication.

Furthermore, both the cognitive and multimodal side of any communicative event are acknowledged in the CDS, through the two poles inherent to any usage event, i.e. conceptualisation and vocalisation. Finally, the crucial role played by the pragmatic context and discourse in general is also highlighted, comprising the immediate circumstances around the communicative event, cultural and social contexts and conventions, and mutually assumed and accessible shared knowledge. In my view, the current discourse space provides a comprehensive yet simple framework to analyse any instance of communication.

My aim with this analysis is twofold: on the one side, to account for humour drawing on basic concepts in cognitive linguistics; on the other side, to reinforce the explanatory power of such a framework, demonstrating its validity to explain highly marked forms of communication such as humour. I do not believe that humour needs to be explained resorting to specific humorous communication mechanisms (Veale et al., 2015; Dynel, 2018). I think any account of communication must be able to explain humour too. The challenge, then, is to pinpoint where humour lies within those mechanisms in order to establish what exactly renders an instance of communication humorous and what makes it different from a non-humorous utterance that can be explained drawing on the same theoretical framework and principles.

7.3.2 Construal mechanisms and humour types

What follows is an account of certain examples extracted from my sample, taking into account the most frequent construal mechanisms and humour types found. The three types of humour with the highest number of occurrences in the sample are anecdote (33 instances; 26.83 %), irony (17 instances; 13.82 %), and parody (16 instances; 13.00%). As for construal mechanisms, viewpoint has been identified in 46 utterances (38.01 %); profiling underlies 23 humorous instances (19.00 %), and subjectivity has been found in 14 instances (11.57%).

Table 11 summarises the construal mechanisms at play in each of the three most frequent types of humour encountered. Notice that the number of construal mechanisms identified may be higher than the total number of occurrences in each humour type. The reason is that, although a minimalist annotation approach has been adopted (cf. section 5.6.1 in Chapter 5), at times two construal mechanisms were deemed clearly salient in a single humorous instance, in which case both types were annotated.

Anecdote	Irony	Parody
Viewpoint (22)	Profiling (6)	Viewpoint (10)
Profiling (6)	Viewpoint (3)	Deixis (4)
Subjectivity (5)	Subjectivity (2)	Categorisation (1)
Dynamic attention (4)	Metaphor (2)	Dynamic attention (1)

Deixis (3)	Deixis (2)
Categorisation (1)	Categorisation (1)
Metaphor (1)	Dynamic attention(1)

Table 11: Construal mechanisms underlying most frequent humour types

I then proceed to analyse an example of each of the three most common types of humour-construal combination, namely: an anecdote based on viewpoint; an instance of irony with profiling as an underlying construal operation, and an example of parody relying on viewpoint.

7.3.2.1 Anecdote and viewpoint

It is perhaps not surprising that anecdote and viewpoint go hand in hand in such a high number of occurrences. Anecdotes are stories based on personal experience or on other people's lives (Dyner, 2009), so it is understandable that they be construed from the standpoint of the speaker.

Let's consider example (76) from the sample. Cristela Alonzo is talking about her childhood. She spent a lot of time at home watching TV, mostly comedy shows, and that is how she became fascinated by stand-up comedy. She explains how her mother, a Mexican immigrant, hardly knew any English and therefore let her watch comedy shows not intended for children, as she was unable to recognise inappropriate content in them. On the contrary, when she saw her daughter amused and she recognised comedians from the Saturday Night Live

programme, which she used to watch with her children, thought that the TV programme was fully appropriate for children.

(76) Cristela Alonzo: *The Specials* came out and she's like "Ay, it's the guy from *SNL*; you can watch this" [laughter in the audience; pause]. And then he starts talking and I'm like "Oh, damn. I can watch this" [laughter in the audience].

This example has been annotated as two distinct humorous utterances, as there is a pause and two bouts of laughter in the audience as a response. Both instances have been identified as anecdotes, and both with viewpoint as the main underlying construal mechanism. Clearly, she is telling a personal story and she brings in two different viewpoints, namely her mother's and Cristela Alonzo's own perspective. In terms of Langacker's (2001) CDS, we could then identify two usage events, each corresponding to one of the humorous instances identified:

UE1: *The Specials* came out and she's like "Ay, it's the guy from *SNL*; you can watch this".

UE2: And then he starts talking and I'm like "Oh, damn. I can watch this".

I will sketch the main elements that, in my view, need to be present in this CDS for communication to be successful. I do not claim it is a comprehensive list including every single one of the elements in each of the components of the CDS. There is no way around the fact that individual construals of this situation may yield somewhat different elements. I will therefore strive to list those items which

I believe form the bare minimum to ensure successful communication and a shared apprehension of the speech event.

The CDS includes both these usage events, plus the previous ones, such as the host's question, all of them setting expectations about what is going to be said next. The shared knowledge in UE1 is the same as in UE2. Shared knowledge contains what comedy and comedians are, what kind of programmes are suitable for children, the role of parents in monitoring what their children watch on TV, the situation of Mexican migrants not proficient in English, what *SNL* and *The Specials* are, and, crucially, what difference in tone exist between both programmes, even when having the same comedian appearing in both, etc. The context, regarding the cultural, mental, social and physical circumstances, comprises the figure of Cristela Alonzo as the first successful mainstream Latino comedian, her role as a comedian, and the need to come across as funny, therefore setting expectations about the likelihood of humour arising in her speech.

The ground, defined as the immediate circumstances surrounding the speech event, contains Cristela Alonzo herself, Stephen Colbert as the host, and the audience, as well as the genre of the programme and the setting of the interview, which is meant to be amusing and entertaining. Finally, the viewing frame, as directed by Cristela Alonzo's speech, comprises her past as a child of a Mexican immigrant mother, having the TV as her main companion—she had actually mentioned that in a previous statement during the interview—, and her watching *The Specials* with the same comedian from SNL, a programme her family used to watch together.

All of the above-listed elements are shared both by UE1 and UE2. However, they differ in the focus of attention. In UE1, the focus of attention is placed on *The Specials* and on the comedian from the point of view of Cristela Alonzo's mother, who allows her daughter to watch the programme as she does not understand what they say and thinks that the show was innocuous for children, having recognised the comedian from SNL, a family-friendly programme the family watched together on Saturdays. In the next usage event, UE2, however, the focus of attention is again on *The Specials*, but this time it is not the comedian that is profiled, but the words he is saying and the exhilaration felt by Cristela Alonzo when she is allowed to watch the programme.

A more fine-grained analysis of the viewing frame can establish that the focus of attention in UE1 is composed of the mother's point of view on *The Specials* and the comedian as the objective situation, as well as the actual words and literal quotation from her mother uttered by Cristela. In UE2, the objective situation is in turn her own point of view and feelings when she realises how inappropriate the programme is for children, and how good she feels she can watch it. Again, in the vocalisation pole, her words, now quoting herself, form the segmental content.

Finally, as for the other two conceptualisation channels, she makes a pause after UE1, even though she keeps the floor, to allow the audience to laugh as an acknowledgement of the incongruity of what she is telling. Information structure is organised in terms of delivering her mother's point of view first, in order to contrast it later with her own view of the situation at the time. In UE2 she actually conveys that latter perspective and then she keeps the floor to explain

that the fact that she watched so much TV comedy as a child had a deep impact on her desire to become a comedian herself.

Lastly, on the vocalisation pole, both intonation and gestures are used to reinforce the different viewpoints from which the situation is construed. She acts out her mother's and her own words at the time to make the message even clearer. Body posture and intonation also mark the passage from one usage event to the next.

The previous analysis shows the validity of Langacker's CDS model to explain how communication unfolds. It does not, though, account for the humorous nature of the utterances themselves. Humour can be said to be conveyed when it is acknowledged by participants in the communication event (Hay, 2001). In this case, the bar to measure whether the utterance has been humorous lies in the reaction by the audience. If they laugh, one can conclude that the utterance is perceived as humorous and has therefore been successful from that point of view (Morreal, 1983; Attardo, Pickering, & Baker, 2011; Bryant & Gibbs, 2015).

With regards to the question of what confers to UE1 and UE2 their humorous nature, drawing on the idea of incongruity, or the notion of contrast of ideas or of expectations, as advocated by incongruity-resolution theories, I argue that humour lies in the following elements. In UE1, there is a clear clash between what parents are expected to do from a socially acceptable perspective, i.e. prevent their children from accessing inappropriate content for their age, and the fact that Cristela Alonzo's mother explicitly gives her permission to watch that

inappropriate show as she recognises the comedian from SNL. Hence, the contrast arises between an element from the shared knowledge (socially acceptable parenting) and an element from the focus of attention (the mother allowing Cristela to watch a programme with inappropriate language). In UE2, Cristela's own point of view is confronted with her mother's, highlighting the mismatch between how they both construed the situation.

7.3.2.2 Irony and profiling

Irony is understood as a type of humour for the purposes of this analysis, although it is a controversial issue (Attardo, 2000a, 2000b; Dynel, 2009; etc.). Irony could be defined as a violation of norms and expectations in a pretence space (Giora, 1998) in which the speaker expresses detachment from the utterance said. In irony not only the utterance itself, but also the evaluation of the speaker on that same utterance is communicated. Most examples of irony in the sample occur with profiling as the main construal mechanism; therefore, one such example will be analysed resorting to the notion of current discourse space (Langacker, 2001).

Profiling is the construal operation whereby only a certain element of the situation construed is made salient. Consider example (77). Susan Sarandon is talking about Bette Davis, as she impersonates her in one of her latest films. She is discussing the highly confrontational relationship between Bette Davis and Joan Crawford. Among the reasons for such enmity, she mentioned that it was partly instigated by a famous columnist, who wrote about their confrontation in

her column, and also the fact that both women had had many sexual partners in common.

(77) Stephen Colbert: Joan Crawford and Bette Davis?

Susan Sarandon: Yeah.

Stephen Colbert: Not at the same time, right?

Susan Sarandon: I don't think so. That would have been, maybe, a bonding experience.

In this case, the humorous utterance identified in the sample is Susan Sarandon's last remark. Susan Sarandon's utterance is considered ironic as there is a clear difference between the intended and literal meaning of the utterance. The possibility that they could have had sex with the same partner simultaneously is very remote, given the seemingly protracted and intractable conflict between both women, not to mention the taboo of such sex practices at the time, i.e. the 30's. In spite of that, Susan Sarandon decides to play on the host's question and present the hypothetical —pretence— scenario, where having sex with the same partner simultaneously could have been a possible recipe to solve the confrontation between Bette Davis and Joan Crawford.

The utterance identified as humorous in this instance is Susan Sarandon's last remark: "I don't think so. That would have been, maybe, a bonding experience". In this case, the CDS contains her previous statements about the antagonism between both actresses, Stephen Colbert's remarks, and her own ironic comment. The shared knowledge includes the public figures of Bette Davis and Joan Crawford, their work as actresses, morality at the time, sex taboos, etc.

As for the context, it comprises the apprehension of Susan Sarandon as a famous actress, and her role as Bette Davis in the film, which allegedly has made her thoroughly study the diva and become an expert on Bette Davis' life.

In the ground, in turn, we find the expectations linked to the programme genre and the informal tone of the interview, as well as the interaction itself between Susan Sarandon and Stephen Colbert, with the host's remarks prompting the ironic response in Susan Sarandon, expected to live up to the witticism and humour in her counterpart's comments. The viewing frame in this particular usage event is the bitter relationship between Joan Crawford and Bette Davis, their life in general and their sex life in particular. Stephen Colbert's question had shifted the focus of attention to a pretence space in which both women had sex with the same partner at the same time. Susan Sarandon's remark zooms into the same focus of attention to profile therein the bonding character that experience could have had for the two actresses, possibly helping to mitigate the bad feelings between them.

In other words, Susan Sarandon keeps the objective situation —both women having sex with the same partner at the same time— that Stephen Colbert has previously established as the conceptualisation pole of the focus of attention, but she directs the attention to one particular element she sees in that objective situation, i.e. the bonding character the experience might have had for Bette Davis and Joan Crawford. Segmental content is, as always, made of the words uttered. The other elements in the vocalisation pole, namely intonation and gestures, do not starkly differ in this case from those accompanying previous remarks, apart from directing her gaze to the host at the end of the comment to

signal end of turn. This takes us to the remaining components of the conceptualisation pole, i.e. speech management—she yields the turn at the end of her statement—, and information structure, where she introduces her view that such an experience could possibly have brought the two actresses closer together.

In my view, humour in this case lies in a contrast within the viewing frame, in which the salient element in the immediately previous usage events had been the animosity between both divas. The focus placed by Stephen Colbert on a pretence space in which both actresses were having sex with the same partner at the same time had already elicited laughter in the audience, possibly because of the incongruous nature of the situation. Susan Sarandon's retort results to be even more humorous—provided we take the higher intensity of laughter in the audience as a valid measure—. I argue that the reason is that, instead of being taken aback by the host's question, she is able to exploit it and delve into the humorous nature of that comment highlighting, i.e. profiling, an even more unexpected feature of that already unlikely event.

Although I believe Susan Sarandon's utterance is fundamentally ironic, it could also be defined as hyperunderstanding (Brône, 2008), in the sense that it exploits what has been said by the previous speaker changing their intended meaning. Although there is no confrontation nor verbal feud of any kind between Susan Sarandon and Stephen Colbert, i.e. neither of them is the target of the other's humour, Susan Sarandon succeeds at reversing the host's question to add an acute and unexpected perspective.

As mentioned in Chapter 5, boundaries between humour types are not always clear-cut and several categories of humour may converge in a single instance. Something similar occurs with construal mechanisms. In this example, it can also be argued that viewpoint plays a major role when it comes to construing and presenting the situation. Susan Sarandon is able to profile the bonding experience as a salient element of the viewing frame because she has conceptualised the situation from her own standpoint, which differs slightly from the one entailed in Stephen Colbert's remarks. As explained before, I have opted to annotate only what I believe are the most salient types of humour and construal mechanisms in each humorous instance, for the sake of simplicity in the analysis and in order to gain an insight into possible patterns emerging between humour types and construal mechanisms.

7.3.2.3 Parody and viewpoint

Parody implies the playful imitation of someone in a given situation, exaggerating certain elements in order to produce a humorous effect. Therefore, the link between parody and viewpoint is unavoidable, as by definition, speakers clearly convey their own construal of the person and the situation subjected to parody. I argue that profiling also plays a major role, as in order to make the parody, the speaker necessarily needs to choose certain traits and features to be exaggerated. However, as I see it, viewpoint is more prominent than profiling, as the choice of features to be made salient to the audience through parody are subordinate to conveying the imitated person and situation from the speaker's standpoint. Not only that, but in parody, the speaker seems to assume the parodied character viewpoint as they act out that person words or behaviour,

when in fact, what is truly conveyed is how the speaker construes the person and situation parodied.

Let's examine example (78), extracted from Condola Rashad's interview. She is known by her role in the TV series *Billions*, where she impersonates a successful and ambitious assistant to the district attorney. When asked about how she prepares for the role and whether there is a big amount of research on legal issues involved, she refers to her sister, an attorney working for the Mayor in Atlanta, as her main source of information on those matters. The subsequent exchange is as follows:

(78) Stephen Colbert: Does she ever get frustrated because you have the job that...?

Condola Rashad: Oh, totally. She was like, "Why do you get the job that I wanted? But you're playing it and I'm the real one here".

I don't think she is trying to make fun of her sister with this parody. Rather, she is acting out her alleged sister's frustration to reinforce the humorous nature of the interview and to acknowledge and support the host's remarks. She decides to imitate her sister, highlighting her annoyance at having Condola coming across as a successful attorney in a relevant position (albeit in fiction), which, as Condola construes and presents the situation, her sister thinks should be her own exclusive purview.

The humorous utterance identified in the sample corresponds to Condola Rashad remarks as referred in (78). The CDS includes Condola Rashad's usage

event as stated above, plus Stephen Colbert's previous remarks, preceded by Condola Rashad's own account of her role as an assistant DA in *Billions*. The shared knowledge comprises the job performed by attorneys; the different ranks within the administration, i.e. that a district attorney—even an assistant—ranks higher than an attorney working for the mayor; sibling relationships; how actors prepare for a role, etc. Elements listed in the context of speech are Condola's work as an actress, the popularity of the series *Billions*, the recognition achieved by Condola thanks to this character, and possibly also the salary entailed, which is presumably higher than a mayor attorney's salary.

The ground, as in the previous examples, is characterised by the tone of the interview, the programme genre, and the casual conversation in which guests are expected to be funny and play along with the host. The viewing frame includes Condola's role in *Billions*, her preparation for the role, and the assistance provided by her sister. The focus of attention, i.e. the salient element in the viewing frame, is placed on her sister's frustration at Condola playing a character with a position she would like to hold herself. Notice that it is Stephen Colbert who has directed the focus of attention to that feeling of irritation in Condola's sister. Condola keeps the focus of attention bringing her sister's point of view on the situation, as Condola herself construes it, conveying it through parody.

The conceptualisation pole of the viewing frame entails the turn taking process as managed by Condola, replying to the host's questions and subsequently yielding the turn. As for information structure, she picks up on Stephen Colbert's remarks, confirming that bit of information and adding her sister's reaction from Condola's viewpoint. The objective situation is precisely her

sister's feelings, expressed by the words uttered by Condola as if she were her own sister. This latest element already belongs to the segmental content in the vocalisation pole, accompanied by the other two channels, intonation and gestures, which very clearly delimit the part in which she is impersonating her sister by adopting a different stance, bodily posture, hand movements pointing in a reproachful attitude, mockingly portraying her allegedly angry sister.

With regards to where humour lies in this example of parody, I claim it lies in the vocalisation pole. The way Condola Rashad acts out her sister's words and attitude leads to the humorous effect. Not only the mock angry tone in which Condola delivers her sister's speech but the actual words, i.e. the segmental content, she decides to use. Especially, the sister's supposed claim "I'm the real one here" marks a contrast between the successful but fake Condola-assistant DA and the hard-working, but unknown sister-real attorney. So in this case, I believe humour relies on the interplay between deliberate gestures and mock angry tone, as well as the somewhat exaggerated opposition fake/real.

7.3.3 Mental spaces

Section 7.3.2 showed how cognitive linguistics can be elicited to account for humorous communication in conversation. I make the case for using Langacker's (2001) notion of current discourse space as the core framework to explain interaction and discourse, with a clear link to conceptualisation. Once this overarching framework has been established, we can zoom in to analyse more in detail the conceptualisation process itself. To do so, I resort to the notion of mental spaces (Fauconnier, 1981, 1994, 1997) and conceptual blending

(Fauconnier & Turner, 1998, 2002, 2003; Coulson, 2005a, 2005b; Coulson & Oakley, 2005; Dynel, 2011). Mental spaces are “small conceptual packets constructed as we think and talk, for purposes of local understanding and action” (Fauconnier & Turner, 2002, p.40). Blending refers to the process of merging different input spaces into a new space, i.e. the blend, that contains some elements from the input spaces, plus new emergent structure. Blends and input spaces conform a conceptual integration network (Fauconnier & Turner, 2002). I will now make some terminological clarifications to set out how, for the purposes of this study, I understand certain basic notions in cognitive linguistics which sometimes overlap and which could be apprehended in slightly different ways: domain, mental space, frame, image schema.

Domain is the conceptualisation of an experience that provides a background framework of knowledge for understanding concepts. Langacker (1987, p.488) defines domain as “a coherent area of conceptualisation relative to which semantic units may be characterised”. Therefore, although Langacker definition of the CDS refers to it as a mental space (2001, p. 144), I don't equate the CDS to a mental space in Fauconnier and Turner's (2002) terms, but to a more encompassing and larger notion of domain (Cienki, 2007). Mental spaces, in turn, are partial and dynamic representations of relevant information in a particular domain (Coulson, 2005a), whereas frames are schematic entrenched mental spaces which can structure blending and the resulting conceptual integration network (Turner, 2015); Finally, image schemas are pre-conceptual patterns stemming from our spatiotemporal experience and elicited to map special structure onto conceptual structure (Oakley, 2007).

We can now turn to applying an analysis based on blending and mental spaces to the humorous instances studied in order to delve into the conceptualisation process entailed. As said before, I consider the CDS to be a domain. Mental spaces and the conceptualisation process of blending, therefore occur against the backdrop of the CDS. More specifically, the mental spaces elicited during conversation and the blending operations will be directly related to the objective situation channel in the conceptualisation pole of the viewing frame, as it is the substantive element being construed (Langacker, 2001). Mental spaces, though, can be prompted drawing on any of the components of the CDS, by eliciting an element of the ground, the context, the shared knowledge, previous usage events or expectations on subsequent usage events.

As explained above, mental spaces are organised in conceptual integration networks (Fauconnier & Turner, 2002; Coulson, 2005b). Typically, the most basic form of network will include at least two input spaces, and the blend, onto which certain elements of each input space are mapped. Fauconnier and Turner (2002) advocate a generic space too, containing the structure shared by the input spaces. However, the need for a generic space has been questioned by some authors (Coulson, 2005a; Brandt & Brandt, 2005; Dore, 2015). Fauconnier and Turner argue that the generic space may lead to entrenched structured spaces that can be elicited directly. As I see it, what may become entrenched is the blend itself, so I will endorse the view that the generic space is, at least, optional (Dore, 2015), and I will not include it in my analysis. Let's consider again Cristela Alonzo's anecdote:

(79) Cristela Alonzo: *The Specials* came out and she's like "Ay, it's the guy from SNL; you can watch this" [laughter in the audience; pause]. And then he starts talking and I'm like "Oh, damn. I can watch this" [laughter in the audience].

This example was annotated as having viewpoint as the main underlying mechanism. Furthermore, I have argued that humour lies in the clash between what children are supposed to be allowed to watch under socially acceptable parenting, and Cristela Alonzo's mother unknowingly granting her permission to watch inappropriate content.

The conceptualisation process itself can be illustrated resorting to blending and mental spaces as follows. I will represent mental spaces as headings of columns in the table, with elements of each mental space under the relevant headings. Elements in one mental space map onto the elements in another space that are in the same row (Coulson, 2005b).

SNL	Specials	Blend
Comedy programme	Comedy programme	Comedy programme
Comedian	Comedian	Comedian
Family-friendly		Family-friendly
Children can watch		Children can watch

Table 12. Mother's viewpoint

SNL	Specials	Blend
Comedy programme	Comedy programme	Comedy programme
Comedian	Comedian	Comedian
Family-friendly	Not suitable for children	Not suitable for children
Children can watch	Children can't watch	Children can watch, but mother allows

Table 13. Cristela Alonzo's viewpoint

In this instance, the viewpoint of Cristela Alonzo's mother is presented first. As shown in Table 11, the mother's viewpoint elicits two input spaces, one corresponding to the programme *Saturday Night Live*, that we have previously been told she watched with her family every week, so that she considered it family-friendly. The second input space corresponds to *The Specials* programme. Cristela's mother viewpoint is presented as the mother identifying the comedy genre, as well as the same comedian from *SNL*, so she creates a blend in which she maps not only counterpart features, but also *SNL* features for which she finds no counterpart in *The Specials* space. She therefore considers the programme to be family-friendly and grants permission for Cristela to watch it.

When Cristela's viewpoint and construal of the situation is presented, though, the input spaces do have corresponding counterparts for all the elements, as her knowledge of English enables her to recognise that the programme contains inappropriate language for children and that therefore she is not

intended as an audience. That is mapped onto the blend, which in this case contains a new emergent structure, i.e. her mother allows her to watch the programme as she deems it child-friendly. Therefore, in this analysis, the contrast that I claim leads to the humorous effect of this utterance is already present in both blends. Let's now turn to Susan Sarandon's example of irony:

(80) Stephen Colbert: Joan Crawford and Bette Davis?

Susan Sarandon: Yeah.

Stephen Colbert: Not at the same time, right?

Susan Sarandon: I don't think so. That would have been, maybe,
a bonding experience.

Bette Davis	Joan Crawford	Blend
Bette Davis	Joan Crawford	Bette Davis + Joan Crawford
Sexual partner	Sexual partner	Sexual partner
intercourse	Sexual intercourse	Sexual intercourse
		Simultaneously

Table 14. Stephen Colbert's conceptual integration network

Bette Davis	Joan Crawford	Stephen Colbert's blend	Susan Sarandon's blend
Bette Davis	Joan Crawford	Bette Davis + Joan Crawford	Bette Davis + Joan Crawford
Same sex.part.	Same sexual partner	Sexual partner	Sexual partner
Sex. intercourse	Sexual intercourse	Sexual intercourse	Sexual intercourse
Increased enmity	Increased enmity	Simultaneously	Simultaneously Bonding experience

Table 15. Susan Sarandon's conceptual integration network

In this instance, Susan Sarandon takes Stephen Colbert's blend as one of the input spaces in her conceptual integration network. The host's blend adds the new feature of having sex with the same partner simultaneously, which is then recruited in Susan Sarandon's own mental space. She merges this blend with her own two input spaces, in which the animosity between the divas was an element aggravated by having had sex with the same people. She brings the new element of simultaneity in the intercourse with the same partner to turn it into a way of bringing both actresses together. In her blend, Susan Sarandon highlights the relationship between Joan Crawford and Bette Davis, as a profiled element that we can assume was missing in Stephen Colbert's blend, thus enabling the ironic comment.

Finally, let's apply this mental-space analysis to the last example studied above, about Condola Rashad's sister attorney.

(81) Stephen Colbert: Does she ever get frustrated because you have the job that...?

Condola Rashad: Oh, totally. She was like, "Why do you get the job that I wanted? But you're playing it and I'm the real one here".

Condola Rashad	Sister	Blend
Role as assistant DA	Attorney	Fake / real attorney
Difficult job	Difficult job	Difficult job
Preparation / research	Preparation / research	Prep./ research for sister's benefit
Recognition as actress		Wish to be recognised
		Sister's frustration

Table 16. Stephen Colbert's conceptual integration network (2)

Condola Rashad	Stephen Colbert's blend	Blend
Role as assistant DA	Fake / real attorney	Fake / real attorney
Difficult job	Difficult job	
Preparation / research	Research for sister's benefit	
	Wish to be recognised	
Recognition as actress	Sister's frustration	Sister's frustration

Table 17. Condola Rashad's conceptual integration network

In my view, in this example, Stephen Colbert's blend serves as an input for Condola Rashad's own conceptual network. She incorporates that full blend and maps onto her own blend only the elements of the sister's feeling of frustration, as well as the opposition fake attorney / real attorney, that she conveys mimicking her alleged sister's words. Again, we can see the contrast leading to humour present in the blend, but as opposed to the previous examples, in which the blend is enriched with new elements, in this last humorous instance studied the blend run by Condola Rashad is streamlined and left only with those relevant items she decides to convey through parody.

According to Fauconnier and Turner (2002), emergent structure in the blend is possible through three operations involved in its construction: composition, completion, and elaboration. Composition refers to the establishment of relations between elements in the blend that are not present in the input spaces. For example, in (80) Bette Davis and Joan Crawford are brought together to have sex with the same person at the same time. So their individual distinct sexual relationships with the same partner are fused together to have them occur simultaneously. Completion, in turn, refers to the process of recruiting background structure and knowledge to complete the blend. The knowledge recruited is deeply entrenched and does not need us to set up an additional input space. For instance, in Cristela Alonzo's anecdote, the knowledge that parents do not let children watch inappropriate content is drawn upon to complete her blend, thus rendering the contrast with her mother permission starker. Finally, elaboration implies the imaginary development of the blend following its own logic and principles. Susan Sarandon's suggestion that Bette Davis and Joan

Crawford's threesome could have been a bonding experience entails an imaginary development of the blend.

7.3.4 The space structuring model

Coulson and Oakley (2005) also combined mental spaces with Langacker's (2001) notion of ground to account for the role played by context in meaning construction. Their conceptual integration network model includes not only input spaces, but also what they called a grounding box, including the elements of Langacker's ground in the CDS, i.e. interactants, interaction, speech event and immediate circumstances. They claimed that the grounding box is not a mental space and may be non-representational.

The grounding box contains the analyst's list of important contextual assumptions, which need not be explicitly represented by speakers, though they influence the way that meaning construction proceeds. When those assumptions are explicitly represented by speakers, they are represented as models in mental spaces within the integration network. They posited that contextual assumptions and concerns affect meaning construction because the grounding box can be used to specify roles, values, and experiences that in turn will contribute to grounding speakers' subsequent representations (Coulson & Oakley, 2005, p. 1517).

They equated Langacker's (2001) notion of ground and their own grounding box to Brandt and Brandt's (2005) semiotic space, which includes the construal of the immediate situation of ongoing discourse and the assumption of mutually-shared knowledge by interlocutors. I agree that the semiotic space as

described by Brandt and Brandt (2005) (cf. Section 4.5, Chapter 4) can be considered equivalent to Langacker's ground. I also endorse Coulson and Oakley's (2005) claim that mental spaces draw heavily on the ground for implicit conceptual assumptions.

Furthermore, I believe that when any of those implicit assumptions is represented explicitly by any of the interlocutors, then it becomes part of a mental space in the ongoing conceptual integration network. Going back to Cristela Alonzo's anecdote, for example, I posit that the notion of socially acceptable parenting, with regards to content allowed to be accessed by children, is an implicit notion. When it is explicitly represented by Cristela Alonzo through her remarks expressing surprise at being allowed to watch the inappropriate programme, that notion of socially acceptable parenting is part of her blended space and her conceptual integration network.

Having said that, I argue that not just the ground, but the whole current discourse space can be drawn upon for background knowledge and conceptual assumptions. As I see it, that notion of good parenting does not belong to the immediate circumstances of the speech event, but rather to the presumed shared knowledge among participants in the communication event. It could also be claimed that this socially acceptable parenting notion falls within the context of speech. In fact, the line between shared knowledge and context as defined by Langacker within the current discourse space is not always clear-cut. I see the context as more closely related to the speech event itself, including salient but non-explicit elements (Giora, 1997), while shared knowledge concerns a more general and comprehensive background encompassing non-salient items.

In any case, the elements that can help to “specify roles, values, and experiences that ground subsequent representations” (Coulson & Oakley, 2005, p. 1534) can be drawn from any of the components of the CDS. The CDS constrains the kind of mental spaces that can be elicited in a usage event, but that it is a dynamic process, whereby recruited mental spaces and elements trigger changes in the CDS too, opening new possibilities for conceptualisation and communication in subsequent usage events.

Coulson (2005a, 2005b) also resorted to blending theory to offer her own account of humour, sarcasm and irony, named the space structuring model. She proposed that “comprehension of a single event frequently requires speakers to set up multiple models of the same object in different mental spaces, in order to capture the differences between the object's properties in different contexts” (Coulson, 2005b, p. 134). She claimed that her model allows to go beyond the traditional definition of irony as the opposite of what is being said (Grice, 1989), to establish that both humour and sarcasm are indirect forms of communication. They entail the simultaneous construction of multiple mental spaces and blended models that often highlight disanalogies between the models elicited by the input mental spaces. Under this perspective, Susan Sarandon’s irony implies the simultaneous construction of several mental spaces: one in which Joan Crawford and Bette Davis sleep with the same people, on different occasions, which eventually reinforces the bad blood between them, and another mental space in which they would have sex with the same partner at the same time and therefore weave a tighter and more friendly relationship.

Coulson claimed that humour in this case might boil down to the unexpected perspective on the situation offered by this conceptual integration network. Furthermore, she posits that the process enabling joke comprehension within her space structuring model is frame-shifting (Coulson, 2015), defined as a “semantic and pragmatic reanalysis in which elements of the existing message-level representation are mapped into a new frame retrieved from long-term memory” (Coulson, Urbach, & Kutas, 2006, p. 229). She further argued that linguistic and non-linguistic information is integrated rapidly and does not require prior representation of the propositional content of an utterance. For Coulson, meaning construction stems from a series of routines that involve the creation of cognitive models —mental spaces— enabling “interpretation, action, and interaction” (Coulson et al., 2006 p. 247).

Coulson’s space structuring model is compatible with my previous account of humorous instances. The simultaneous construction of multiple mental spaces advocated by the space structuring model is entirely possible within the CDS, always associated to the conceptualisation pole of the utterance, as specified in the three relevant channels in the viewing frame. The grounding box presented in her model of conceptual integration really points to the significance of context in how we construe, represent, and communicate a given situation, so my proposal simply extends that notion to include all elements pertaining the current discourse space. Finally, I understand frame-shifting as a type of blending, where the ongoing conceptual integration network created online during interaction brings in a frame, i.e. an entrenched mental space, to confront it with the current representation of discourse. Coulson argues that frames are elicited from long-

term memory; I further claim that they are retrievable because they are part of the shared knowledge within the current discourse space.

I endorse Coulson's view that inputs from language, perception, social context and the interlocutor's cognitive state contribute to meaning and to construing the discourse situation in a particular manner. Therefore, changes in any of those elements result in different conceptualisation and meaning construction. I consider that all those elements are encompassed in Langacker's (2001) depiction of the current discourse space. Thus, the CDS is the basic framework for communication and conceptualisation, including humour (Veale et al., 2006).

7.4 Convergence with other approaches to humour

7.4.1 General Theory of Verbal Humour

Both Raskin's (1985) SSTH and Attardo's (2001) GTVH have been discussed in Chapter 2. The SSTH is essentially a semantic theory, which applies solely to single jokes, not to larger texts or other types of humour, and focused on humour competence, defined as "the capacity of a speaker to process semantically a given text and to locate a set of relationships among its components, such that he/she would identify the text —or part of it— as humorous in an ideal situation" (Attardo, 2001, p. 167). Raskin's theory predicts that humour will occur if the joke text is compatible with two opposing scripts (or frames), which fully or partially overlap. Attardo (1994; 2001) portrayed his General Theory of Verbal Humour (GTVH) as a revision of Raskin's SSTH in order to account for any kind of humorous texts, not only jokes. In doing so, he

took GTVH beyond semantics and placed it in the realm of linguistics at large (Brône & Feyaerts, 2003).

Brône and Feyaerts (2003) claimed that the GTVH is a cognitive theory, insofar as it explores the interface between language and cognition by studying how different parameters relate and contribute to creating humorous meaning. In particular, they pointed to the fundamentally cognitive nature of the Logical Mechanism, one of the six knowledge resources posited by Attardo (2001) as parameters that define the humorous text. The Logical Mechanism allows for the partial resolution of the incongruity at the core of the joke.

Attardo (2001, 2015) also resorted to mental spaces in his account of humour and irony, where he advocates the adoption of a certain ironic or humorous mode leading to a possible world –akin to a mental space, in Attardo’s view– where interlocutors can then operate. Attardo defines mode adoption as the “acceptance on H’s [hearer’s] part of a possible world, as defined by S [speaker], which differs from [the real world], i.e., the world that S and H mutually know” (Attardo, 2001 p. 176). Utterances that can trigger mode adoption, e.g. irony, humour, metaphor, etc. will enable the construal of a new mental space next to the base/reality space, so that the hearer does not have to reject the utterance as ill-formed. Attardo thus placed this account within pragmatics, to try to explain why speakers may decide to express their message indirectly, despite the higher cognitive effort required for the hearer to process it, but he drew on cognitive linguistics to argue that the construction of an ironical mental space avoids conflict between the presuppositions of an utterance and those of the speaker and hearer (Brône & Feyaerts, 2003).

The GTVH can be seen as compatible with my own account. I assume that mental spaces are always construed, elicited, and blended as we communicate. Once humour is recognised in a given construal of a situation, as represented through the relevant mental spaces, humour or irony can be further developed by means of elaborating that mental space (cf. section 7.3.2 above), which is in line with Attardo's mode adoption proposal. Furthermore, I view the six knowledge resources posited by GTVH as potential elements in mental spaces; the Logical Mechanism would govern the mapping of relevant items between mental spaces, and the relations established therein, so that the resulting blend is recognised as novel and humorous.

Let's consider Condola Rashad's parody of her sister (81) under this perspective. As seen before, Stephen Colbert sets up a mental space in which Condola's sister is frustrated to see that Condola is perceived as a successful attorney in her role in *Billions*, holding a position that the sister would want for herself in real life. It could be argued that this mental space, including the novel element of the sister's frustration, triggers the adoption of a playful and humorous mode, in Attardo's terminology. One could safely assume that, rather than frustrated, the sister would be happy for Condola's success. Moreover, the position of assistant DA is only a role, not a real job as an attorney, so there cannot be real frustration linked to the position allegedly coveted by Condola's sister, as Condola does not actually hold that position. However, Condola Rashad adopts the mode set by Stephen Colbert's mental space and reinforces it assuming and acting out the role of the frustrated sister.

In the resulting blend, on which the representation of Condola's sister frustration is based, one could identify the following KR: Script Opposition – the contrast between Condola's role as an attorney in *Billions* and her sister as a real attorney. That is the fundamental basis for this particular humorous instance, as if the sister had been a biologist, for example, no grounds for Stephen Colbert's construal of the sister's frustration would exist. The Situation evoked in the parody is that of a conversation between both sisters. In this case, there is no Target of the joke, as it is not really directed against anyone. As for the Narrative Strategy, the humorous instance takes the form of a monologue by the sister in the imaginary conversation elicited as the situation. Finally, the Language Knowledge resource is made of the actual words said by Condola while parodying the sister, and I dare to say that the way she says those words is also part of the language resource, thus establishing a link between this knowledge resource and the vocalisation pole in the viewing frame (Langacker, 2001).

As I see it, the Logical Mechanism governs the correspondences established between the input spaces from which these elements are mapped, which enrich the highly schematic conceptual integration network described above, and the resulting new mental space. The LM is based on a relation of analogy, as the core of the joke is the duality or opposition between the sister's actual attorney job and the fake position taken by Condola Rashad in the role she plays in *Billions*. Notice that in this example there is no incongruity to be resolved. In fact, as I will discuss in more detail in section 5 of this chapter, there is a growing body of literature that questions the very notion of incongruity and

incongruity resolution as a necessary condition for humour (Forabosco, 2008; Veale, 2015; Gibbs et al., 2018).

Furthermore, in his elaboration upon the GTVH, Attardo (2001, p. 47 – 49) referred to a storage area with mutually assumed and shared information as a main feature of the theory. To me, this storage area is akin to the current discourse space. Attardo further posited that the information is stored in “clusters of information (scripts, frames) that come surrounded by a web of associations and links to other clusters of information” (ibid.). In other words, a conceptual integration network. Attardo acknowledged that “the GTVH has been mainly developed on the basis of canned jokes and that its application to conversational humour is less than straightforward” (ibid., p. 68). He claimed that conversational jokes are created online and depend heavily on context, and that speakers constantly negotiate and update what they assumed to be mutually known and what they consider to be relevant at any point in the interaction. I argue that Langacker’s (2001) current discourse space can account for all the elements that enable such negotiation and updating process. Furthermore, it provides a framework where the contribution made by different approaches can converge, as the role of semantics, pragmatics and discourse to enable successful communication is fully acknowledged, on the basis of a highly overlapping construal of the speech event by all participants in the interaction.

7.4.2 Conceptual subversion

Veale (2015) put forward a complementary approach to what he calls juxtapositional theories of humour, which are largely based on mental spaces

(Fauconnier, 1994; 1997), frame-shifting (Coulson & Kutas, 2001; Coulson, 2015), and script opposition (Raskin, 1985; Attardo, 2001). He acknowledged the inherently juxtapositional nature of humour, as it always entails some form of comparison in order to yield an innovative view with regards to what is considered normative or is expected. He argued, though, that the crucial element for humour production and comprehension is the nature of the structures that are juxtaposed. For Veale, humour lies in the conceptual subversion of a given category or mapping, rather than in the combination of different but overlapping categories. In other words, Veale argued that humour production and comprehension is not based on blending or cross-mapping various different input spaces (frames, scripts or mental spaces), but rather in manipulating and exploring the boundaries of a single input structure in order to identify novel and interesting counter-examples. He refers to this process as subversion.

He further posited that the appropriate counter-example to the category subverted is constructed “by stripping away the layers of conventionality and habitual thinking that have accreted around a category” (Veale, 2015, p. 77). In my view, the counter-example advocated by Veale as necessary to construe a novel and creative view of a given category is reminiscent of Coulson’s space structuring model (2005b). This approach advances that the simultaneous mental spaces of the object, category, event, etc. to be construed are created as different versions of that concept. This allows to identify the differences between those possible conceptualisations. Veale’s counterexample, therefore, would be one of those mental models created and confronted to the traditional frame (as an entrenched mental space) depiction of the category.

Let's consider one of the examples used by Veale (2015) to illustrate his point, in which he reports a witticism by Zsa Gabor, a woman known for her multiple marriages and divorces:

(82) Darlink [*sic.*], actually I am an excellent housekeeper. Whenever I leave a man, I keep the house! (ibid., p. 77)

Veale claimed that the humour in this remark stems from a novel reading of the term 'housekeeper', which introduces a new member into that category, i.e. a woman who literally keeps the house rather than a woman that strives to maintain the house in good condition and make everything work for the family. According to Veale, the category of 'housekeeper' is thus subverted to introduce a novel reading. I agree with his claim that the boundaries of the category are exploited creatively for humorous purposes, but I still believe that in order to subvert such category a novel blended space is construed from multiple inputs, which could schematically be described in Table 18 below.

Zsa Zsa Gabor	Housekeeper	House - keeper
Failed marriage	Successful marriage	Failed marriage
Divorce		Divorce
House ownership	House work	
	Housekeeping	Keeps/owns house

Table 18. Conceptual integration network leading to subversion of the category

In this particular example, comprehension of the subversion entailed in Zsa Zsa Gabor's comment is facilitated and made salient by her own words. Furthermore, subversion of the category actually relies on the contrast between the typical construal of a diligent and hardworking housekeeper, and the depiction of an independent and non-conventional woman that collects houses upon divorcing one husband after another. Consequently, in my view, conceptual subversion is one of the possible outcomes resulting from conceptual integration networks. Completely novel concepts, new category members or boundaries, or unexpected mappings between inputs may arise as a result of emergent structure in the blend.

Veale's proposal helps to better understand one of those phenomena. In addition, I posit that any of the different components of Langacker's (2001) CDS, —viewing frame, ground, context, shared knowledge, previous usage events and expectations on subsequent usage events— can be recruited to subvert the concept at hand. In that, my view converges with Veale's, as he considers that humour does not lie in the subversion itself, but in the pragmatic and social uses of that subversion.

7.4.3 Layering model and pretence space

Clark's (1996) model of layered meaning has been used by some authors to account for humour (Brône, 2008; Tabacaru, 2014). Clark posited that communication can occur at different levels, as there may be several layers of conceptualisation which are like "theatre stages built one on top of the other" (Clark, 1996, p. 16). Conversations may just occur on the first layer, but additional

layers may be necessary in instances of indirect speech, such as irony or humour. According to Brône (2008), layers represent different discourse worlds based on the actual utterances, which conform the surface level. In humour, sarcasm or irony, the first layer is the discourse base space, with the literal interpretation of an utterance, whereas a pretence space is created in a second layer, in which the intended meaning of the utterance becomes apparent (Brône, 2008; Tabacaru, 2014). Brône, therefore, equated layers to mental spaces, and so does Coulson (2005b). Tabacaru and Lemmens (2014) argued that layering and mental spaces show how people can access and process information, and how communication is made possible by resorting to meaning derived from previous discourse or background knowledge.

This is fully compatible with my account of humorous communication based on Langacker's (2001) CDS and Fauconnier and Turner's (2002) blending. Parallelisms can be drawn between the pretence space (Clark, 1996; Brône, 2008) and the notion of counterexample leading to subversion (Veale, 2015) as part of the multiple mental spaces construed in interaction to confront and conceptualise the possible various readings of a given object or event (Coulson, 2005b). As for the relevance of discourse and background knowledge, it is fully explained by the role that the CDS and all elements therein play in communication at large.

7.5 What makes something humorous?

The approaches to humour listed above describe the mechanisms whereby humour is produced and understood from a cognitive point of view.

Nevertheless, the question remains as to where the humorous nature of a given utterance, text, situation, etc. lies.

The most widespread paradigm in current humour studies is the incongruity-resolution (IR) principle. The concept of incongruity, therefore, is present in such ideas as script-opposition (Raskin, 1985; Attardo, 2001) –although Raskin rejects having his SSTH theory categorised as an IR theory (Raskin, Hempelman, & Taylor 2009)—, frame-shifting (Coulson et al. 2006; Coulson, 2015), relevant inappropriateness (Attardo, 2000a), discrepancies (de Jongste, 2016), pretence theory (Clark, 1996; Brône, 2008).etc.

These concepts differ to some extent but mostly overlap around the idea of incongruity as a mismatch in expectations (Bergen & Bindsten, 2004), a surprise (Giora, 1991), a violation of conventional wisdom (Coulson et al. 2006; Coulson, 2015), etc. Forabosco (2008 p.45) stated that a stimulus is perceived as incongruous when it diverts from the cognitive model of reference. This very much resonates with the process underlying the frame-shifting model (Coulson et al., 2006; Coulson, 2015). Broadly speaking, incongruity-resolution models assume that behind any instance of humour lies an incongruity that has to be resolved, at least in part. Coulson and Kutas (2001), though, reject the two-stage process of humour interpretation, namely the identification of a surprise or incongruity, followed by its resolution, i.e. the re-establishment of coherence.

Notice that incongruity does not *per se* entails the notion of outright opposition. What incongruity implies is a perceived lack of coherence, a breach of expectations, not necessarily by confronting opposing readings or mental

frames. Having said that, for humour to arise expectations need to be broken. That clash in expectations may be due to the presentation of opposing ideas, the subversion of a category, the unwinding of cultural conventions, the momentarily deception with regards to assumptions on future usage events in a current discourse space, etc. In my view, that clash of expectations is manifested through a radical disruption in the CDS, in any of the components therein; i.e. the focus of attention, the ground, context or shared knowledge. Furthermore, that mismatch in expectations can be based both on the conceptualisation and on the vocalisation channel, as multimodality in face-to-face communication, e.g. verbal, gestures, intonation, can be recruited to reinforce and signal the humorous nature of a given usage event as construed by the speaker (Attardo et al., 2003; Attardo, 2008; Attardo, Pickering, & Baker, 2011; Tabacaru & Lemmens, 2014; Bryant & Gibbs, 2015).

Having said that, the presence of an incongruity alone is not sufficient for humour to arise, as incongruities exist in a variety of situations where humour will most probably not be the outcome (Morreal, 1983; Morreal 2009a, 2009b; Veale, 2015). If a dog suddenly appears and fiercely barks at an absent-minded passer-by, fear rather than humour is likely to be triggered. So the question remains as to what other conditions are necessary, and, possibly, sufficient for humour to occur. Giora argued that we find pleasure in stimuli that are innovative but still familiar (Giora et al., 2009). We find a remark, picture, cartoon, etc. pleasurable when we recognise the familiarity in the novel. Her Revised Optimal Innovation Hypothesis (Giora et al., 2017, p.10) states that a stimulus is optimally innovative if

[It] involves a non-default response to a given stimulus, [...], while allowing for the automatic recoverability of the default response(s) related to that stimulus, so that both the default and non-default responses may be weighed against each other, their similarity and differences assessable.

Default responses are automatically retrieved and processed, as they are salient and based on coded meaning, i.e. coded in the lexicon, whereas non-default responses are non-salient, usually context-dependent, and take more time to be processed. According to Giora, therefore, we find pleasure when confronted to a non-default, i.e. unexpected, stimulus, as long as we can recall the expected default response to assess the (dis)analogy between them.

In my view, this hypothesis is reminiscent of incongruity, as there is a notion of surprise and inconsistency, as well as of partial resolution of that incoherence upon the comparison of the default and non-default stimuli. Notice that Giora does not specifically refer to humour, but to a pleasurable effect. Indeed, upon reading a whodunnit we find pleasure, but not humour, in finally resolving the mystery, that is, the incongruity, running through the story (Forabosco, 2008; Morreal, 2009b; Veale, 2015). Some authors argue that the incongruity in humour is only partially solved, as opposed to other phenomena with incongruity as a core element, e.g., detective stories or some poetry, and that it is precisely the residual incongruity left what renders a particular text, utterance, event humorous (Forabosco, 2008; Morreal, 2009b; Hempelman & Attardo, 2011).

Apart from the presence of an incongruity, the need to resolve it only partially, with a stimulus that presents the right mix between innovation and familiarity, another condition claimed as necessary for humour is the need for interlocutors to be in the right playful and humorous mode (Canestrari, 2010; de Jongste, 2018). Morreal (2009b) considered humour to be a reaction to something that violates our mental patterns and expectations, i.e. an incongruity, that we not only experience but enjoy. He acknowledges, though, that this is not sufficient for humour, as incongruities can also be enjoyed “in other ways than amusement, as in the grotesque, the macabre, the horrible, the bizarre, and the fantastic” (p. 241).

Morreal further posited that humour is a cognitive phenomenon –as it involves perception, thought, mental patterns and expectations– that triggers a sudden cognitive shift that we find pleasurable because we perceive it as a kind of play. For that we need to be in “a play mode, disengaged from practical and noetic concerns” (p. 243). In other words, humour can only arise if we have a certain predisposition towards it and our concern for the truth is suspended.

De Jongste (2018) claimed that the basic requirement for successful humorous communication is that interlocutors shared or can quickly switch into a paratelic, i.e. non-serious, state, in which the orientation towards essential objectives is suspended. This allows for a challenging but pleasurable form of play in which interlocutors negotiate and coordinate their construal of the situation. De Jongste further posited that when people are not in a paratelic state or cannot switch to it because of certain emotions, for example, they may still

recognise a certain stimulus as humorous, but they will reject it and not engage in it.

This opens the question as to whether humour relies on the intent of the speaker or the apprehension of the content as humorous by the hearer (Hay, 2001; Canestrari, 2010). De Jongste (2016) propounded that for humorous discourse to be successful, the humorous intention of the speaker needs to be identified and accepted by the hearer. He further argued that the humorous intent must be partially hidden by the producer, although it is meant to be detected, thus conforming an enjoyable form of play. Humour intent can be signalled through verbal and non-verbal cues, but it is the hearer who ultimately decides whether that is perceived as funny or not. Veale (2015) questioned whether incongruity resolution is the main driver in humour creation, and gave pre-eminence to the social logic behind a humorous instance.

In order to account for the social dimension of humour, Canestrari (2010) added a seventh knowledge resource to the GTVH (Attardo, 2001), that she named the Meta-Knowledge resource. Canestrari claimed that an interaction is humorous when at least one of the interlocutors is aware of its humorous nature. This awareness, in turn, can be signalled through verbal or non-verbal cues, which serve to communicate the metamessage that the instance at hand has to be taken as humorous. She defined “the Meta-Knowledge Resource as the signal that refers to the speaker’s intention of being humorous and to the hearer’s recognition of such intention” (Canestrari, 2010, p.330). Feyaerts (2013) also highlighted the social dimension of humour. He defined intersubjectivity as our ability to comprehend the interlocutor’s perspective and to model their mental

states, as a crucial element in communication, and key for humour to be produced and recognised (ibid., p. 245).

Finally, moving beyond the idea of incongruity, Gibbs et al. (2018) advocated that humour arises from “disturbing the body in some non-serious ways” (Gibbs et al., 2018, p.10). They aligned their view with the benign-violation hypothesis on humour (McGraw & Warren, 2010), which “suggests that anything that is threatening to one’s sense of how the world ‘ought to be’ will be humorous, as long as the threatening situation also seems benign” (ibid. p. 1142). Gibbs et al. (2018) equated this threat perceived as benign to rough-and-tumble play fight and made the case for the embodied foundation of ironic humour.

Although they claimed that benign violation is a more fundamental element to humour than incongruity, I believe that the notion of incongruity itself is embedded in the benign-violation hypothesis, provided we understand incongruity in a broad sense, as a diversion from the cognitive model of reference (Forabosco, 2008 p.45). In my view, the benign-violation hypothesis adds the perception of that diversion as a threat, but it does not invalidate the notion of incongruity as a violation of the cognitive model being threatened (social convention, belief, the perception of one-self or others, etc.). It is also reminiscent of the subversive role of humour with regards to the undermining of “habitually held world views” (Veale, 2015, p. 87).

Humour is such a complex phenomenon that it resists being reduced to sufficient and necessary conditions (Veale, 2015). After reviewing all the approaches studied in this chapter, I still consider that incongruity, loosely

understood as a diversion from the cognitive model of reference (Forabosco, 2008, p.45), is a core element of humour. For humour to occur, a sudden cognitive shift is needed (Morreal, 2009b), which I equate to a sudden disruption in the current discourse space.

In my view, sudden changes in cognitive states are brought about by a clash between the expectations held on an utterance, and its actual instantiation. That may take different forms (Veale et al., 2015), as a violation of conventional wisdom (Veale, 2015), an juxtaposition of opposing ideas (Raskin, 1985; Attardo, 2001), a threat to our own world-view perceived as benign (McGraw & Warren, 2010; Gibbs et al., 2018), or a mismatch with regards to frames or cognitive models stored in our memory (Coulson et al., 2006; Coulson, 2015; Forabosco, 2008; de Jongste, 2018).

Resolving the incongruity is not always necessary for humour to be successful (Veale, 2015; Gibbs et al., 2018). Absurd humour, for example, is based on extreme incongruities which are left unresolved. Having said that, when a connexion can be made between what is novel, i.e. incongruous, and what it is found as familiar, it is easier to find pleasure, as it requires less cognitive effort (Giora et al., 2017).

I fully endorse the view that a paratelic state shared by interlocutors or, at the very least, a cognitive and emotional mode or state which enables to quickly switch to that paratelic state is a necessary requirement for humour to be successful. Ultimately, humour is a fundamentally social phenomenon (Veale et al., 2006; Brône, 2008; Feyaerts, 2013; Veale, 2015) and it can only be

considered successful when acknowledged and accepted by the hearer (Hay, 2001; Canestrari, 2010).

From a cognitive perspective, there may be usage events conceived as humorous by the speaker, but humour will fail if it is not apprehended or approved by the hearer, which means that the paratelic is not shared. On the other hand, there may be unintentional humour, which is perceived as amusing by the hearer although the humorous intent was absent in the speaker's production. In that case, for humour to be ultimately successful, the speaker will need to acknowledge and accept humour and join in the paratelic state.

Furthermore, humour heavily relies on the circumstances surrounding communication, i.e. the pragmatic context (Curcó, 1997; Attardo, 2008; Yus, 2016). For humorous interaction to be successful, both interlocutors need to share mutual assumptions about how the counterpart is construing the speech event, so that they can engage in the negotiation, alignment and re-alignment of their intersubjective viewpoints (Feyaerts, 2013), drawing upon the different elements of the current discourse space, which fully acknowledges the role of the pragmatic and discourse dimension in determining whether humour succeeds or fails (Veale, 2015).

7.6 Conclusion

In the previous sections, I provide a streamlined and simplified model to account for humorous communication based on Langacker's (2001) notion of current discourse space and Fauconnier and Turner's (2002) theory on blending and conceptual integration networks. In doing so, I join many authors that claim

that there are not linguistic mechanisms specific to humour; rather, it can be described by the same linguistic principles governing communication at large (Bergen & Bindsten, 2004; Brône & Feyaerts, 2003; Brône et al., 2006; Hempelman & Attardo, 2011; Tabacaru, 2014; Veale et al., 2015; Dynel, 2018).

In addition, I bring together different approaches to humour with a more or less explicit cognitive stance, in order to seek similarities and convergence. The purpose of this exercise was to distil the essence of the different models and notions put forward to find the common denominator, as a way to map out the core features of humorous communication.

Finally, I pondered what it is that renders an instance of communication humorous, again by exploring the explanations offered by major authors in the field, to conclude that humour is a highly complex and varied phenomenon that is hard to categorize. In essence, humour is a cognitive-social phenomenon (Feyaerts, Brône, & De Ceukelaire, 2015) based on a pleasurable sudden cognitive shift (Morreal, 2009a, 2009b) brought about by a perceived incongruity (Forabosco, 2008), which in turn can be manifested through different forms.

Successful humour requires interlocutors align their construal of a communicative event (Canestrari, 2010; Feyaerts, 2013), and that they share or can rapidly switch to a playful mode (Forabosco, 2008). The pragmatic and discourse context, as represented by the full current discourse space, plays a decisive role both in the conceptualisation and expression of humour, as well as in its acknowledgment by the hearer (Attardo, 2008; Antonopoulou et al., 2015; Yus, 2016).

Furthermore, as seen in chapter 6, multimodal cues can also be used to prompt such alignment and to signal the humorous intent of a given usage event (Attardo et al., 2003; Attardo, 2008; Attardo et al., 2011; Tabacaru & Lemmens, 2014); i.e., to signal the paratelic state in which the utterance has been said (Bergen & Bindsten, 2015; Tabacaru, 2014; Bryant & Gibbs, 2015; Gibbs et al., 2018).

Chapter 8

Conclusion

8.1 Introduction

The research presented in this thesis broadly aimed at gaining an insight into how spontaneous humour is multimodally conveyed and cognitively construed in face-to-face interaction. To that end, a study of prosody, face displays and head movements co-occurring with humorous speech was conducted, with a view to disentangling the function of such multimodal cues and determine whether they could be considered markers of humour. From the cognitive perspective of the analysis, I explored construal mechanisms underlying humorous utterances, and delved into the creativity implied in humorous statements resorting to conceptual blending. The present chapter summarises the results and conclusions obtained in my research and outlined through the previous chapters, before stating the major findings of the study, both with regards to the research questions asked in the introduction, and in terms of the contribution made to scientific knowledge on the topic. Finally, avenues for future research are also suggested.

8.2 Summary of the dissertation

In Chapter 1 the main purpose of the study is presented, subsequently translated into the research questions guiding the analysis. The theoretical framework underpinning the study is briefly sketched, in order to explain the relevance of the multimodal and cognitive paradigm to study humour, as well as to provide the rationale for the study, demonstrating the need to fill a gap in the

literature, as far as multimodal and cognitive studies of spontaneous humorous communication are concerned.

Chapter 2 reviews the literature concerning humour studies, specifically focusing on multimodal studies and on linguistic theories from a semantic, pragmatic, and cognitive perspective. A comprehensive survey of the literature on multimodality and humour is offered, highlighting the conflicting results stemming from multimodal studies of irony, and sarcasm (Attardo et al., 2003), and the lack of conclusive findings on multimodal markers of non-ironical, non-sarcastic humour (Gironzetti, 2017). Regarding linguistic studies, common ground is found among seemingly disparate theories, especially around the notion of incongruity, spelled in different approaches from a clash between opposing ideas (Raskin, 1985; Attardo & Raskin, 1991), to a breach of expectations with regards to our usual worldview (Giora, 1991; Brône & Feyaerts, 2003; Coulson, 2005b; Veale et al., 2006).

Chapter 3 revolves around the literature on multimodality, exploring the relation between this multidisciplinary approach and theoretical paradigms, such as semiotics, pragmatics, and cognitive linguistics. Given the focus on spontaneous humorous communication adopted for the present study, as opposed to staged or scripted humour, the next section is devoted to multimodality in face-to-face interaction, as a prototypical and inherently multimodal form of human communication, in which gestures, prosody, and speech interplay. It compiles findings by previous studies, which assign both interactional, and discursive functions to gestures, and prosody (Muller, 2013b). In addition, gestures may also be representational (Kendon, 2004; McNeill,

2013), while prosody typically conveys affective meaning (Brazil, 1997; Wennerstrom, 2011).

Chapter 4 delves into cognitive linguistics, outlining its major tenets to demonstrate the validity of combining a multimodal and cognitive analysis of communication, drawing mainly on the experiential basis for language advocated by cognitive linguists (Johnson, 1987; Lakoff, 1987; Langacker, 2001; Talmy, 2000a, 2000b, etc.). Blending (Fauconnier & Turner, 2002) and construal mechanisms (Croft & Cruse, 2004) are put forward as core elements of cognitive linguistics which can provide an insight into how humour is construed. Furthermore, the notion of current discourse space (Langacker, 2001) is thoroughly analysed and advanced as a framework to account for communication at large, and for humour in particular, in which different humour approaches can converge.

Chapter 5 presents the results of the qualitative empirical analysis of the sample, composed of 14 interviews from *The Late Show with Stephen Colbert*, in which 109 humorous utterances are identified, on the basis of laughter elicited in the audience. The methodology of the study, using Elan, Praat, and SPSS as analytical tools, is explained in detail. Annotations have been made in Elan on mean pitch and mean intensity values (both for humorous and non-humorous utterances), as well as on humour type, construal mechanisms underlying the humorous utterances, gestures, and transcription. A breakdown of the different humour types, construal mechanisms, and gestures found is included. The most frequent humour types are anecdote, irony, and parody. In addition, the outcome of the prosodic analysis is outlined, which concludes that no prosodic contrast

(Bryant, 2010) between humorous and non-humorous utterances has been detected. In other words, humour in the sample is not prosodically marked. Tilts, nods, raised eyebrows, and smiles are the most frequent types of gestures annotated, followed by head turns and head shakes. As for construal operations, the highest number of occurrences are linked to viewpoint, profiling, and subjectivity.

Chapter 6 includes a multimodal analysis of a set of humorous utterances extracted from the sample. I have interpreted the uses and functions of the most frequently found face gestures and head movements, namely raised eyebrows (14.54%), smiles (9.17%), nods (20.36%), head tilts (16.67%), turns (11.86%) and shakes (11.86%). The analysis fails to find any correlation between the type of gesture performed and the kind of construal mechanism or humour type identified in the relevant humorous utterance. Furthermore, although in some instances, gestures and speech align to make a certain element of the utterance more prominent, and this element happens on some occasions to be key to the humorous interpretation of such remark, no recurrent pattern of association between gestures and humour has been identified. In conclusion, according to the analysis of the examples selected from the sample, it cannot be inferred that face gestures and head movements in the interviews studied are used as markers of humour. Instead, they can be interpreted as having the same functions as in non-humorous communication, which are also recruited to convey humour.

Finally, Chapter 7 offers a thorough cognitive analysis of several humorous instances found in the sample, on the basis of the most frequently occurring construal mechanisms –viewpoint (38.01%), profiling (19.00%), subjectivity

(16.94%)—, and their combination with the most frequent humour types —anecdote (26.83%), irony (13.82%), parody (13.00%)—. All humorous instances are analysed resorting to Langacker’s (2001) current discourse space, as well as accounting for the new meaning created through conceptual blending (Fauconnier & Turner, 2002). I claim that both notions add to a simplified model which can be used to account for humorous communication, in which various approaches to humour can converge. In light of the analysis performed, I further endorse other authors in advocating that humour does not need specific models. It can be explained through the same principles governing non-humorous communication (Bergen & Bindsten, 2004; Brône & Feyaerts, 2003; Hempelman & Attardo, 2011; Brône et al., 2006; Tabacaru, 2014; Veale et al., 2015; Dynel, 2018). A final reflection on what renders a given utterance humorous leads me to conclude that humour lies in the interplay of different elements within the current discourse space, taking the latter as the basic framework for communication in face-to-face interaction.

8.3 Main findings

8.3.1 Answering the research questions

8.3.1.1 Are there multimodal cues that contribute to signal the humorous nature of an utterance?

Multimodal cues at large serve to facilitate communication (Ekman, 1982; McNeill, 2000; Kendon, 2004; Bryant, 2010; Muller, Ladewig, & Bressemer, 2013; Attardo et al., 2013; Cienki, 2013b; Ladewig, 2014a; Wagner et al., 2014; Tabacaru and Lemmens, 2014; Gironzetti, 2017; etc.). In this capacity, they can

be occasionally elicited to convey humorous meaning, enabling the apprehension of a given turn as humorous. For example, raised eyebrows, nods or intonation can be used to give more saliency to a certain element in an utterance (Flecha-García, 2010; Ishi et al., 2014; Wennerstrom, 2001). This may help to interpret it as humorous, rather than take it at face value. However, this function is not directly related to the humorous nature of the utterance *per se*. Instead, the results of the present study point to an *ad-hoc* use of these multimodal cues to convey humour, as they have a more general function linked to information structure or discourse management. Multimodal signs may at times be useful to stress a humorous element, and to signal a marked form of communication, as opposed to uttering some run-of-the-mill statements that are to be taken at face value.

Ultimately, when it comes to spontaneous face-to-face communication, according to the results of this study, the humorous character of an utterance does not determine its multimodal expression. Instead, the latter will hinge upon context and discursive elements, such as the degree of familiarity with the interlocutors (Ishi et al., 2014), or the genre shaping communication (Bateman, 2008), i.e. a humour-prone setting —a late-night show or stand-up comedy— as opposed to a more serious context, like a job interview, for example. In sum, the multimodal expression of an utterance, whether humorous or not, will depend on the configuration and interplay of the different elements within the current discourse space.

8.3.1.2 *Can they be considered markers, i.e., are they consistently associated with humour and do they help to predict it?*

As explained in the answer to the previous question, my claim is that there are not multimodal markers of humour, to the extent that markers are defined as consistent multimodal cues recurrently co-occurring with humour and which help to predict it (Attardo, Wagner, & Urios-Aparisi, 2011). As seen in the analysis of various examples in the sample, the same gesture can have different functions, even within a single instance of humorous communication, not necessarily linked to its humorous nature. In addition, the multimodal cues found are not exclusive to humour and do not seem to perform distinct functions when co-occurring with humour, as compared to those listed in studies of non-humorous communication. In conclusion, I posit that multimodal cues can be recruited to facilitate the comprehension of humorous utterances as humorous, not due to the fact that the usage event (Langacker, 2001) at hand is humorous, but as part of the usual function of multimodal cues as facilitators of communication at large.

8.3.1.3 *What are the cognitive mechanisms involved in the production of humour?*

The cognitive analysis of the sample studied in this dissertation has applied a range of notions from cognitive linguistics. First, Langacker's (2001) current discourse space is used as the basic framework underpinning any communicative event, from which meaning is construed and ultimately expressed. In addition, a more fine-grained analysis of the online creativity

implied in face-to-face humorous communication has been conducted resorting to conceptual blending, and showing how the blends construed by each interlocutor in a given usage event have in turn an impact on the blend that will be elicited by the counterpart in the interaction.

Finally, a study of humorous utterances has been conducted to try to discover the main underlying construal mechanisms. The results of this multi-layered cognitive analysis lead me to conclude that, as is the case with the multimodal expression of humour, no specific cognitive mechanism —among those studied in the sample— can be associated with the construal of humour. In other words, despite humour clearly being acknowledged as a marked form of communication, to the extent that humorous remarks are not simply to be taken at face value, the cognitive processes applied to the production of humour in this sample do not differ from those governing non-humorous communication.

8.3.1.4 Where does the humorous nature of an instance of communication lie?

In light of the findings reported above, the question remains as to what makes a certain utterance humorous. This is by no means a simple matter. To begin with, this study has focused on the production of humour, but has taken laughter in the audience as the main indicator to identify humorous utterances, in order to avoid bias, and following standard practice in the literature (Morreal, 1983; Attardo, Pickering, & Baker, 2011; Archakis & Tsakona, 2005; Flamson et al., 2011; Tabacaru, 2014; Bryant & Gibbs, 2015).

That opens the question of whether humour is linked to the humorous intent of the speaker, or to the apprehension of a given instance as humorous by the audience. As shown in our study, in some utterances annotated as humorous, as they elicit laughter in the audience, the humorous intent on the part of the speaker is, at best, questionable. Only the producer of a humorous utterance could ultimately confirm whether the statement has been intended to be humorous. If we admit the widely-held view in the literature that humour stems from a form of incongruity which produces a pleasurable cognitive shift (Morreal, 1983), then humour would rather lie in the listener's apprehension, as the interlocutor experiencing the cognitive shift. On the other hand, if a certain utterance is produced with a humorous intent, a multimodal, cognitive, discourse or linguistic analysis focused on the speaker must consider that particular instance as humorous, even if humour fails (Hay, 2001).

In sum, humour is an extremely complex cognitive-social phenomenon (Feyaerts et al., 2015), which to be successful requires interlocutors align how they view and construe a particular communicative event (Canestrari, 2010; Feyaerts, 2013). I argue that the pragmatic and discourse context, as encompassed in the current discourse space (Langacker, 2001), plays a decisive role both in the conceptualisation and expression of humour, as far as speakers are concerned, as well as in the necessary acknowledgment of humour by the hearer (Attardo, 2008; Antonopoulou et al., 2015; Yus, 2016).

8.3.2 Contribution to knowledge

To my knowledge, this is the first multi-faceted study conducted on the production of spontaneous (not-scripted, not-rehearsed) humorous utterances, in which a multimodal and a cognitive approach are combined. Face and head gestures, along with various cognitive mechanisms, are analysed and confronted to all types of humour found in the sample, not just focusing on one particular kind. This study confirms the validity of applying the same frameworks and principles used in the research of non-humorous communication to the analysis of humorous instances. It convincingly advocates the relevance of conducting studies on spontaneous humour, as results may differ widely from those obtained in the research of scripted humour, given the major role played by the current discourse space (Langacker, 2001), i.e., the large context in which communication takes place.

The fully spontaneous nature of the interviewees' speech could be questioned, as most of them are people used to speaking in public and may therefore be seen as merely acting out their public persona during the show. However, even if that was the case, their speech could still be considered spontaneous to the extent that it had not been previously scripted and rehearsed. The research conducted for this thesis aimed at confronting spontaneous humorous utterances to those taken from sitcoms, TV shows or stand-up comedies in previous literature.

Regarding the multimodal analysis, the purpose of the study was to look into possible recurrent patterns between certain gestures and types of humour,

as well as to see whether gestures could be considered markers of humour, with functions specific to humorous communication. To do so, the use of the gestures found in humorous instances were compared to uses found in the literature for non-humorous communication. The research conducted on this sample bears out previous studies on the function of gestures, which seems to be the same both in humorous and non-humorous communication. Along the same line, this research reaffirms the explanatory power of cognitive linguistic notions, such as the current discourse space (Langacker, 2001), conceptual blending (Fauconnier & Turner, 2002), and construal mechanisms (Croft & Cruse, 2004), by applying them to a marked form of communication such as humour.

8.4 Suggestions for further research

The present study opens more questions that it helps to resolve, especially with regards to the humorous nature of any given utterance. In order to tackle this issue, further research into the multimodality and cognitive grounds of humour in spontaneous face-to-face interaction, adding also the point of view of the listener, is warranted, to gain an insight into how both interlocutors cooperate and may align their construal and multimodal expression of a humorous event. Such an approach would ideally widen the scope of the study from the individual to all the participants in the interaction, in order to gain an insight into how meaning is co-constructed among interactants at any given point in the current discourse space.

A further line of enquiry that would shed light into the multimodality of humour may entail a similar study but with a larger sample, in which not only face

displays and head movements, but also body posture, hand gestures and gaze would be annotated, to have a more comprehensive multimodal picture of how humorous communication unfolds. In addition, taking whole turns, rather than utterances, as the unit of study, may lead to better understanding of the role played by different modes and possibly help to detect certain multimodal patterns.

In light of the above, a more fine-grained analysis of gestures, accounting for repetitions, pace and amplitude of movements, etc. might contribute to better understanding the functions of specific types of gestures, such as shakes or nods. A semi-controlled experimental environment would be needed to conduct such analysis. Finally, a step forward in a multimodal study would look into the kinetic dynamics of all the interlocutors involved in a usage event.

Furthermore, multiple coders should ideally be involved in the study of a single sample to gain reliability and to reduce subjectivity. Also, research focused on outlining a systematic approach on utterance segmentation would greatly benefit this type of analyses and allow for better comparability of research outcomes.

Additionally, besides construal mechanisms and blending operations, a cognitive analysis can be enriched looking into the relation between the various figures of thought identified, e.g., metaphors of metonymic origin, irony based on metaphor or metonymy, differences between humorous and non-humorous irony on the basis of how the underlying figure is construed, etc.

Finally, on a different note, more related to my professional field of expertise, i.e., interpreting, further research could also look into how humour is

transferred from source to target language during interpretation, looking both at multimodal cues produced by the original speaker and the interpreter, as well as to similarities and differences in the construal of the humorous instance between interpreter and speaker, as to how it is rendered in both languages. The study of humour conveyed in settings in which English is used as *lingua franca* can also yield interesting insights as far as intercultural communication is concerned.

8.5 Conclusion

Communication is inherently multimodal. In addition, we cannot fail but to communicate according to patterns of belief, attitudes and behaviours (Poyatos, 2002). All our experiences of the world within our given culture will determine how we construe that world and how we express it through language (Johnson, 1987; Langacker, 2001; etc.). Hence, in order to gain an insight into communication between human beings, a multimodal perspective and a cognitive approach based on embodiment as the basic way in which we relate to the world around us is essential.

Humour is a social phenomenon. As such, an analysis of humour must take into account all participants in a communicative event. While engaged in interaction, we exchange messages through different modalities. To develop a comprehensive theory of interaction, we need to understand how these modalities are related, and look into the discourse mechanisms implied in, for example, cooperation between interlocutors. Multimodal elements are flexible and dynamic, and evolve during interaction. They are linked to our

conceptualisation of a particular event and provide clues as to how we construe what we mean to convey. Understanding language entails not only decoding words, grasping the context of communication, and assuming how interlocutors construe such context themselves. It also involves performing visual and motor imagery. As a result, a complete understanding of how we interact needs to take into account how all multimodal resources available interplay as well as how meaning is construed both individually and cooperatively among interlocutors.

References

- Adachi, T. (1996). Sarcasm in Japanese. *Studies in Language*, 20(1), 1-36. doi: 10.1075/sl.20.1.02ada
- Adami, E. (2016). Multimodality. In O. García, N. Flores, & M. Spotti (Eds.), *Oxford Handbook of Language and Society* (pp. 451-472). Oxford: Oxford University Press.
- Adami, E., & Kress, G. (2014). Introduction: multimodality, meaning making, and the issue of 'text'. *Text & Talk*, 34(3), 231-237. doi: 10.1515/text-2014-0007
- Akakin, H. Ç., & Sankur, B. (2011). Robust classification of face and head gestures in video. *Image and Vision Computing*, 29(7), 470-483. doi: 10.1016/j.imavis.2011.03.001
- Alibali, M. W., DiRusso, A. A. (1999). The function of gesture in learning to count: more than keeping track. *Cognitive Development*, 14(1), 37-56. doi: 10.1016/S0885-2014(99)80017-3
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15(6), 593-613. doi: 10.1080/016909600750040571
- Altorfer, A., Jossen, S., Würmle, O., Käsermann, M.L., Foppa, K., & Zimmermann, H. (2000). Measurement and meaning of head movements in everyday face-to-face communicative interaction. *Behaviour Research*

- Methods Instruments and Computers* 32 (1), 17-32.
<https://doi.org/10.3758/BF03200785>
- Ambrazaitis, G., & Bruce, G. (2009). Perception of South Swedish word accents. *Working Papers in Linguistics*, 52, 5-8. Retrieved from <https://pdfs.semanticscholar.org/2361/b31da94ac83a7134ee24c6ccdf4ebd6b37a2.pdf>
- Anolli, L., Ciceri, R., & Infantino, M.G. (2000), Irony as a game of implicitness: Acoustic profiles of ironic communication. *Journal of Psycholinguistic Research*, 29(3), 275-311. <https://doi.org/10.1023/A:1005100221723>
- Antonopoulou, A., Nikiforidou, K., & Tsakona, V. (2015). Construction grammar and discursal incongruity. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 13-48). Berlin/Boston: Walter de Gruyter.
- Archakis, A., & Tsakona, V. (2005). Analyzing conversational data in GTVH terms: A new approach to the issue of identity construction via humour. *Humour: International Journal of Humour Research*, 18(1), 41-68. doi: 10.1515/humr.2005.18.1.41
- Archakis, A., Giakomelou, M., Papazachariou, D., & Tsakona, V. (2010). The prosodic framing of humour in conversational narratives: evidence from Greek data. *Journal of Greek Linguistics*, 10, 187-212. doi: 10.1163/156658410X531375

- Attardo, S. (1993). Violation of conversational maxims and cooperation: The case of jokes. *Journal of Pragmatics*, 19(6), 537-558. [https://doi.org/10.1016/0378-2166\(93\)90111-2](https://doi.org/10.1016/0378-2166(93)90111-2)
- Attardo, S. (1994). *Linguistic Theories of Humor*. Berlin: Walter de Gruyter.
- Attardo, S. (2000a). Irony and relevant inappropriateness. *Journal of Pragmatics*, 32(6), 793-826. doi: 10.1016/S0378-2166(99)00070-3
- Attardo S. (2000b). Irony markers and functions: Toward a goal-oriented theory of irony and processing. *Rask. Internationalt Tidsskrift for Sprog og Kommunikation*, 12, 3-20.
- Attardo, S. (2001). *Humorous Texts: a semantic and pragmatic analysis*. Berlin: Walter de Gruyter.
- Attardo, S. (2003). Introduction: the pragmatics of humor. *Journal of Pragmatics*, 35(9), 1287-1294. doi: 10.1016/S0378-2166(02)00178-9
- Attardo, S. (2008). A primer for the linguistics of humour. In V. Raskin (Ed.), *The Primer of Humour Research* (pp. 101-156). Berlin: Mouton de Gruyter.
- Attardo, S. (2015). Humorous Metaphors. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 91-111). Berlin/Boston: Walter de Gruyter.
- Attardo, S., & Raskin, V. (1991). Script theory revisited: joke similarity and joke representation model. *Humor: The International Journal of Humor Research*, 4(3/4), 293-347.

- Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor: International Journal of Humor Research*, 16(2), 243-260. doi: 10.1515/humr.2003.012
- Attardo S., Pickering, L., & Baker, A. (2011). Prosodic and multimodal markers of humor in conversation. In S. Attardo, M. Wagner, & E. Urios-Aparisi (Eds.), *Prosody and Humor* (pp. 224-247). Amsterdam / Philadelphia: John Benjamins.
- Attardo, S., Wagner, M., & Urios-Aparisi, E. (2011). Prosody and humor. *Pragmatics & Cognition*, 19(2), 189-201. doi:10.1075/pc.19.2.01att
- Attardo, S., Pickering, L., Lomotey, F., & Menjo, S. (2013). Multimodality in Conversational Humor. *Review of Cognitive Linguistics*, 11(2), 402-416. doi: 10.1075/rcl.11.2.12att
- Bateman, J. (2008). *Multimodality and Genre. A Foundation for the Systematic Analysis of Multimodal Documents*. London: Palgrave.
- Baumann, S. (2006). Information Structure and Prosody: Linguistic Categories for Spoken Language Annotation. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in Empirical Prosody Research, Edition: Language, Context, and Cognition* 3 (pp.153-180). Berlin: Mouton De Gruyter. doi: 10.1515/9783110914641.153
- Baxter, J. (2002). Jokers in the pack: why boys are more adept than girls to speak in public settings. *Language and Education* 16(2), 81-96. <https://doi.org/10.1080/09500780208666821>

- Bergen, B., & Binsted, K. (2004). The Cognitive Linguistics of Scalar Humor. In Achard, M., and Kemmer, S. (Eds.), *Language, Culture, and Mind* (pp. 79-92). Stanford: CSLI Publications
- Bergen, B. & Bindsten, K. (2015). Embodied grammar and humor. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 49-68). Berlin/Boston: Walter de Gruyter.
- Bergman, M. (2007). The secret of rendering signs effective: the import of CS Peirce's semiotic rhetoric. *Public Journal of Semiotics*, 1(2), 2-11.
Retrieved from <http://www.helsinki.fi/science/commens/papers/renderingsigns.pdf>
- Bergmann K., Kahl S., & Kopp S. (2013) Modeling the Semantic Coordination of Speech and Gesture under Cognitive and Linguistic Constraints. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *Intelligent Virtual Agents. IVA 2013. Lecture Notes in Computer Science* (Vol 8108, pp- 203-216). Berlin/Heideber: Springer.
- Bertrand, R. & Priego-Valverde, B. (2011). Does prosody play a specific role in conversational humor?. *Pragmatics and Cognition*, 19(2), 333-356. doi: 10.1075/pc.19.2.08ber
- Bezemer, J., Jewitt, C., Diamantopoulou, S., Kress, G., & Mavers, D. (2012). Using a social semiotic approach to multimodality: Researching learning in schools, museums and hospitals. NCRM working paper (Unpublished). Retrieved from <http://eprints.ncrm.ac.uk/2258/>

- Bird, C. (2011). The prosody of riddle openings. In S. Attardo, M. Wagner, & E. Urios-Aparisi (Eds.) *Prosody and Humor* (pp. 268-290). Amsterdam: John Benjamins.
- Bohle, U. (2014). Body posture and movement in interaction: Participation management. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1301-1310). Berlin: de Gruyter Mouton.
- Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press.
- Bolinger, D. (1989). *Intonation and its uses : melody in grammar and discourse*. Stanford: Stanford University Press.
- Brandt, L., & Brandt, P. A. (2005). Making sense of a blend: A cognitive-semiotic approach to metaphor. *Annual Review of Cognitive Linguistics*, 3(1), 216-249. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2898881
- Brazil, D. (1984). The intonation of sentences read aloud. In D. Gibbon, & H. Richter (Eds.), *Intonation, accent and rhythm. Studies in discourse phonology* (pp. 44-66). Berlin: Walter de Gruyter.
- Brazil, D. (1997). *The communicative value of intonation in English*. 2nd edition. Cambridge: Cambridge University Press.

- Bressemer, J. (2012). *Repetitions in gesture: Structures, functions, and cognitive aspects* (Unpublished doctoral dissertation). Europa-Universität Viadrina, Germany.
- Bressemer, J., & Muller, C. (2014). The family of Away gestures: Negation, refusal, and negative assessment. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1592-1604). Berlin: de Gruyter Mouton.
- Brône, G. (2008). Hyper- and misunderstanding in interactional humor. *Journal of Pragmatics*, 40(12), 2027-2061. doi: 10.1016/j.pragma.2008.04.011
- Brône, G., & Feyaerts, K. (2003). The cognitive linguistics of incongruity resolution: Marked reference-point structures in humor. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.4193&rep=rep1&type=pdf>
- Brône, G., Feyaerts, K., & Veale, T. (2006). Introduction: Cognitive linguistic approaches to humor. *International Journal of Humor Research*, 19(3), 0933-1719. doi: 10.1515/HUMOR.2006.012
- Brône, G., Feyaerts, K., & Veale, T. (Eds.). (2015). *Cognitive Linguistics and Humour Research*. Berlin/Boston: Walter de Gruyter.
- Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of Intonation*. University Park Press.

- Bryant, G. (2010). Prosodic Contrast in Ironic Speech. *Discourse Processes*, 47 (7), 545-566. doi: 10.1080/01638530903531972
- Bryant, G. (2011). Verbal irony in the wild. *Pragmatics & Cognition* 19(2), 291-309. doi: 10.1075/pc.19.2.06bry
- Bryant, G., & Gibbs, R. (2015). Behavioral complexities in ironic humor. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 147-166). Berlin/Boston: Walter de Gruyter.
- Caballero, R., & Díaz, V. J. E. (2013). *Sensuous Cognition, Explorations into Human Sentience: Imagination, (E)motion and Perception*. Berlin, Boston: De Gruyter Mouton.
- Calbris, G. (2003). From cutting an object to a clear-cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture* 3(1): 19-46.
<https://doi.org/10.1075/gest.3.1.03cal>
- Canestrari, C. (2010). Meta-communicative signals and humorous verbal interchanges: A case study. *Humor*, 23(3), 327-349. doi: 10.1515/HUMR.2010.015
- Cap, P. (2017). Spatial Cognition and Critical Discourse Research. In J. Flowerdew, & J. E. Richardson (Eds.), *Handbook of Critical Discourse Analysis* (pp. 92-105). London: Routledge.

- Casasanto, D. (2013). Different bodies, different minds: The body specificity of language and thought. In R. Caballero & V. J. E. Díaz (Eds.), *Sensuous Cognition, Explorations into Human Sentience: Imagination, (E)motion and Perception* (pp. 9-18). Berlin, Boston: De Gruyter Mouton.
- Castelfranchi, C., & Poggi, I. (1990). Blushing as a discourse: was Darwin wrong? In R. Crozier (Ed.), *Shyness and Embarrassment. Perspective from Social Psychology* (pp. 230-251). Cambridge: Cambridge University Press.
- Cavé, C., Gua, I., Bertrand, R., Santi, S., Harley, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. *International Conference on Spoken Language Processing*, 2175-2178. doi: 10.1109/ICSLP.1996.607235
- Chandler, D. (2007). *Semiotics: The basics*. (2nd Ed.). London/New York: Routledge.
- Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, 126(3), 1394-1405. doi: 10.1121/1.3177275
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Cienki, A. (1997). Some properties and groupings of image schemas. In M. Verspoor, K. Dong Lee, & E. Sweetser (Eds.), *Lexical and Syntactical*

- Constructions and the Construction of Meaning* (pp. 3-15). Amsterdam/Philadelphia: John Benjamins.
- Clenki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), *From perception to meaning: Image Schemas in Cognitive Linguistics* (pp. 421-441). Berlin: Mouton de Gruyter.
- Cienki, A. (2007). Frames, Idealized Cognitive Models, and Domains. In D. Geraeerts, & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp.170-187). Oxford: Oxford University Press.
- Cienki, A. (2008). Why study metaphor and gesture? In Cienki, A. & Müller, C. (Eds.), *Metaphor and Gesture* (pp. 5-25). Amsterdam, Philadelphia: John Benjamins.
- Cienki, A. (2013a). Cognitive Linguistics: Spoken language and gesture as expressions of conceptualization. In C. Müller, A. Cienki, S. Ladewig, D. McNeill & S. Teßendorf (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1 pp. 182-201). Berlin: de Gruyter Mouton.
- Cienki, A. (2013b). Gesture, space, grammar, and cognition. In P. Auer, M. Hilpert, A. Stukenbrock, & B. Szmrecsanyi (Eds.), *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives* (pp. 667–686). Berlin: Walter de Gruyter.
- Cienki, A. & Müller, C. (2008). *Metaphor and Gesture*. Amsterdam / Philadelphia: John Benjamins.

- Cienki, A. & Müller, C. (2014). Ways of viewing metaphor in gesture. In C. Muller, A.J. Cienki, E. Fricke, D. McNeill. & J. Bressemer (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1766-1781). Berlin: De Gruyter Mouton.
- Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clausner, T., & Croft, W. (1999). Domains and image schemas. *Cognitive Linguistics* 10 (1), 1-31. doi: <https://doi.org/10.1515/cogl.1999.001>
- Coulson, S. (2005a). Extemporaneous blending: Conceptual integration in humorous discourse from talk radio. *Style*, 39(2), 107-121. Retrieved from <http://www.cogsci.ucsd.edu/~coulson/spaces/coulson-style.pdf>
- Coulson, S. (2005b). Sarcasm and the Space Structuring Model. In S. Coulson & B. Lewandowska-Tomaszczyk (Eds.) *The Literal and the Nonliteral in Language and Thought* (pp. 129-144). Berlin: Peter Lang.
- Coulson, S. (2005c). What's so funny? Cognitive semantics and jokes. *Cognitive Psychopathology/Psicopatologia cognitive* 2(3), 67-78. Retrieved from <http://www.cogsci.ucsd.edu/~coulson/funstuff/funny.html>
- Coulson, S. (2015). Frame-shifting and frame semantics: Joke comprehension on the space structuring model. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 167-190). Berlin/Boston: Walter de Gruyter.
- Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters*, 316, 71-

74. Retrieved from <http://www.cogsci.ucsd.edu/~coulson/Papers/coulson-kutas01.pdf>
- Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of Pragmatics*, 37(10), 1510-1536. Retrieved from <http://www.cogsci.ucsd.edu/~coulson/Papers/coak-jop.pdf>
- Coulson, S., Urbach, T., & Kutas, M. (2006). Looking back: Joke comprehension and the space structuring model. *Humor*, 19(3), 229-250. doi: 10.1515/HUMOR.2006.013
- Coulson, S., & Wu, Y.C. (2014). Multimodal discourse comprehension. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1922-1929). Berlin: de Gruyter Mouton.
- Croft, W. & Cruse, A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Cruttenden, A. (1986) *Intonation*. Cambridge: Cambridge University Press.
- Curcó Cobos, M. D. C. (1997). *The pragmatics of humorous interpretations: a relevance-theoretic approach* (Unpublished doctoral dissertation). University of London. Retrieved from <http://discovery.ucl.ac.uk/1317855/>

- Cvejic, E., Kim, J., & Davis, C. (2010). Prosody off the top of the head: Prosodic contrasts can be discriminated by head motion. *Speech Communication*, 52(6), 555-564. doi: 10.1016/j.specom.2010.02.006
- Cvejic, E., Kim, J., & Davis, C. (2012). Effects of seeing the interlocutor on the production of prosodic contrasts. *The Journal of the Acoustical Society of America*, 131(2), 1011-1014. doi: 10.1121/1.3676605
- Darwin, C. (1872). *The Expression of the Emotion in Man and Animals*. London: John Murray.
- Daylight, R. (2012). The Difference between Semiotics and Semiology. *Gamma: Journal of Theory & Criticism*, 20, 37-50 Retrieved from <http://www.enl.auth.gr/gramma/gramma12/B.1.pdf>
- de Jongste, H. (2016). Mental models and humorous intent. *Journal of Pragmatics*, 95, 107–119. <https://doi.org/10.1016/j.pragma.2016.01.006>
- de Jongste, H. (2018). Mental models, humorous texts and humour evaluation. *Review of Cognitive Linguistics*, 16(1), 97-127. doi: 10.1075/rcl.00005.jon
- de Ruiter, J.P., Bangerter, A., & Dings, P. (2012). Interplay between gesture and speech in the production of referring expressions: investigating the trade-off hypothesis. *Topics in Cognitive Science*, 4(2), 232-248. doi: 10.1111/j.1756-8765.2012.01183.x
- Deely, J. (1990). *Basics of Semiotics*. Bloomington: Indiana UP.

- Del Saz-Rubio, M. & Pennock-Speck, B. (2009). Constructing female identities through feminine hygiene TV commercials. *Journal of Pragmatics*, 41(12), 2535 – 2556. <https://doi.org/10.1016/j.pragma.2009.04.005>
- Deppermann, A. (2013a). Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics*, 46(1), 1-7. doi: 10.1016/j.pragma.2012.11.014
- Deppermann, A. (2013b). Turn-design at turn-beginnings: Multimodal resources to deal with tasks of turn-construction in German. *Journal of Pragmatics*, 46(1), 91-121. doi: 10.1016/j.pragma.2012.07.010
- Dore, M. (2015). Metaphor, humour and characterisation in the TV comedy programme Friends. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 191-214). Berlin/Boston: Walter de Gruyter.
- Drahota, A., Costall, A., & Reddy, V. (2008). The vocal communication of different kinds of smile. *Speech Communication*, 50(4), 278-287. doi: /10.1016/j.specom.2007.10.001
- Dynel, M. (2009). Beyond a Joke: Types of Conversational Humour. *Language and Linguistics Compass*, 3(5), 1284-1299. doi: 10.1111/j.1749-818X.2009.00152.x
- Dynel, M. (2011). Blending the Incongruity-Resolution Model and the Conceptual Integration Theory: The Case of Blends in Pictorial Advertising. *Review of Pragmatics* 3, 59-83. doi: 10.1163/187731011X561009

- Dynel, M. (2018). Taking cognisance of cognitive linguistic research on humour. *Review of Cognitive Linguistics*, 16(1), 1-18. doi: 10.1075/rcl.00001.dyn
- Efron, D. (1941). *Gesture and Environment*. Morningside Heights: King's Crown Press.
- 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. <https://doi.org/10.1016/j.pragma.2004.12.003>
- Ekman, P. (1979). About brows—emotional and conversational signals. In von Cranach, M. K. Foppa, W. Lepenies, & D. Ploog, (Eds.), *Human Ethology*: 169-248. Cambridge: Cambridge University Press.
- Ekman, P. (1982). *Emotion in the Human Face*. New York: Cambridge University Press.
- Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124-129. doi: 10.1037/h0030377
- El Kaliouby, R., & Robinson, P. (2004). Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. In B. Kiskanin, P. Pavlovic, & T. Huang (Eds.), *Real-Time Vision for Human-Computer Interaction* (pp. 181-200). Boston: Springer.
- ELAN (Version 5.1) [Computer software]. (2017, April 18). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>

- Elleström, L. (2002). *Divine madness: On interpreting literature, music, and the visual arts ironically*. Cranbury/London/Mississauga: Bucknell University Press.
- Enfield, N.J., & Sidnell, J. (2014). Language presupposes an enchronic infrastructure for social interaction. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language* (pp. 92-104). Oxford: Oxford University Press.
- Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Fauconnier, G. (1981). Pragmatic functions and mental spaces. *Cognition*, 10(1/3), 85-88. [https://doi.org/10.1016/0010-0277\(81\)90029-9](https://doi.org/10.1016/0010-0277(81)90029-9)
- Fauconnier, G. (1994). *Mental spaces. Aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.
- Fauconnier, G. (1997). *Mappings in Thought and Language*. Cambridge: Cambridge University Press.
- Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, 22(2), 133-187. doi: 10.1207/s15516709cog2202_1
- Fauconnier, G., & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Minds Hidden Complexities*. New York: Basic Books.
- Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches En Communication*, 19(19), 57-86. Retrieved from <http://tecfa.unige.ch/tecfa/maltt/cofor-1/textes/Fauconnier-Turner03.pdf>

- Feng, D., & O'Halloran, K. (2015). The visual representation of metaphor: a social semiotic approach. In M. Pinar-Sanz (Ed.), *Multimodality and Cognitive Linguistics* (pp. 99-114). Amsterdam: John Benjamins.
- Ferré, G. (2010). Timing relationships between speech and co-verbal gestures in spontaneous French. In *Language Resources and Evaluation Conference (LREC). Workshop on Multimodal Corpora, Malta. W6* (pp.86-91). May 2010, Malta. Retrieved from <https://hal.archives-ouvertes.fr/hal-00485797>
- Feyaerts, K. (2013). Tackling the complexity of spontaneous humorous interaction: An integrated classroom-modeled corpus approach. In L. Ruiz Gurillo, & M. B. Alvarado Ortega (Eds.), *Pragmatics & Beyond New Series* (Vol. 231, pp. 243–268). Amsterdam: John Benjamins Publishing Company.
- Feyaerts, K., Speelman, D., Brône, G., & Oben, B. (2010). *Corinth: Corpus Interactionele Humor*. University of Leuven.
- Feyaerts, K., Brône, G., & De Ceukelaire, R. (2015). The Art of Teasing. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 215-242). Berlin/Boston: Walter de Gruyter.
- Fillmore, C. J. (1968). The case for case. In E. W. Bach, & R. T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1-88). New York: Holt, Rinehart & Winston.
- Fillmore, C. J. (1975). An Alternative to Checklist Theories of Meaning. *Proceedings of the First Annual Meeting, Berkeley Linguistics Society*, 123-31. doi: <http://dx.doi.org/10.3765/bls.v1i0.2315>

- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and the Development of Language and Speech*, 260, 20-32. <https://doi.org/10.1111/j.1749-6632.1976.tb25467.x>
- Flamson, T., Bryant, G., & Barret, H. (2011). Prosody in spontaneous humour. *Pragmatics & Cognition*, 19(2), 189-201. doi: 10.1075/pc.19.2.01att
- Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, 52(6), 542-554. doi: 10.1016/j.specom.2009.12.003
- Forabosco, G. (2008). Is the concept of incongruity still a useful construct for the advancement of humour research?. *Lodz Papers in Pragmatics*, 4(1), 45-62. doi: 10.2478/v10016-008-0003-5
- Forceville, C. (2009). Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In C. Forceville, & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 19-42). Berlin: Walter de Gruyter.
- Forceville, C. (2014). Relevance Theory as model for analyzing visual and multimodal communication. In D. Machin (Ed.), *Visual Communication* (pp. 51-70). Amsterdam Berlin: Walter de Gruyter.
- Forceville, C. (2015). Metaphor and symbol: SEARCHING FOR ONE'S IDENTITY IS LOOKING FOR A HOME in animation film. In M. Pinar-Sanz (Ed.), *Multimodality and Cognitive Linguistics* (pp. 27-44). Amsterdam: John Benjamins.

- Forceville, C. (2016). Chapter 11. Mixing in pictorial and multimodal metaphors? In R. W. Gibbs (Ed.), *Metaphor in Language, Cognition, and Communication* (Vol. 6, pp. 223-240). Amsterdam: John Benjamins Publishing Company.
- Forceville, C., & Urios-Aparisi, E. (Eds.) (2009). *Multimodal Metaphor*. Berlin: Walter de Gruyter.
- Forceville, C., & Clark, B. (2014). CAN PICTURES HAVE EXPLICATURES? *Linguagem Em (Dis)curso*, 14(3), 451-472. doi: 10.1590/1982-4017-140301-0114
- Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 64(1), 83-93. <http://dx.doi.org/10.1037/0022-3514.64.1.83>
- Geraeerts, D. (Ed.). (2008). *Cognitive Linguistics: Basic Readings*. Berlin, Boston: De Gruyter Mouton.
- Geraeerts, D., & Cuyckens, H. (Eds.). (2007). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press.
- Gibbs, R., Samermit, P., & Karzmark, C. (2018). Humor, irony, and the body. *Review of Cognitive Linguistics*, 16(1), 72-96. doi: 10.1075/rcl.00004.gib
- Giora, R. (1991). On the cognitive aspects of the joke. *Journal of Pragmatics*, 16(5), 465-485. [https://doi.org/10.1016/0378-2166\(91\)90137-M](https://doi.org/10.1016/0378-2166(91)90137-M)

- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 8(3), 183-206. doi: 10.1515/cogl.1997.8.3.183
- Giora, R. (1998). Irony. In J. Verschueren, J. Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of Pragmatics* (pp.1-21). Amsterdam/Philadelphia: John Benjamins Publ. Co.
- Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, 34(2), 487-506. doi: 10.1016/S0378-2166(01)00045-5
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2009). Weapons of mass distraction: Optimal Innovation and pleasure ratings. *Metaphor and Symbol*, 19(2), 115-141.
https://doi.org/10.1207/s15327868ms1902_2
- Giora, R., Drucker, A., Fein, O., & Mendelson, I. (2015). Default Sarcastic Interpretations: On the Priority of Nonsalient Interpretations. *Discourse Processes*, 52(3), 173-200. doi: 10.1080/0163853X.2014.954951
- Giora, R., Givoni, S., & Fein, O. (2015). Defaultness Reigns: The Case of Sarcasm. *Metaphor and Symbol*, 30(4), 290-313. doi: 10.1080/10926488.2015.1074804
- Giora, R., Fein, O., Kotler, N., & Shuval, N. (2015). *Know Hope*: Metaphor, optimal innovation and pleasure. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 129-146). Berlin/Boston: Walter de Gruyter.

- Giora, R., Givoni, S., Heruti, V., Fein, O. (2017). The role of defaultness in affecting pleasure: the Optimal Innovation Hypothesis revised. *Metaphor and symbol*, 32(1), 1-18. doi: 10.1080/10926488.2017.1272934
- Giora, R., Jaffe, I., Becker, I., & Fein, O. (2018). Strongly attenuating highly positive concepts: The case of default sarcastic interpretations. *Review of Cognitive Linguistics*, 16(1), 19-47. doi: 10.1075/rcl.00002.gio
- Gironzetti, E. (2017). Prosodic and Multimodal Markers of Humor. In S. Attardo (Ed.), *The Routledge Handbook of Language and Humor* (pp. 253-254). London/New York: Routledge.
- Gironzetti, E., & Menjo, S. (2014). Smiling as a discourse marker of humor. *Proceedings from 2nd Conference of the American Pragmatics Association (AMPRA)*. Los Angeles: University of California.
- Gironzetti, E., Attardo, S., Pickering, L., Huang, M., Menjo, S., & Zhang, Y. (2015). Multimodal analysis of humour in conversation. *Proceedings from the 27th Conference of the International Society of Humor Studies (ISHS)*, Oakland: Holy Names University.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- 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. doi: 10.1016/j.pragma.2015.10.002

- Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze. *Sociological Inquiry*, 50(3/4), 272-302.
<https://doi.org/10.1111/j.1475-682X.1980.tb00023.x>
- Goodwin, C. (2013). The cooperative, transformative organisation of human action and knowledge. *Journal of Pragmatics*, 46(1), 8-23.
<https://doi.org/10.1016/j.pragma.2012.09.003>
- Goodwin, C., & Goodwin, M. H. (1992). Gesture and co-participation in the activity of searching for a word. *Semiotica*, 62(1/2), 51-75.
<https://doi.org/10.1515/semi.1986.62.1-2.51>
- Goodwin, M.H. & Cekaite, A. (2013). Calibration in directive/response sequences in family interaction. *Journal of Pragmatics*, 46(1), 122-138.
<https://doi.org/10.1016/j.pragma.2012.07.008>
- Grady, J., Oakley, T., & Coulson, S (1999). Conceptual blending and metaphor. In R. Gibbs (Ed.), *Metaphor in Cognitive Linguistics* (pp. 420-440). Amsterdam: John Benjamins.
- Grice, P. (1989). *Studies in the way of words*. Cambridge: Harvard University Press.
- Guaitella, I., Santi, S., Lagrue, B., & Cave, C. (2009). Are Eyebrow Movements Linked to Voice Variations and Turn-taking in Dialogue? An Experimental Investigation. *Language and Speech*, 52(2/3), 207-222. doi: 10.1177/0023830909103167

- Hadar, U., Steiner, T.J., Grant, E.C., & Clifford Rose, F. (1983). Kinematics of head movements accompanying speech during conversation. *Human Movement Science*, 2, 35-46. [https://doi.org/10.1016/0167-9457\(83\)90004-0](https://doi.org/10.1016/0167-9457(83)90004-0)
- Hadar, U., Steiner, T.J., & Grant, E.C. (1984). The timing of shifts of head postures during conversation. *Human Movement Science*, 3, 237-245. [https://doi.org/10.1016/0167-9457\(84\)90018-6](https://doi.org/10.1016/0167-9457(84)90018-6)
- Hadar, U., Steiner, T.J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9, 214-228. <https://doi.org/10.1007/BF00986881>
- Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford: Oxford University Press on Demand.
- Halliday, M. (1976). *Intonation and grammar in British English*. The Hague: Mouton.
- Halliday, M. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M. (2004). *An Introduction to Functional Grammar* (3rd ed. revised by C. M. I. M Matthiessen). London: Arnold.
- Hampe, B. (2005a). Image schemas in Cognitive Linguistics: Introduction. In B. Hampe (Ed.), *From perception to meaning: image schemas in cognitive linguistics* (pp. 1-14). Berlin: Mouton de Gruyter.

- Hampe, B. (Ed.). (2005b). *From perception to meaning: image schemas in cognitive linguistics*. Berlin: Mouton de Gruyter.
- Harris, C., & Alvarado, N. (2005). Facial expressions, smile types, and self-report during humour, tickle, and pain. *Cognition & Emotion*, 19(5), 655-669. doi: 10.1080/02699930441000472
- Hay, J. (2001). The pragmatics of humor support. *Humor: International Journal of Humor Research*, 14(1), 55-82. <https://doi.org/10.1515/humr.14.1.55>
- Hempelmann, C., & Salvatore, A. (2011). Resolutions and their incongruities: Further thoughts on Logical Mechanisms. *Humor: International Journal of Humor Research*, 24(2), 125-149. doi: 10.1515/HUMR.2011.008
- Heylen, D. (2008). Listening heads. In Wachsmuth, & Knoblich, G. (Eds.), *Modeling communication* (pp. 241-259). Berlin/Heidelberg: Springer-Verlag.
- Hirst, D., & Di Cristo, A. (Eds.) (1999). *Intonation: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Holmes, J. (2000). Politeness, power and provocation: how humor functions in the workplace. *Discourse Studies*, 2 (2), 159-185. <https://doi.org/10.1177/1461445600002002002>
- Holmes, J., & Marra, M. (2002). Humour as a discursive boundary marker in social interaction. In A. Duszak (Ed.), *Us and Others: Social Identities across Languages, Discourses, and Cultures* (pp.377-400). Amsterdam: John Benjamins.

- Holt, E., & Glenn, P.J. (2013). *Studies of laughter in interaction*. London: Bloomsbury Academy.
- Hoskin, J. (director), & Colbert, S., Spyra, J., Stack, B., & Dinello, P. (writers). (2015-). *The Late Show with Stephen Colbert* [Television show]. Retrieved from <https://www.youtube.com/channel/UCMtFAi84ehTSYSE9XoHefiq>
- Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15(3), 495-514. <https://doi.org/10.3758/PBR.15.3.495>
- Ibarretxe, I., & Valenzuela, J. (Eds.) (2012). *Lingüística Cognitiva*. Barcelona: Anthropos.
- Ikeda, K., & Bysouth, D. (2013). Laughter and turn-taking: Warranting next speakership in multiparty interactions. In E. Holt, & P. Glenn (Eds.), *Studies of laughter in interaction* (pp. 36-64). London: Bloomsbury Academic.
- Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, 57, 233-243. doi: 10.1016/j.specom.2013.06.008
- Jannedy, S., & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, 3, 199-244. Retrieved from https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/781/file/isis03_7jannedy_etal.pdf

- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: The University of Chicago Press.
- Johnson, M. (2005). The philosophical significance of image schemas. In B. Hampe (Ed.), *From perception to meaning: image schemas in cognitive linguistics*. Berlin: Mouton de Gruyter.
- Kaal, A. A. (2012). *Metaphor in Conversation*. Oisterwijk, The Netherlands: Uitgeverij BOXPress.
- Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207-227). The Hague: Mouton and Co.
- Kendon, A. (2002). Some uses of the headshake. *Gesture* 2(2), 147-182. <https://doi.org/10.1075/gest.2.2.03ken>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, England: Cambridge University Press.
- Kendon, A. (2013). Exploring the utterance roles of visible bodily action: A personal account. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 7-28). Berlin: de Gruyter Mouton.
- Khottoff, H. (2007). Oral genres of humour. On the dialectic of genre knowledge and creative authoring. *Journal of Pragmatics* 17(2), 263-296. doi: 10.1075/prag.17.2.04kot

- Kim, J., Cvejic, E., & Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, 57, 317-330. doi:10.1016/j.specom.2013.06.003
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill, (Ed.), *Language and Gesture* (pp. 162-185). Cambridge: Cambridge University Press.
- Kohler, K. (2006). Paradigms in experimental prosodic analysis: from measurement to function. In S. Sudhoff, ... J. Schließer (Eds.), *Methods in Empirical Prosody Research* (Vol.3, pp. 123-152). Berlin/New York: Walter de Gruyter.
- Kopp, S., Bergmann, K., & Kahl, S. (2013). A spreading activation model of the semantic coordination of speech and gesture. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (COGSCI 2013)*. Retrieved from <https://pdfs.semanticscholar.org/aada/1bfeac1f202044c395f8b32fed087786043a.pdf>
- Kousidis, S., Malisz, Z., Wagner, P., & Schlangen, D. (2013). Exploring annotation of head gesture forms in spontaneous human interaction. *TiGeR 2013, Tilburg Gesture Research Meeting*. Retrieved from https://pub.uni-bielefeld.de/download/2567303/2606548/tiger2013_KousidisEtAl.pdf
- Kousidis, S., Malisz, Z., Wagner, P., Schlangen, D. Ladewig, S. (2014). Recurrent gestures. In C. Muller, A. Cienki, E. Fricke, S. Ladewig, D. McNeil, & J.

- Bressemer (Eds.), *Handbooks of Linguistic and Communication Science. Body-Language-Communication* (Vol. 2, pp. 1558-1574). Berlin/Boston: De Gruyter.
- 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. doi: 10.1016/j.jml.2007.06.005
- Krauss, R.M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell, L. Messing (Eds.), *Gesture, Speech, and Sign* (pp. 93-116). Oxford: Oxford University Press.
- Kress, G. (2015). Semiotic work: Applied Linguistics and a social semiotic account of Multimodality. *AILA Review*, 28, 49-71. doi: [10.1075/aila.28.03kre](https://doi.org/10.1075/aila.28.03kre)
- Kress, G., & van Leeuwen, T. (2001). *Multimodal Discourse: The modes and media of contemporary communication*. London: Arnold. New York: Oxford University Press.
- Kyratzis, S. (2003). Laughing metaphorically: Metaphor and humour in discourse. In *8th International Cognitive Linguistics Conference* (pp. 20-25). Citeseer. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.9689&rep=rep1&type=pdf>
- Ladewig, S. (2014a). Creating multimodal utterances. The linear integration of gestures into speech. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J.

- Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1662-1677). Berlin: de Gruyter Mouton.
- Ladewig, S. (2014b). Recurrent gestures. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1558-1574). Berlin: de Gruyter Mouton.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: Chicago University Press.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202-251). Cambridge: Cambridge University Press.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G. & Turner, M. (1989). *More Cool than Reason: A Field Guide to Poetic Metaphor*. Chicago: University of Chicago Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Langacker, R. (1987). *Foundations of cognitive grammar, Vol. I: Theoretical prerequisites*. Stanford/California: Stanford University Press.
- Langacker, R. (1991). *Foundations of cognitive grammar, Vol. II: Descriptive application*. Stanford/California: Stanford University Press.

- Langacker, R. W. (2001). Discourse in Cognitive Grammar. *Cognitive Linguistics*, 12(2), 143-188. doi: 10.1515/cogl.12.2.143
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Lapaire, J.-R. (2011). Grammar, gesture and cognition: Insights from multimodal utterances and applications for gesture analysis. *Visnyk of Lviv University Philology Series*, 52, 87-107. Retrieved from [https://www.researchgate.net/publication/298790338 Grammar gesture and cognition insights from multimodal utterances and applications for gesture analysis](https://www.researchgate.net/publication/298790338_Grammar_gesture_and_cognition_insights_from_multimodal_utterances_and_applications_for_gesture_analysis)
- Lausber, H. (2013). Neuropsychology of gesture production. In C. Müller, A. Cienki, S. Ladewig, D. McNeill & S. Teßendorf (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1, pp. 168–182). Berlin: de Gruyter Mouton.
- Lee, D. (2001). *Cognitive Linguistics: An Introduction*. Oxford: Oxford University Press Australia & New Zealand.
- Lee, J., & Marsella, S. (2006). Nonverbal behavior generator for embodied conversational agents. *Proceedings of Sixth International Conference on Intelligent Virtual Agents (IVA)*, 243-255. Marine del Rey, USA, August 21-23. doi: 10.1007/11821830_20
- Lee, J., & Marsella, S. (2010). Predicting speaker head nods and the effects of affective information. *IEEE Transactions on Multimedia*, 12(6), 552-562. doi: [10.1109/TMM.2010.2051874](https://doi.org/10.1109/TMM.2010.2051874)

- Lévi-Strauss, C. (1972). *Structural Anthropology*. London: Penguin.
- Loehr, D. (2004). *Gesture and intonation* (Unpublished doctoral dissertation). Georgetown University. Retrieved from http://faculty.georgetown.edu/loehrd/pubs_files/Loehr04.pdf
- Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1), 71-89. doi: 10.1515/lp-2012-0006
- Loehr, D. (2014). Gesture and prosody. In C. Muller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer, & D. McNeill (Eds.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2, pp. 1381-1391). Berlin: de Gruyter Mouton.
- Magno-Caldogneto, E., & Poggi, I. (1997). Dall'analisi della multimodalità quotidiana alla costruzione di agenti animati con facce parlanti ed espressive. *Atti delle XI Giornate di Studio del Gruppo di Fonetica Sperimentale. Multimodalità e Multimedialità nella Comunicazione*, 47-55. Padova, 29 novembre-1 dicembre 2000. Padova: Unipress
- Marhula, J., & Rosiński, M. (2014). Identifying metaphor in spoken discourse: Insights of applying MIPVU to radio talk data. *Zeszyt*, 85. 32-42. Retrieved from https://www.academia.edu/10084006/Identifying_metaphor_in_spoken_discourse_Insights_from_applying_MIPVU_to_radio_talk_data

- Maynard, S. (1987). Interactional functions of a nonverbal sign Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics*, 11(5), 589-606. [https://doi.org/10.1016/0378-2166\(87\)90181-0](https://doi.org/10.1016/0378-2166(87)90181-0)
- McClave, E. (1998). Pitch and manual gestures. *Journal of Psycholinguistic Research*, 27(1), 69-89. <https://doi.org/10.1023/A:1023274823974>
- McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32(7), 855-858. [https://doi.org/10.1016/S0378-2166\(99\)00079-X](https://doi.org/10.1016/S0378-2166(99)00079-X)
- McGraw, P., & Warren, C. (2010). Benign Violations: Making Inmoral Behavior Funny. *Psychological Science*, 21(8), 1141-1149. doi: 10.1177/0956797610376073
- McNeill, D. (1985). So You Think that Gestures are Nonverbal? *Psychological Review*. (92)3, 350-371. doi: 10.1037/0033-295X.92.3.350
- McNeill, D. (1992) *Hand and mind: what gestures reveal about thought*. Chicago & London: The University of Chicago Press.
- McNeill, D. (Ed.) (2000), *Language and Gesture*. Cambridge: Cambridge University Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- McNeill, D. (2007). Gesture and thought. In *The Summer Institute on Verbal and Non-verbal Communication and the Biometrical Principle*, Sept. 2-12, 2006, Vietri sul Mare (Italy), organized by Anna Esposito. Retrieved from http://mcneilllab.uchicago.edu/pdfs/dmcn_vietri_sul_mare.pdf

- McNeill, D. (2008). Imagery for speaking. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özçaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 517-530). London: Taylor & Francis.
- McNeill, D. (2013). Gesture as a window onto mind and brain, and the relationship to linguistic relativity and ontogenesis. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, S. (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 28-54) . Berlin: de Gruyter Mouton.
- McNeill, D. (2014). Gesture–speech unity: Phylogenesis, ontogenesis, and microgenesis. *Language, Interaction and Acquisition*, 5(2), 137-184. doi: 10.1075/lia.5.2.01mcn
- McNeill, D., & Levy, E. T. (1993). Cohesion and gesture. *Discourse Processes*, 16(4), 363-386. doi: 10.1080/01638539309544845
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 115-154). Amsterdam/ Philadelphia: John Benjamins.
- Mittelberg, I. (2017). Embodied frames and scenes. *Gesture*, 16(2), 203-224. DOI: 10.1075/pc.19.2.08ber
- Mondada, L. (2013). Conversation analysis: Talk and bodily resources for the organisation of social interaction. In C. Müller, A. Cienki, S. Ladewig, D. McNeill & S. Teßendorf (Eds.), *Body–Language–Communication: An*

- International Handbook on Multimodality in Human Interaction* (Vol. 1, pp. 228 - 226). Berlin: de Gruyter Mouton.
- Moran, J. M., Wig, G. S., Adams, R. B., Janata, P., & Kelley, W. M. (2004). Neural correlates of humor detection and appreciation. *NeuroImage*, 21(3), 1055-1060. doi: 10.1016/j.neuroimage.2003.10.017
- Morreal, J. (1983). *Taking Laughter Seriously*. New York: State University of New York Press.
- Morreal, J. (2009a). *Comic Relief: A Comprehensive Philosophy of Humour*. Chichester, U.K.; Malden, MA: Wiley-Blackwell.
- Morreal, J. (2009b). Humor as cognitive play. *Journal of Literary Theory*, 3(2), 241-260. <http://dx.doi.org/10.1515/JLT.2009.014>
- Muecke, D. (1978). Irony markers. *Poetics*, 7(4), 363-375. [https://doi.org/10.1016/0304-422X\(78\)90011-6](https://doi.org/10.1016/0304-422X(78)90011-6)
- Muller, C. (2009). Gesture and Language. In K. Malmkjaer (Ed.), *Routledge's Linguistics Encyclopedia* (pp. 214-217). London: Routledge
- Muller, C. (2013a). Gestures as a medium of expression: The linguistic potential of gestures. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 200-217). Berlin: de Gruyter Mouton.
- Muller, C. (2013b). Introduction. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An*

- International Handbook on Multimodality in Human Interaction* (Vol.1 pp. 1-6). Berlin: de Gruyter Mouton.
- Muller, C. & Cienki, A. (2009). Words, gestures and beyond: Forms of multimodal metaphor in the use of spoken language. In Forceville, Ch. & Urios-Aparisi, E. (Eds.), *Multimodal Metaphor* (pp. 297-328). Berlin: de Gruyter Mouton.
- Müller, C. Cienki, A., Ladewig, S., McNeill, D & Teßendorf, S. (Eds.). (2013). *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1). Berlin: de Gruyter Mouton.
- Muller, C., Ladewig, S., Bressemer, J. (2013). Gestures and speech from a linguistic perspective: A new field and its history. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 55-81). Berlin: de Gruyter Mouton.
- Muller, C., Ladewig, S., Cienki, A., Fricke, E., Bressemer, J. & McNeill, D. (2014). *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*. (Vol. 2). Berlin: de Gruyter Mouton.
- Norricks, N. R. (1993). *Conversational joking*. Bloomington: Indiana University Press.
- Norris, S. (2004a). *Analysing Multimodal Interaction: A Methodological Framework*. London and New York: Routledge.
- Norris, S. (2004b). Multimodal discourse analysis: A methodological framework. In P. LeVine, & R. Scollon (Eds.), *Discourse and Technology: Multimodal*

- Discourse Analysis* (pp. 101-115). Washington: Georgetown University Press.
- Norris, S. (2013). Multimodal (Inter)action Analysis: An Integrative methodology. In C. Müller, A. Cienki, S. Ladewig, D. McNeill & S. Teßendorf (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1, pp. 271-282). Berlin: de Gruyter Mouton.
- Norris, S. (2014). Modal Density and Modal Configurations. In C. Jewitt, (Ed.). *The Routledge Handbook of Multimodal Analysis* (2nd edition). London/New York: Routledge.
- Nuyts, J. (2007). Cognitive linguistics and functional linguistics. In D. Geeraerts, H. Cuyckens (Eds.), *Handbook of cognitive linguistics* (pp. 543-565). Oxford: Oxford University Press.
- O'Halloran, K. (2004). *Multimodal discourse analysis: Systemic functional perspectives*. London: Continuum.
- Oakley, T. (2007). Image Schemas. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 214-235). Oxford: Oxford University Press.
- Pelachaud, C., Badler, N., & Steedman, M. (1996). Generating facial expressions for speech. *Cognitive Science*, 20, 1-46. doi: 10.1207/s15516709cog2001_1

- Pennock-Speck, B. & Del Saz-Rubio, M. (2009). Voice-overs in standardized English and Spanish commercials. *Atlantis. Journal of the Spanish Association of English-American Studies* 31(1), 117 – 127.
- Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A., & Attardo, S. (2009). Prosodic markers of saliency in humorous narratives. *Discourse Processes*, 46(6), 517-540. doi: 10.1080/01638530902959604
- Peirce, C. P. (1965). Basic Concepts of Peircean Sign Theory. In Gottdiener, M., Boklund-Lagopoulou, K. & Lagopoulos, A.P. (2003). *Semiotics*. London: Sage Publications.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. (Unpublished doctoral dissertation). MIT. Retrieved from <https://dspace.mit.edu/handle/1721.1/16065#files-area>
- Pinar Sanz, M.J. (Ed.). (2015). *Multimodality and Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Poggi, I. (2013). Minds, hands, face and body: A sketch of a goal and belief view of multimodal communication. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 627-647). Berlin: de Gruyter Mouton.
- Poggi, I. & Pelachaud, C. (1998). Performative faces. *Speech Communication*, 26(1/2), 5-21. [https://doi.org/10.1016/S0167-6393\(98\)00047-8](https://doi.org/10.1016/S0167-6393(98)00047-8)

- Poggi, I., D'Errico, F., & Vincze, L. (2010). Types of Nods. The Polysemy of a Social Signal. In LREC. Retrieved from http://www.academia.edu/download/4683379/596_paper.pdf
- Poyatos, F. (2002). *Nonverbal Communication Across Disciplines*. Amsterdam: John Benjamins Publishing Co.
- Provine, R. (2000). *Laughter. A scientific Investigation*. New York: Viking.
- Purandare, A., & Litman, D. (2006). Humor: prosody analysis and automatic recognition for F* R* I* E* N* D* S. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 208-215. Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1610107>
- Raskin, V. (1985). *Semantic Mechanisms of Humor*. Dordrecht: D. Reidel Publishing company.
- Raskin, V., Hempelman, C., & Taylor, J (2009). How to understand and assess a theory: The evolution of the SSTH into the GTVH and now into the OSTH. *Journal of Literary Theory*, 3(2), 285 -312. doi: 10.1515/JLT.2009.016
- Ritchie, G. (2006). Reinterpretation and viewpoints. *Humor: The International Journal of Humor Research*, 19(3), 251-270. <https://doi.org/10.1515/HUMOR.2006.014>
- Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, 29(5), 483-495. <https://doi.org/10.1023/A:1005120109296>

- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573-605. [DOI: 10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
- Ruiz de Mendoza Ibáñez, F. J. (2014). Mapping concepts: Understanding figurative thought from a cognitive-linguistic perspective. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 27(1), 187-207. doi: org/10.1075/resla.27.1.08rui
- Ruiz de Mendoza, F. & Galera, A. (2014). *Cognitive modeling: A linguistic perspective*. Amsterdam and Philadelphia: John Benjamins
- Saussure, F. (1993). *Cours de linguistique générale*. Paris: Payot.
- Schaffer, R. (1982). Are there consistent vocal clues for irony?. In C. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior*. Chicago, IL: Chicago Linguistic Society: 204-210.
- Schegloff, E.A. (1987). Analyzing single episodes of interaction. An exercise in conversation analysis. *Social Psychology Quarterly*, 50(2), 101-114. <http://dx.doi.org/10.2307/2786745>
- Slobin, D. (1987). Thinking for speaking. *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society* (pp. 435-445). Berkeley. <http://dx.doi.org/10.3765/bls.v13i0.1826>
- Smoski, M.J., & Bachorowski, J.A. (2003). Antiphonal laughter between friends and strangers. *Cognition & Emotion*, 17(2), 327-340. doi: [10.1080/02699930302296](https://doi.org/10.1080/02699930302296)

- So, W.C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cognitive Science* 33(1), 115-125. doi: [10.1111/j.1551-6709.2008.01006.x](https://doi.org/10.1111/j.1551-6709.2008.01006.x)
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Steen, G. J. (2011). Metaphor, language, and discourse processes. *Discourse Processes*, 48(8), 585-591. doi: 10.1080/0163853X.2011.606424
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU* (Vol. 14). Amsterdam: John Benjamins Publishing.
- Stivers, T., & Sidnell, J. (2005). Introduction: multimodal interaction. *Semiotica*, 156 (1/4), 1-20. Retrieved from: http://www.sscnet.ucla.edu/soc/faculty/stivers/Publications_files/Stivers_Sidnell_intro.pdf
- Swerts, M., & Kraemer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, 36(2), 219-238. doi: 10.1016/j.wocn.2007.05.001
- Swerts, M., & Kraemer, E. (2010). Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics*, 38, pp. 197-206. doi:10.1016/j.wocn.2009.10.002

Tabacaru, S. (2014). *Humorous implications and meanings: a multi-modal approach to sarcasm in interactional humor* (Unpublished doctoral dissertation). Université Charles de Gaulle - Lille 3. Retrieved from <http://www.theses.fr/2014LIL30015>

Tabacaru, S., & Lemmens, M. (2014). Raised eyebrows as gestural triggers in humour: The case of sarcasm and hyper-understanding. *The European Journal of Humour Research*, 2(2), 11-31. <http://dx.doi.org/10.7592/EJHR2014.2.2.tabacaru>

Tabacaru, S. & Feyaerts, K. (2016). The power of metonymy in humour: stretching contiguous relations across different layers of meaning. *European Journal of Humour Research*, 4(2), 1-18, doi: 10.7592/EJHR2016.4.2.tabacaru

Talmy, L. (2000a). *Toward a cognitive semantics*. (Vol. 1): *Concept structuring systems*. Cambridge, MA: MIT Press.

Talmy, L. (2000b). *Toward a cognitive semantics*. (Vol. 2): *Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Teßendorf, S. (2013). Emblems, quotable gestures, or conventionalised body movements. In C. Müller, A. Cienki, S. Ladewig, D. McNeill, & S. Teßendorf, (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction* (Vol.1, pp. 82-100). Berlin: de Gruyter Mouton.

Tsakona, V., & Chovanec, J. (Eds.). (2018). *The Dynamics of Interactional Humour*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Turner, M. (2015). Blending in language and communication. In E. Dabrowska, & D. Divjak (Eds.), *Handbook of Cognitive Linguistics*, (pp. 211-232). Berlin: De Gruyter Mouton.
- Urios-Aparisi, E., & Wagner, M. (2011). Prosody of humor in *Sex and the City*. *Pragmatics & Cognition*, 19(3), 507-529. doi: 10.1075/pc.19.3.06uri
- Veale, T. (2015). The humour of exceptional cases: jokes as compressed thought experiments. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 69-90). Berlin/Boston: Walter de Gruyter.
- Veale, T., Feyaerts, K. & Brône, G. (2006). The cognitive mechanisms of adversarial humor. *Humor: International Journal of Humor Research*, 19(3), 305-338. doi: 10.1515/HUMOR.2006.016
- Veale, T. Brône, G., Feyaerts, K. (2015). Humour as the *killer app* of language. In G. Brône, K. Feyaerts, & T. Veale (Eds.), *Cognitive Linguistics and Humour Research* (pp. 1-12). Berlin/Boston: Walter de Gruyter.
- Velasco-Sacristán, M. (2010). Metonymic grounding of ideological metaphors: Evidence from advertising gender metaphors. *Journal of Pragmatics*, 42(1), 64-96. <https://doi.org/10.1016/j.pragma.2009.05.019>
- Velasco-Sacristán, M., Fuertes-Oliveira, P., & Samaniego-Fernández, E. (2005). La metáfora cultural de género en el discurso publicitario: ejemplo de proyección metafórica de origen metonímico. In J.L. Otal Campo, I. Navarro Ferrando, & B. Bellés Fortuño (Eds.), *Cognitive and Discourse*

- Approaches to Metaphor and Metonymy*, (pp. 157-173). Castellón de la Plana: Universitat Jaume I.
- Velasco-Sacristán, M., & Fuertes-Olivera, P. (2006). Olfactory and olfactory-mixed metaphors in print ads and perfume. *Annual Review of Cognitive Linguistics*, 4(1), 217-252. doi: 10.1075/arcl.4.09vel
- Verhagen, A. (2007). Construal and Perspectivization. In D. Geraeerts, & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 48-81). Oxford: Oxford University Press.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209-232. doi:10.1016/j.specom.2013.09.008
- Weeks, M. (2016). The enigma of solitary laughter. *European Journal of Humour Research*, 4(3), pp. 76-87. <http://dx.doi.org/10.7592/EJHR2016.4.3.weeks>
- Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press.
- Wennerstrom, A. (2011). Rich pitch: The humorous effects of deaccent and L+H* pitch accent. *Pragmatics & Cognition*, 19 (2), 310-332. <http://dx.doi.org/10.1075/pc.19.2.07wen>
- Wennerstrom, A., & Siegel, A. F. (2003). Keeping the Floor in Multiparty Conversations: Intonation, Syntax, and Pause. *Discourse Processes*, 36(2), 77-107. doi: 10.1207/S15326950DP3602_1

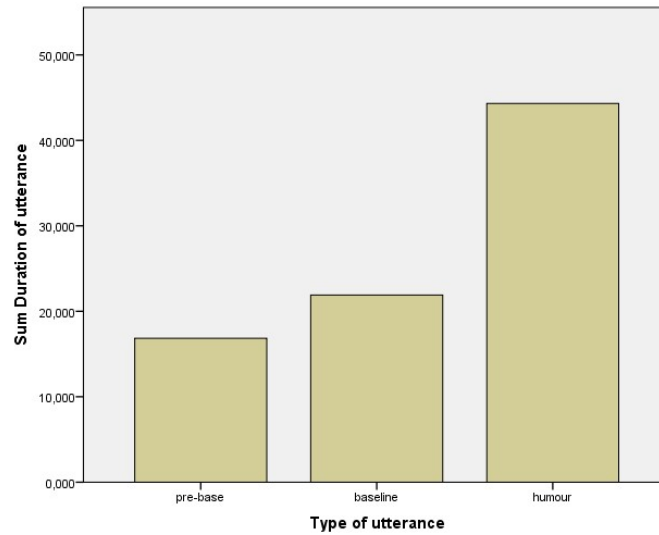
- Williams, R. F. (2008). Gesture as a conceptual mapping tool. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture [Gesture Studies 3]*, pp. 55-92.
- Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn, & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 607-632). Oxford: Blackwell.
- Yamaguchi, H. (1988). How to pull strings with words. Deceptive violations in the garden-path joke. *Journal of Pragmatics*, 12(3), pp. 323-337.
[https://doi.org/10.1016/0378-2166\(88\)90036-7](https://doi.org/10.1016/0378-2166(88)90036-7)
- Yus, F. (2003). Humor and the search for relevance. *Journal of Pragmatics*, 35(9), 1295-1331. doi: 10.1016/S0378-2166(02)00179-0
- Yus, F. (2016). *Humour and Relevance*. Amsterdam: John Benjamins Publishing Company.
- Ziemke, T., Zlatev, J., & Roslyn, F. (Eds.). (2007). *Body, Language, and Mind. Vol 1, Embodiment*. Berlin: Mouton de Gruyter.
- Zlatev, J. (2008). Intersubjectivity, Mimetic Schemas and the Emergence of Language. In C. Fabrice, & L. Kaufmann (Eds.), *Culture and Society: Some Viewpoints of Cognitive Scientists, Intellectica*, 46/47, (pp. 123-151). Retrieved from
http://intellectica.org/SiteArchives/archives/n46_47/46-47-6-Zlatev.pdf
- Zlatev, J. (2012). Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics*, 4(1), 2-24. Retrieved from
<https://pdfs.semanticscholar.org/6ad4/2975f57eba981bab77f9512a4cd1ca1d2b43.pdf>

Annex 1: Prosodic analysis

The following graphs and tables are presented for each interview:

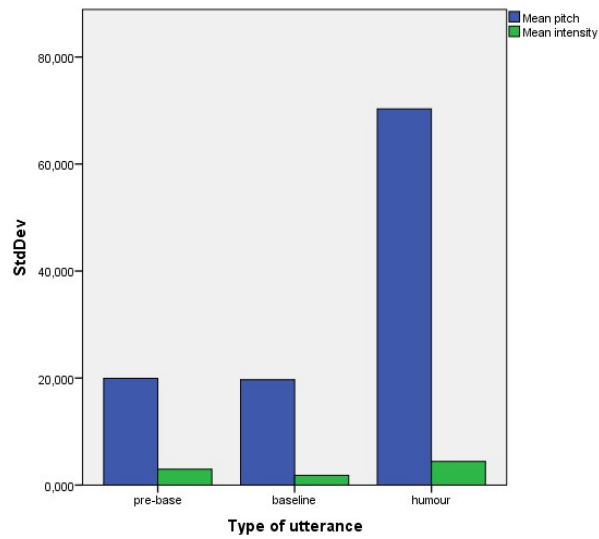
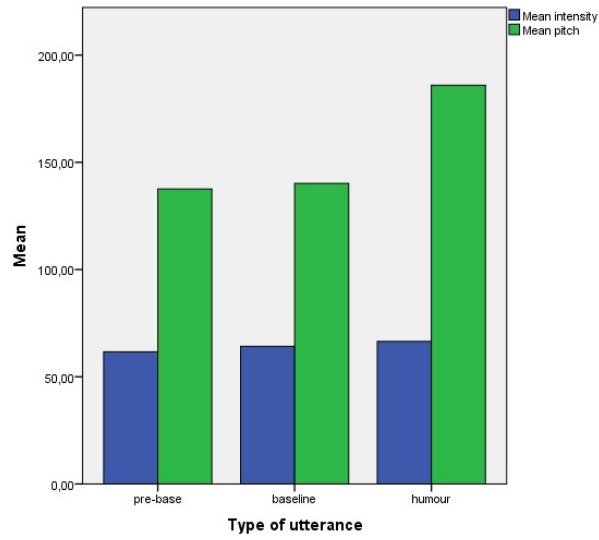
1. Case summary: number and duration of utterances per type: table and graph.
2. Report on mean pitch and intensity values and standard deviation (SD) per type of utterance: table and graph.
3. Normality tests: histogram and one-sample Kolmogorov-Smirnov Test.
4. T-test and/or nonparametric tests as appropriate.

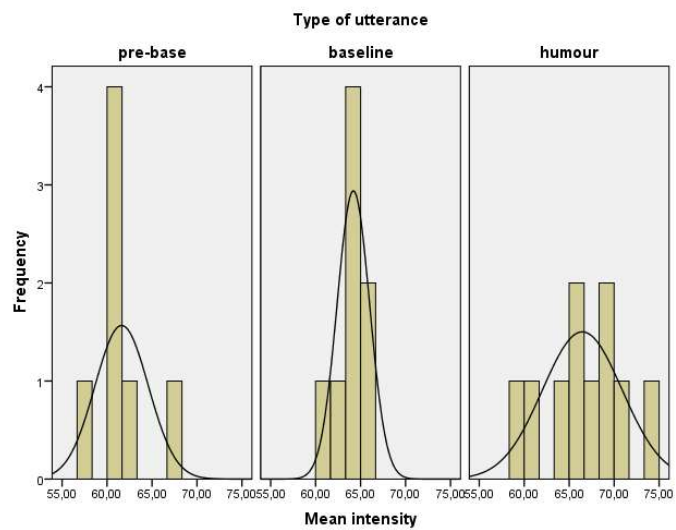
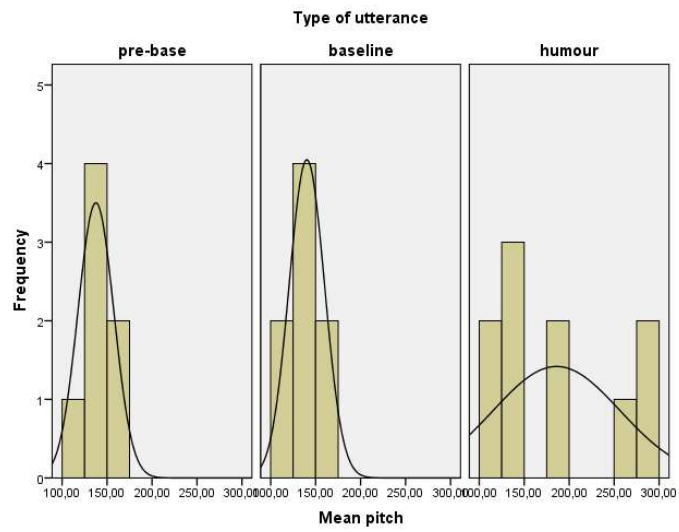
INTERVIEW 1: ALEC BALDWIN



Report

Type of utterance	Utterance ID	Mean pitch	Mean intensity
pre-base	N	8	7
	Mean	137,5686	61,6214
	Std. Deviation	19,93694	2,97154
baseline	N	8	8
	Mean	140,2113	64,1925
	Std. Deviation	19,70421	1,80954
humour	N	13	10
	Mean	185,9670	66,4600
	Std. Deviation	70,30261	4,42888
Total	N	29	25
	Mean	157,7736	64,3796
	Std. Deviation	51,17725	3,81472





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		25	25
Normal Parameters ^{a,b}	Mean	157,7736	64,3796
	Std. Deviation	51,17725	3,81472
Most Extreme Differences	Absolute	,233	,118
	Positive	,233	,118
	Negative	-,176	-,079
Kolmogorov-Smirnov Z		1,164	,592
Asymp. Sig. (2-tailed)		,133	,874

- a. Test distribution is Normal.
- b. Calculated from data.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	pre-base	7	137,5686	19,93694	7,53545
	baseline	8	140,2113	19,70421	6,96649
Mean intensity	pre-base	7	61,6214	2,97154	1,12314
	baseline	8	64,1925	1,80954	,63977

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,162	,694	-,258	13	,801	-2,64268	10,25367	-24,79438	19,50902	
	Equal variances not assumed			-,258	12,692	,801	-2,64268	10,26231	-24,86784	19,58248	
Mean intensity	Equal variances assumed	,555	,469	-2,056	13	,060	-2,57107	1,25056	-5,27274	,13060	
	Equal variances not assumed			-1,989	9,654	,076	-2,57107	1,29257	-5,46515	,32300	

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	baseline	8	140,2113	19,70421	6,96649
	humour	10	185,9670	70,30261	22,23164
Mean intensity	baseline	8	64,1925	1,80954	,63977
	humour	10	66,4600	4,42888	1,40054

Independent Samples Test

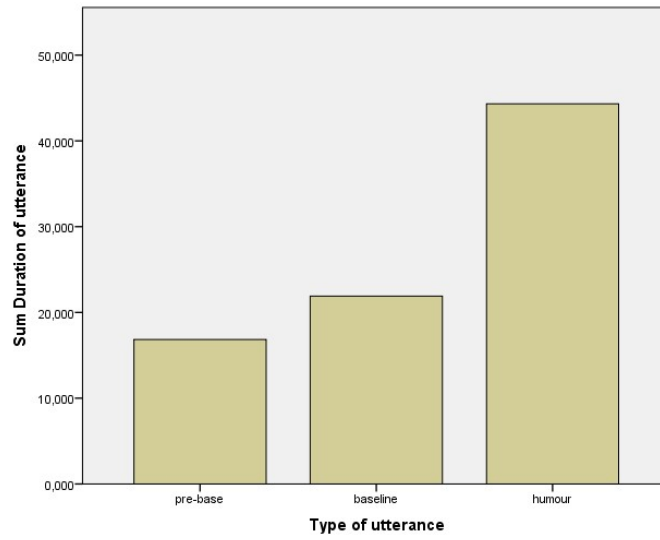
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	10,118	,006	-1,776	16	,095	-45,75575	25,76332	-100,37155	8,86005	
	Equal variances not assumed			-1,964	10,721	,076	-45,75575	23,29759	-97,19646	5,68496	
Mean intensity	Equal variances assumed	6,236	,024	-1,354	16	,195	-2,26750	1,67477	-5,81785	1,28285	
	Equal variances not assumed			-1,473	12,451	,166	-2,26750	1,53974	-5,60889	1,07389	

INTERVIEW 2: ALISON JANNEY

Case Summaries^a

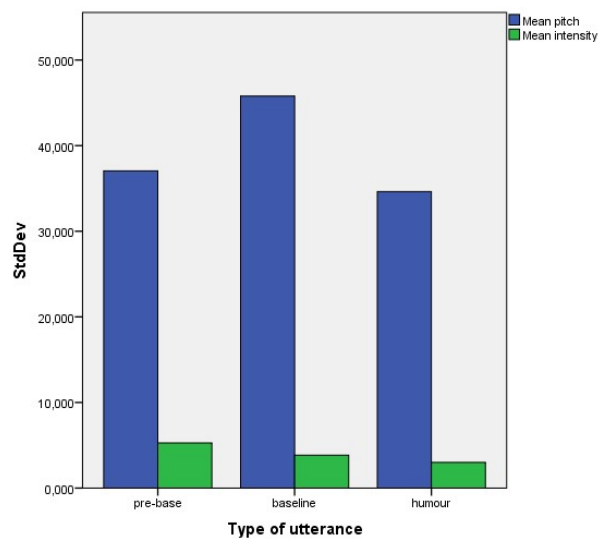
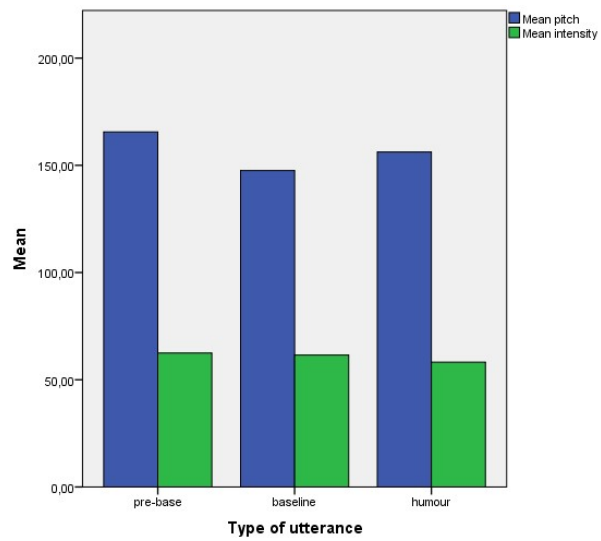
			Duration of utterance	
Type of utterance	pre-base	1	4,391	
		2	,900	
		3	1,445	
		4	3,327	
		5	3,309	
		Total	N	5
			Sum	13,372
	baseline	1	4,983	
		2	4,590	
		3	3,108	
		4	2,072	
		5	2,581	
		Total	N	5
			Sum	17,334
	humour	1	3,490	
		2	8,360	
3		5,918		
4		5,231		
5		3,570		
6		2,637		
Total	N	6		
		Sum	29,206	
Total	N	16		
	Sum	59,912		

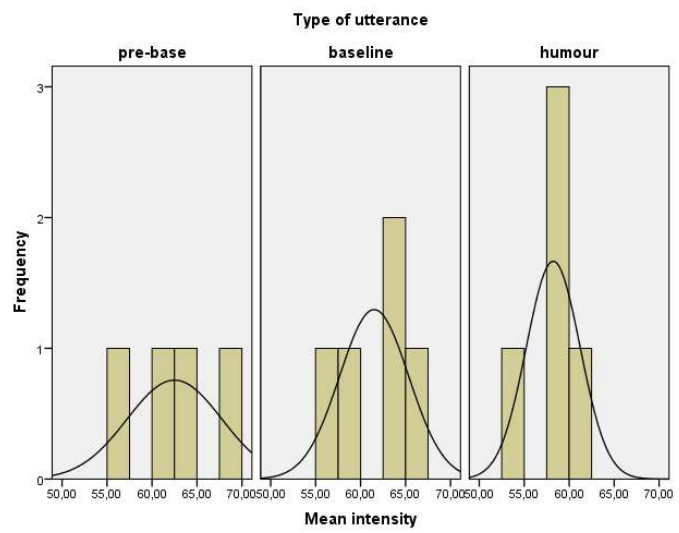
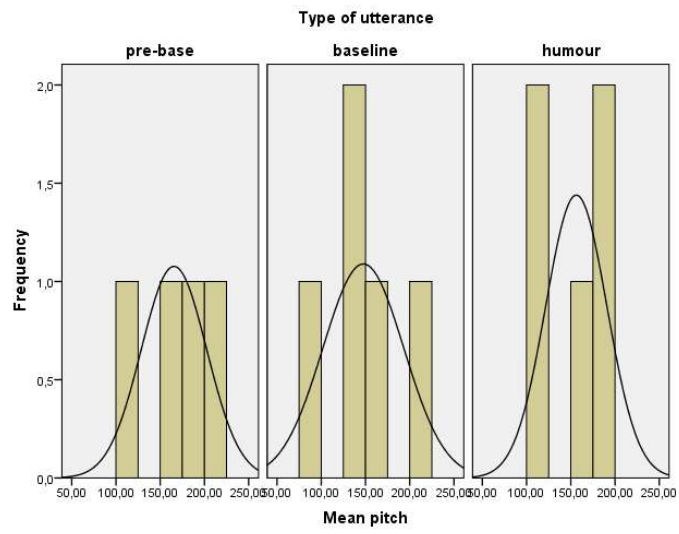
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	165,5650	62,4950
	N	4	4
	Std. Deviation	37,04350	5,27290
baseline	Mean	147,5340	61,5020
	N	5	5
	Std. Deviation	45,78989	3,84737
humour	Mean	156,2040	58,2260
	N	5	5
	Std. Deviation	34,63924	2,99356
Total	Mean	155,7821	60,6157
	N	14	14
	Std. Deviation	37,23821	4,16093





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		14	14
Normal Parameters ^{a,b}	Mean	155,7821	60,6157
	Std. Deviation	37,23821	4,16093
Most Extreme Differences	Absolute	,098	,110
	Positive	,098	,097
	Negative	-,095	-,110
Kolmogorov-Smirnov Z		,366	,412
Asymp. Sig. (2-tailed)		,999	,996

- a. Test distribution is Normal.
- b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	pre-base	4	165,5650	37,04350	18,52175
	baseline	5	147,5340	45,78989	20,47786
Mean intensity	pre-base	4	62,4950	5,27290	2,63645
	baseline	5	61,5020	3,84737	1,72060

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,132	,727	,636	7	,545	18,03100	28,35133	-49,00923	85,07123	
	Equal variances not assumed			,653	6,987	,535	18,03100	27,61156	-47,28471	83,34671	
Mean intensity	Equal variances assumed	,299	,601	,328	7	,753	,99300	3,02794	-6,16693	8,15293	
	Equal variances not assumed			,315	5,369	,764	,99300	3,14823	-6,93523	8,92123	

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	baseline	5	147,5340	45,78989	20,47786
	humour	5	156,2040	34,63924	15,49114
Mean intensity	baseline	5	61,5020	3,84737	1,72060
	humour	5	58,2260	2,99356	1,33876

Independent Samples Test

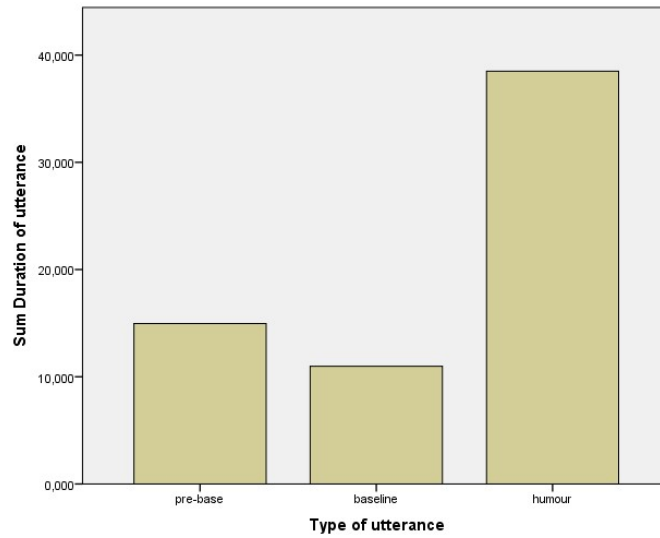
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,086	,777	-,338	8	,744	-8,67000	25,67719	-67,88171	50,54171	
	Equal variances not assumed			-,338	7,449	,745	-8,67000	25,67719	-68,65326	51,31326	
Mean intensity	Equal variances assumed	,935	,362	1,503	8	,171	3,27600	2,18008	-1,75127	8,30327	
	Equal variances not assumed			1,503	7,544	,174	3,27600	2,18008	-1,80463	8,35663	

INTERVIEW 3: AMY SCHUMER

Case Summaries^a

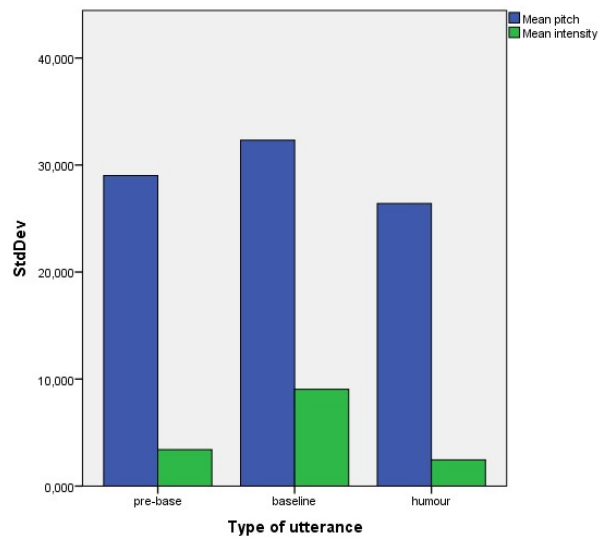
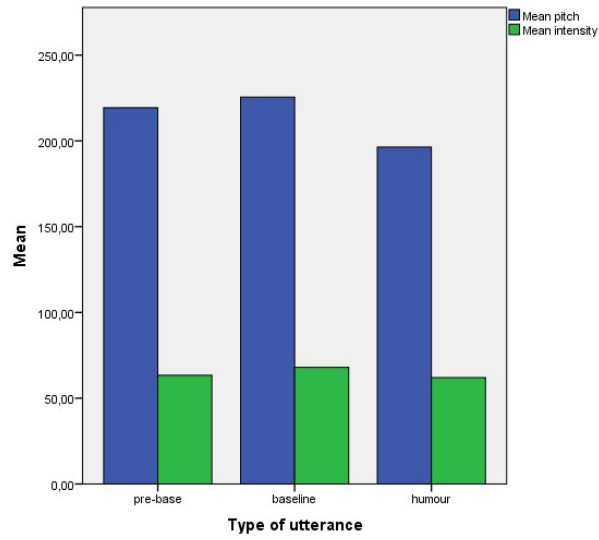
			Duration of utterance
Type of utterance	pre-base	1	1,220
		2	1,590
		3	1,860
		4	2,720
		5	3,090
		6	,620
		7	3,860
	Total	N	7
		Sum	14,960
	baseline	1	,930
		2	1,034
		3	,740
		4	,920
5		3,600	
6		1,330	
7		2,430	
Total	N	7	
	Sum	10,984	
humour	1	2,300	
	2	1,230	
	3	2,275	
	4	3,050	
	5	2,750	
	6	1,950	
	7	2,230	
	8	5,580	
	9	2,710	
	10	5,320	
	11	5,540	
	12	3,560	
Total	N	12	
	Sum	38,495	
Total	N	26	
	Sum	64,439	

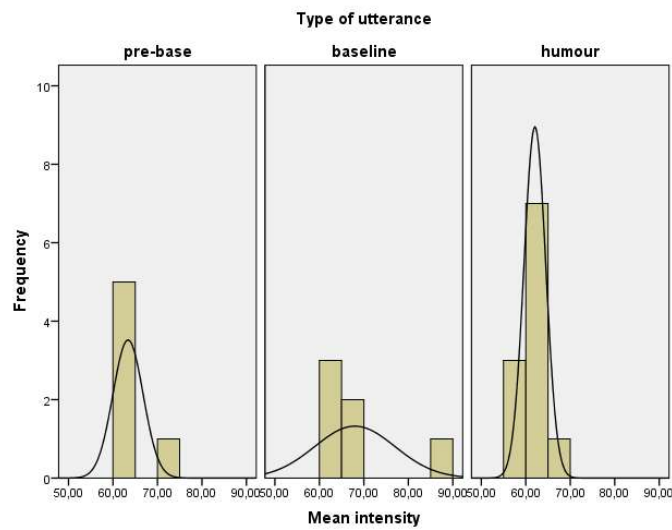
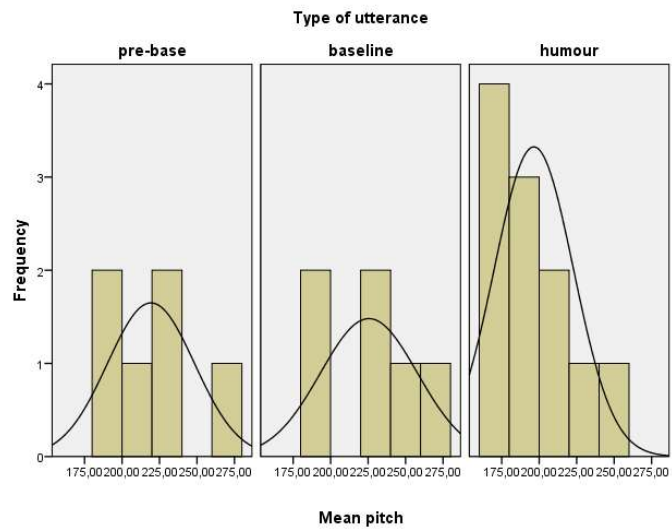
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	219,2950	63,3783
	N	6	6
	Std. Deviation	29,02287	3,39891
baseline	Mean	225,5383	67,9700
	N	6	6
	Std. Deviation	32,32674	9,04214
humour	Mean	196,4200	62,0445
	N	11	11
	Std. Deviation	26,40356	2,45036
Total	Mean	209,9835	63,9383
	N	23	23
	Std. Deviation	30,45399	5,49986





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		23	23
Normal Parameters ^{a,b}	Mean	209,9835	63,9383
	Std. Deviation	30,45399	5,49986
Most Extreme Differences	Absolute	,123	,253
	Positive	,123	,253
	Negative	-,077	-,170
Kolmogorov-Smirnov Z		,590	1,213
Asymp. Sig. (2-tailed)		,877	,106

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch pre-base	6	219,2950	29,02287	11,84854
baseline	6	225,5383	32,32674	13,19733
Mean intensity pre-base	6	63,3783	3,39891	1,38760
baseline	6	67,9700	9,04214	3,69144

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,114	,743	-,352	10	,732	-6,24333	17,73577	-45,76109	33,27442
	Equal variances not assumed			-,352	9,886	,732	-6,24333	17,73577	-45,82296	33,33629
Mean intensity	Equal variances assumed	1,924	,196	-1,164	10	,271	-4,59167	3,94362	-13,37860	4,19527
	Equal variances not assumed			-1,164	6,385	,286	-4,59167	3,94362	-14,10195	4,91861

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch baseline	6	225,5383	32,32674	13,19733
humour	11	196,4200	26,40356	7,96097
Mean intensity baseline	6	67,9700	9,04214	3,69144
humour	11	62,0445	2,45036	,73881

Independent Samples Test

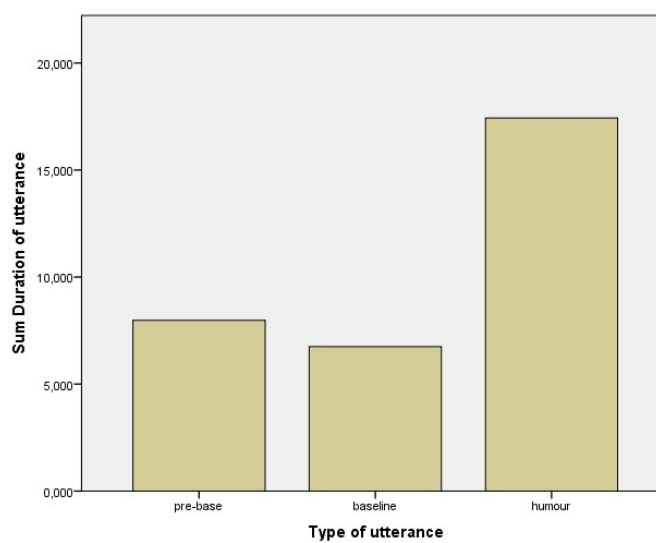
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,198	,663	2,012	15	,063	29,11833	14,47190	-1,72780	59,96447
	Equal variances not assumed			1,889	8,723	,092	29,11833	15,41255	-5,91657	64,15324
Mean intensity	Equal variances assumed	4,484	,051	2,088	15	,054	5,92545	2,83740	-,12233	11,97324
	Equal variances not assumed			1,574	5,404	,172	5,92545	3,76465	-3,53814	15,38905

INTERVIEW 4: CONDOLA RASHAD

Case Summaries^a

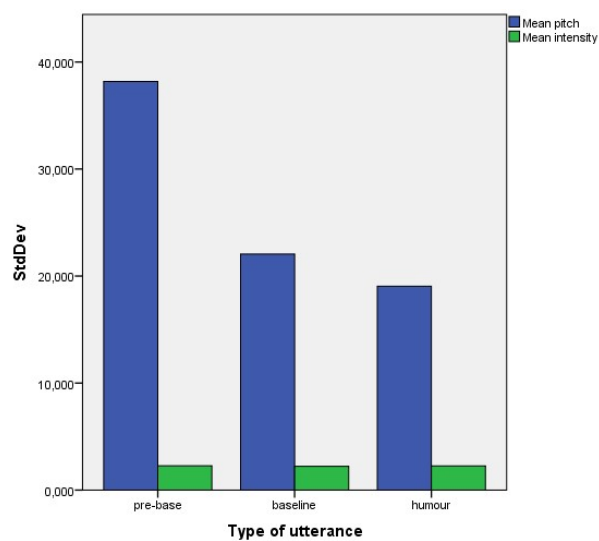
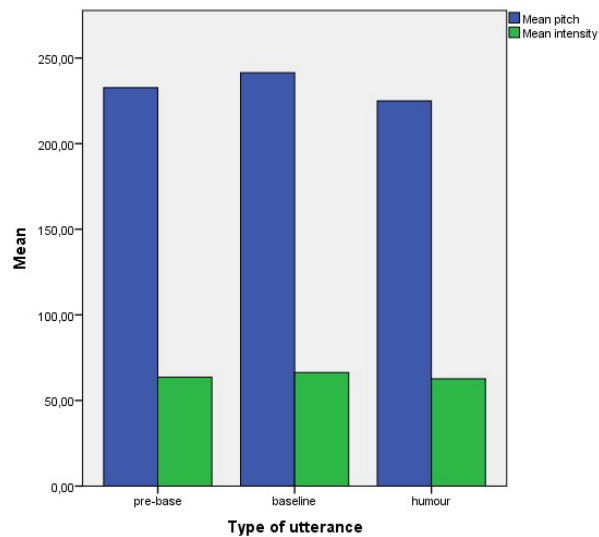
			Duration of utterance	
Type of utterance	pre-base	1	2,890	
		2	1,490	
		3	3,600	
		Total	N	3
			Sum	7,980
	baseline	1	1,440	
		2	3,110	
		3	2,200	
		Total	N	3
			Sum	6,750
humour	1	6,470		
	2	2,550		
	3	3,600		
	4	4,810		
	Total	N	4	
		Sum	17,430	
Total	N		10	
	Sum		32,160	

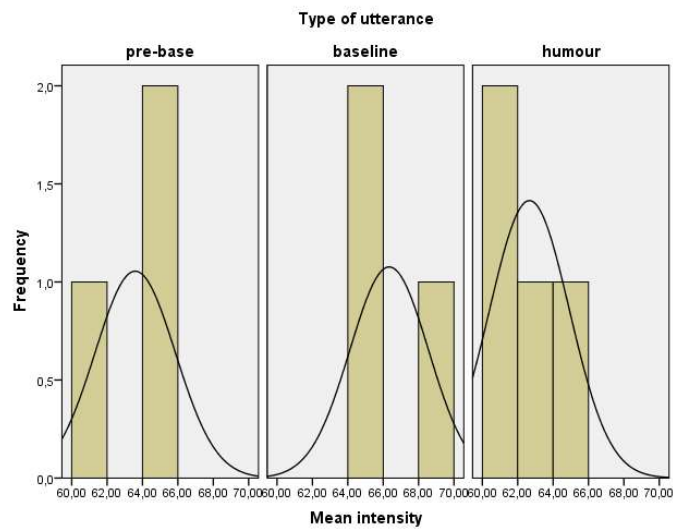
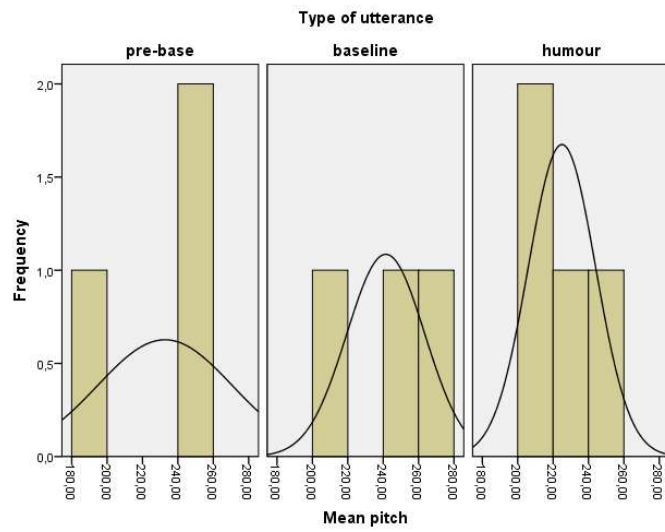
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	232,6767	63,5867
	N	3	3
	Std. Deviation	38,18587	2,26895
baseline	Mean	241,4433	66,3367
	N	3	3
	Std. Deviation	22,05118	2,22338
humour	Mean	224,9875	62,6625
	N	4	4
	Std. Deviation	19,05478	2,25596
Total	Mean	232,2310	64,0420
	N	10	10
	Std. Deviation	24,59258	2,57081





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		10	10
Normal Parameters ^{a,b}	Mean	232,2310	64,0420
	Std. Deviation	24,59258	2,57081
Most Extreme Differences	Absolute	,200	,165
	Positive	,122	,165
	Negative	-,200	-,121
Kolmogorov-Smirnov Z		,634	,520
Asymp. Sig. (2-tailed)		,816	,949

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	pre-base	3	232,6767	38,18587	22,04663
	baseline	3	241,4433	22,05118	12,73125
Mean intensity	pre-base	3	63,5867	2,26895	1,30998
	baseline	3	66,3367	2,22338	1,28367

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	1,953	,235	-,344	4	,748	-8,76667	25,45856	-79,45097	61,91764
	Equal variances not assumed			-,344	3,200	,752	-8,76667	25,45856	-86,99143	69,45810
Mean intensity	Equal variances assumed	,000	,991	-1,499	4	,208	-2,75000	1,83408	-7,84223	2,34223
	Equal variances not assumed			-1,499	3,998	,208	-2,75000	1,83408	-7,84305	2,34305

Group Statistics

	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	baseline	3	241,4433	22,05118	12,73125
	humour	4	224,9875	19,05478	9,52739
Mean intensity	baseline	3	66,3367	2,22338	1,28367
	humour	4	62,6625	2,25596	1,12798

Independent Samples Test

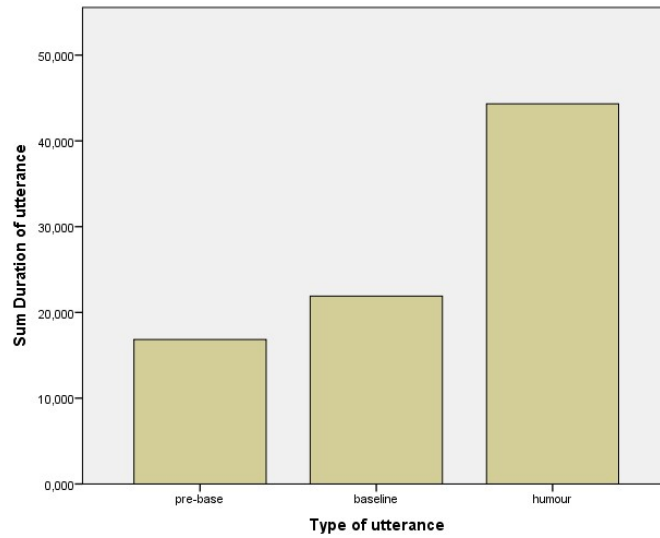
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,001	,975	1,061	5	,337	16,45583	15,50932	-23,41215	56,32381
	Equal variances not assumed			1,035	4,026	,359	16,45583	15,90145	-27,58296	60,49463
Mean intensity	Equal variances assumed	,003	,961	2,145	5	,085	3,67417	1,71311	-,72952	8,07785
	Equal variances not assumed			2,150	4,495	,090	3,67417	1,70884	-,87139	8,21972

INTERVIEW 5: CRISTELA ALONZO

Case Summaries^a

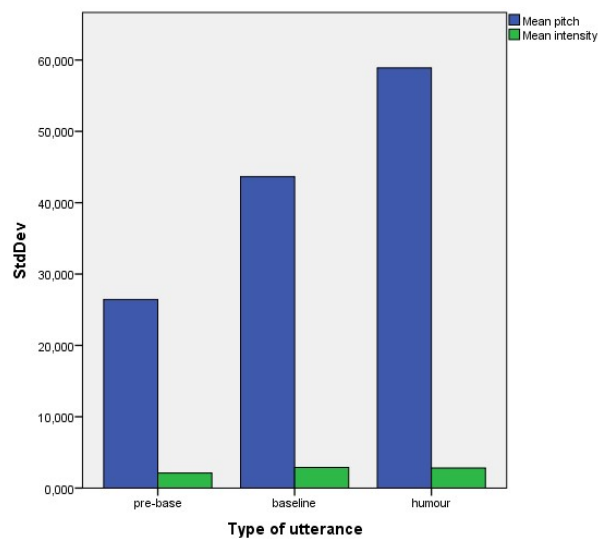
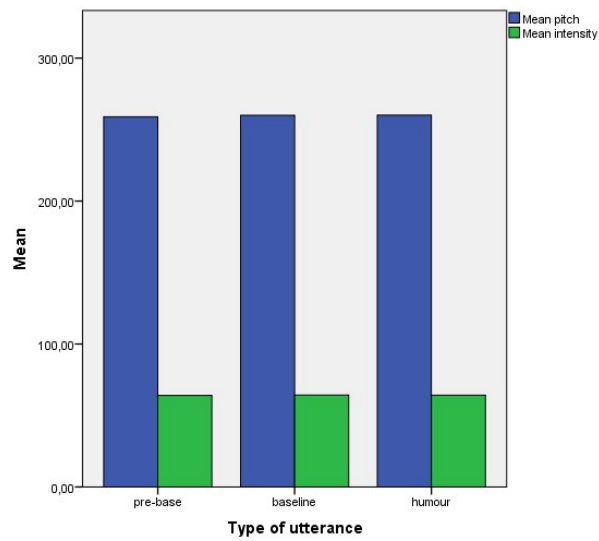
			Duration of utterance
Type of utterance	pre-base	1	2,540
		2	1,280
		3	1,750
		4	1,860
		5	1,750
		6	2,820
		7	2,180
		8	2,660
	Total	N	8
		Sum	16,840
	baseline	1	3,570
		2	1,970
		3	3,420
4		1,410	
5		1,220	
6		3,720	
7		1,810	
8		,800	
9		2,710	
10		1,270	
Total	N	10	
	Sum	21,900	
humour	1	5,060	
	2	6,370	
	3	4,950	
	4	3,895	
	5	2,590	
	6	4,220	
	7	2,770	
	8	1,580	
	9	3,300	
	10	3,560	
	11	1,670	
	12	4,360	
Total	N	12	
	Sum	44,325	
Total	N	30	
	Sum	83,065	

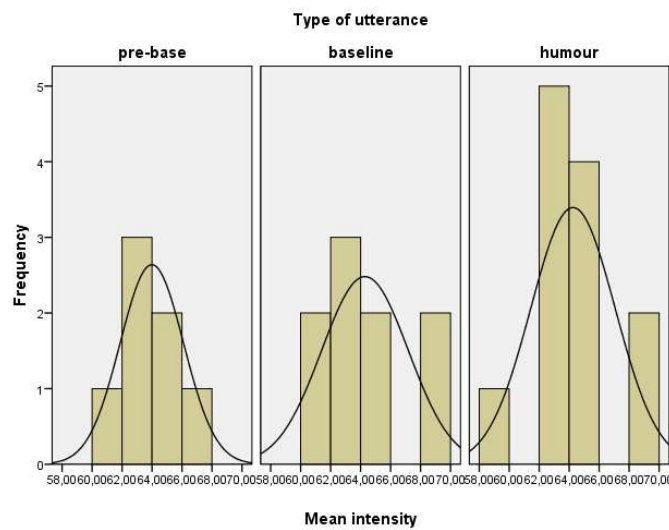
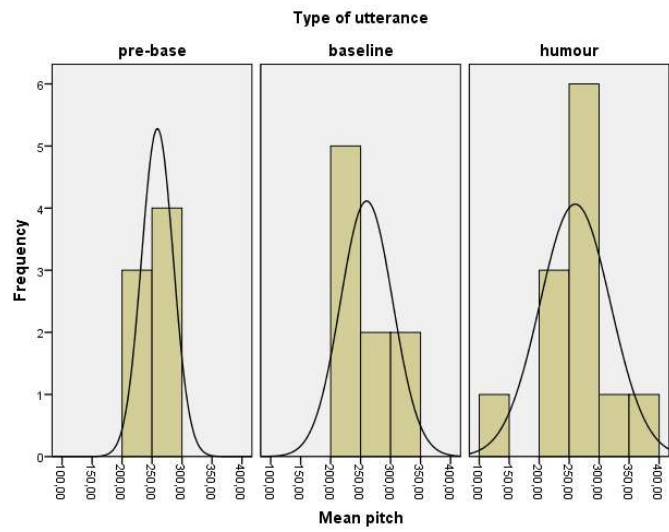
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	258,9414	63,9857
	N	7	7
	Std. Deviation	26,44350	2,11659
baseline	Mean	260,0211	64,2911
	N	9	9
	Std. Deviation	43,63078	2,89421
humour	Mean	260,0708	64,2517
	N	12	12
	Std. Deviation	58,88834	2,82058
Total	Mean	259,7725	64,1979
	N	28	28
	Std. Deviation	46,17895	2,59509





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		28	28
Normal Parameters ^{a,b}	Mean	259,7725	64,1979
	Std. Deviation	46,17895	2,59509
Most Extreme Differences	Absolute	,105	,124
	Positive	,098	,124
	Negative	-,105	-,100
Kolmogorov-Smirnov Z		,558	,657
Asymp. Sig. (2-tailed)		,915	,782

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	pre-base	7	258,9414	26,44350	9,99470
	baseline	9	260,0211	43,63078	14,54359
Mean intensity	pre-base	7	63,9857	2,11659	,80000
	baseline	9	64,2911	2,89421	,96474

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
Mean pitch	Equal variances assumed	5,410	,036	-,058	14	,955	-1,07968	18,77167	-41,34091	39,18155
	Equal variances not assumed			-,061	13,366	,952	-1,07968	17,64682	-39,09750	36,93814
Mean intensity	Equal variances assumed	,875	,365	-,234	14	,818	-,30540	1,30508	-3,10453	2,49373
	Equal variances not assumed			-,244	13,975	,811	-,30540	1,25328	-2,99388	2,38308

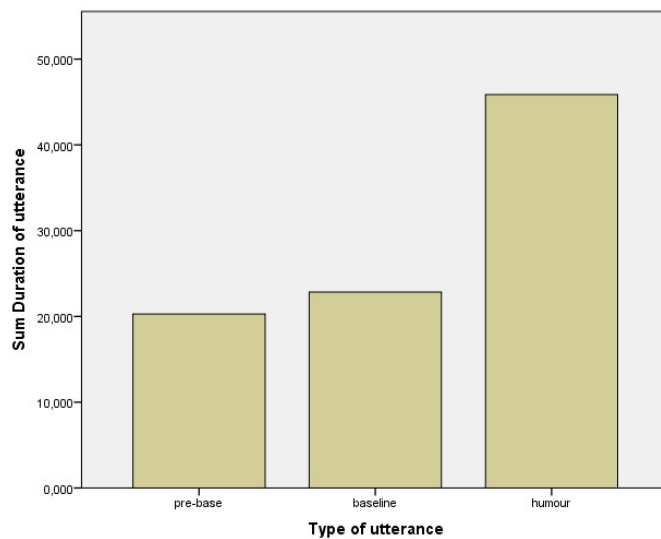
Group Statistics

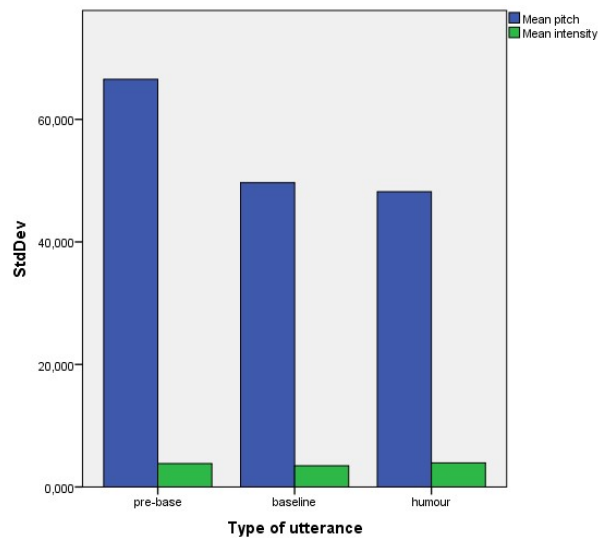
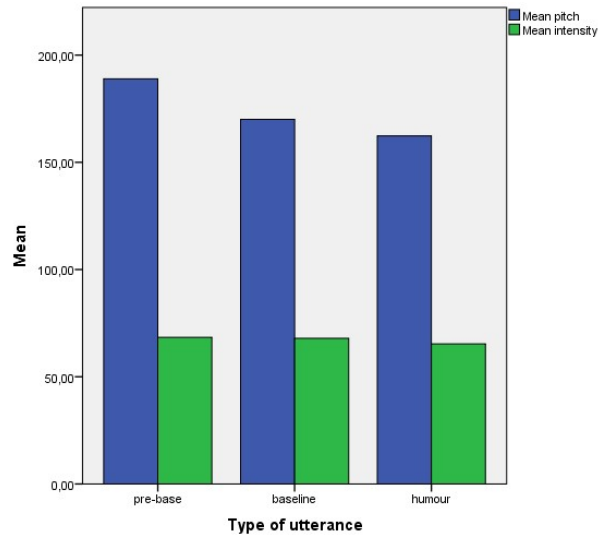
	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	baseline	9	260,0211	43,63078	14,54359
	humour	12	260,0708	58,88834	16,99960
Mean intensity	baseline	9	64,2911	2,89421	,96474
	humour	12	64,2517	2,82058	,81423

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
Mean pitch	Equal variances assumed	,041	,841	-,002	19	,998	-,04972	23,37176	-48,96737	48,86793
	Equal variances not assumed			-,002	19,000	,998	-,04972	22,37191	-46,87470	46,77526
Mean intensity	Equal variances assumed	,000	,983	,031	19	,975	,03944	1,25753	-2,59260	2,67149
	Equal variances not assumed			,031	17,134	,975	,03944	1,26241	-2,62243	2,70132

INTERVIEW 6: DANIEL KALUUYA





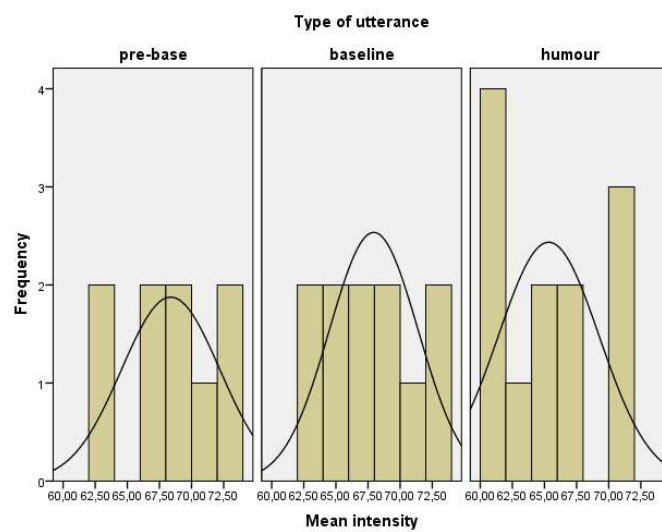
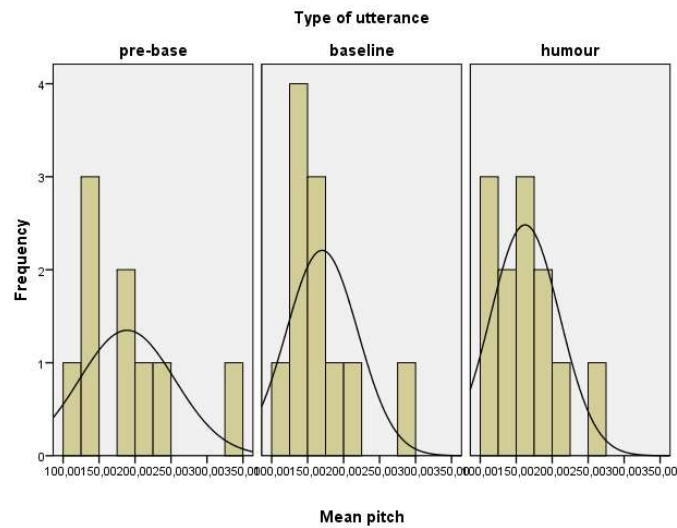
Case Summaries^a

			Duration of utterance
Type of utterance	pre-base	1	1,195
		2	4,640
		3	1,302
		4	3,020
		5	,819
		6	3,780
		7	2,760
		8	1,900
		9	,860
		Total	N
		Sum	20,276
baseline	1	2,600	
	2	1,620	
	3	,700	
	4	,820	
	5	2,140	
	6	1,320	
	7	1,750	
	8	2,280	
	9	1,190	
	10	2,045	
	11	2,240	
	12	2,980	
	13	1,160	
		Total	N
		Sum	22,845
humour	1	2,413	
	2	3,240	
	3	3,480	
	4	4,943	
	5	3,155	
	6	2,740	
	7	3,310	
	8	3,120	
	9	3,240	
	10	3,608	
	11	5,500	
	12	3,100	
	13	1,190	
	14	2,820	
	Total	N	14
		Sum	45,859
Total	N		36
	Sum		88,980

a. Limited to first 100 cases.

Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	188,9444	68,3733
	N	9	9
	Std. Deviation	66,54127	3,82891
baseline	Mean	170,0491	67,9445
	N	11	11
	Std. Deviation	49,67407	3,46133
humour	Mean	162,3283	65,3275
	N	12	12
	Std. Deviation	48,20253	3,93147
Total	Mean	172,4681	67,0838
	N	32	32
	Std. Deviation	53,70071	3,88232



One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		32	32
Normal Parameters ^{a,b}	Mean	172,4681	67,0838
	Std. Deviation	53,70071	3,88232
Most Extreme Differences	Absolute	,139	,101
	Positive	,139	,101
	Negative	-,106	-,078
Kolmogorov-Smirnov Z		,789	,572
Asymp. Sig. (2-tailed)		,562	,899

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	pre-base	9	188,9444	66,54127	22,18042
	baseline	11	170,0491	49,67407	14,97730
Mean intensity	pre-base	9	68,3733	3,82891	1,27630
	baseline	11	67,9445	3,46133	1,04363

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,886	,359	,728	18	,476	18,89535	25,97094	-35,66758	73,45828	
	Equal variances not assumed			,706	14,540	,491	18,89535	26,76360	-38,30743	76,09813	
Mean intensity	Equal variances assumed	,007	,936	,263	18	,796	,42879	1,63125	-2,99834	3,85592	
	Equal variances not assumed			,260	16,407	,798	,42879	1,64867	-3,05921	3,91679	

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	baseline	11	170,0491	49,67407	14,97730
	humour	12	162,3283	48,20253	13,91487
Mean intensity	baseline	11	67,9445	3,46133	1,04363
	humour	12	65,3275	3,93147	1,13492

Independent Samples Test

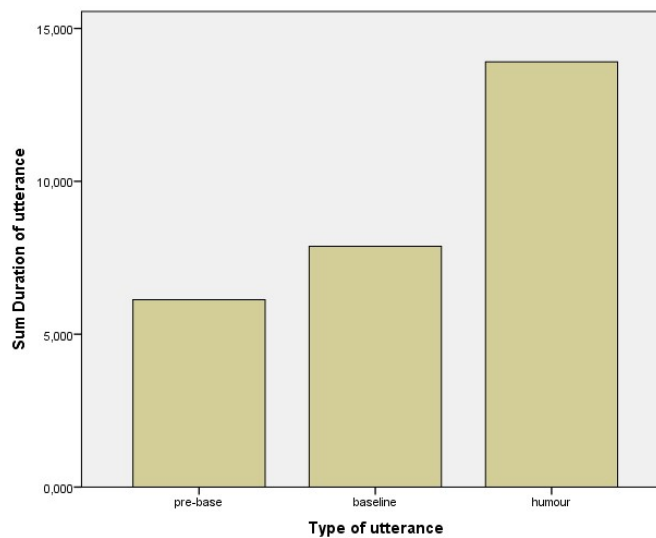
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,020	,890	,378	21	,709	7,72076	20,41568	-34,73597	50,17749	
	Equal variances not assumed			,378	20,696	,710	7,72076	20,44366	-34,83221	50,27373	
Mean intensity	Equal variances assumed	,306	,586	1,688	21	,106	2,61705	1,55074	-,60789	5,84198	
	Equal variances not assumed			1,697	20,973	,104	2,61705	1,54182	-,58959	5,82368	

INTERVIEW 7: ELON MUSK

Case Summaries^a

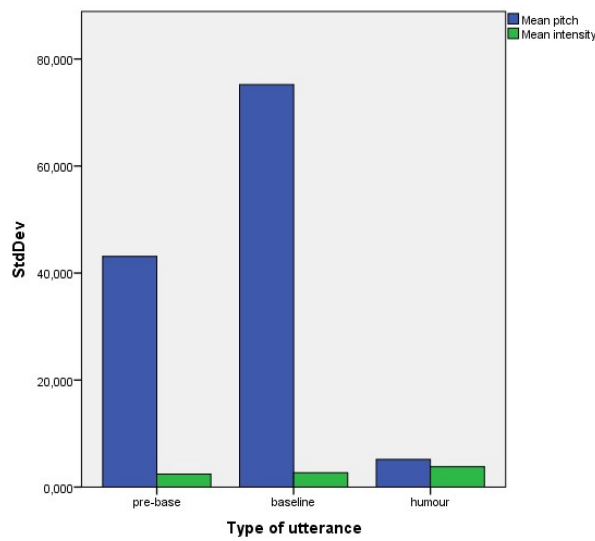
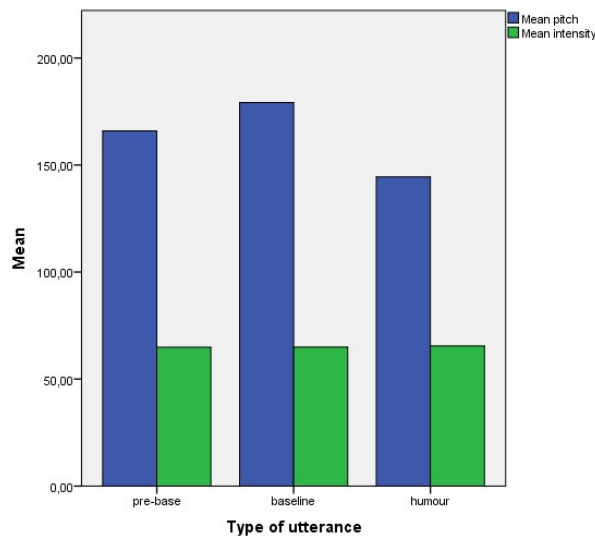
				Duration of utterance
Type of utterance	pre-base	1		3,160
		2		1,890
		3		1,080
		Total	N	3
			Sum	6,130
	baseline	1		1,760
		2		2,205
		3		1,310
		4		2,600
		Total	N	4
			Sum	7,875
humour	1		2,920	
	2		4,795	
	3		2,100	
	4		4,090	
	Total	N	4	
		Sum	13,905	
Total	N		11	
	Sum		27,910	

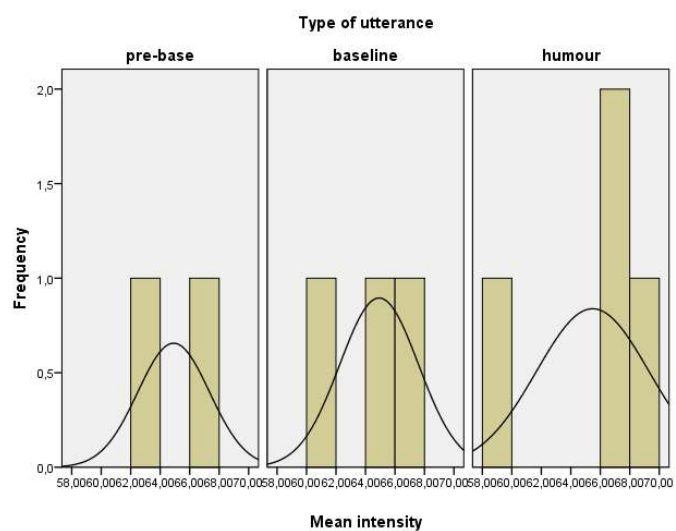
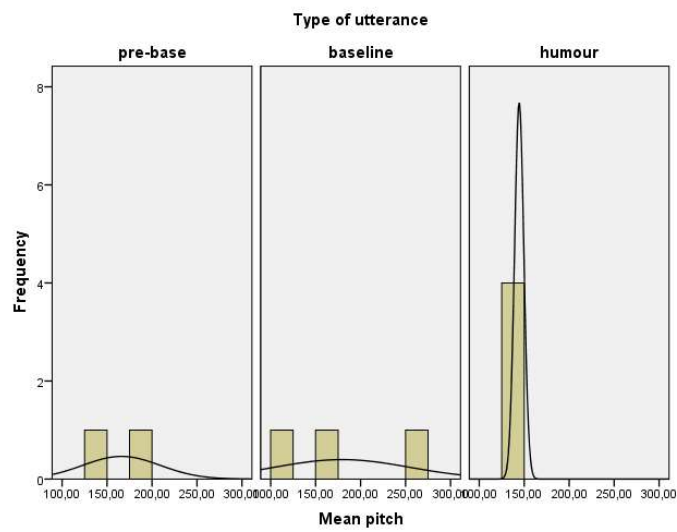
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	165,9900	64,9100
	N	2	2
	Std. Deviation	43,10523	2,43245
baseline	Mean	179,1700	64,9233
	N	3	3
	Std. Deviation	75,22266	2,67395
humour	Mean	144,4275	65,4825
	N	4	4
	Std. Deviation	5,19048	3,80694
Total	Mean	160,8000	65,1689
	N	9	9
	Std. Deviation	43,86664	2,83733





One-Sample Kolmogorov-Smirnov Test

		Duration of utterance	Mean pitch	Mean intensity
N		11	9	9
Normal Parameters ^{a,b}	Mean	2,53727	160,8000	65,1689
	Std. Deviation	1,140222	43,86664	2,83733
Most Extreme Differences	Absolute	,160	,281	,230
	Positive	,160	,281	,128
	Negative	-,101	-,171	-,230
Kolmogorov-Smirnov Z		,531	,843	,689
Asymp. Sig. (2-tailed)		,941	,477	,729

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	pre-base	2	165,9900	43,10523	30,48000
	baseline	3	179,1700	75,22266	43,42982
Mean intensity	pre-base	2	64,9100	2,43245	1,72000
	baseline	3	64,9233	2,67395	1,54381

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,829	,430	-,218	3	,842	-13,18000	60,49554	-205,70380	179,34380
	Equal variances not assumed			-,248	3,000	,820	-13,18000	53,05827	-182,03984	155,67984
Mean intensity	Equal variances assumed	,084	,791	-,006	3	,996	-,01333	2,36977	-7,55499	7,52832
	Equal variances not assumed			-,006	2,461	,996	-,01333	2,31122	-8,36839	8,34172

Group Statistics

	Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch	baseline	3	179,1700	75,22266	43,42982
	humour	4	144,4275	5,19048	2,59524
Mean intensity	baseline	3	64,9233	2,67395	1,54381
	humour	4	65,4825	3,80694	1,90347

Independent Samples Test

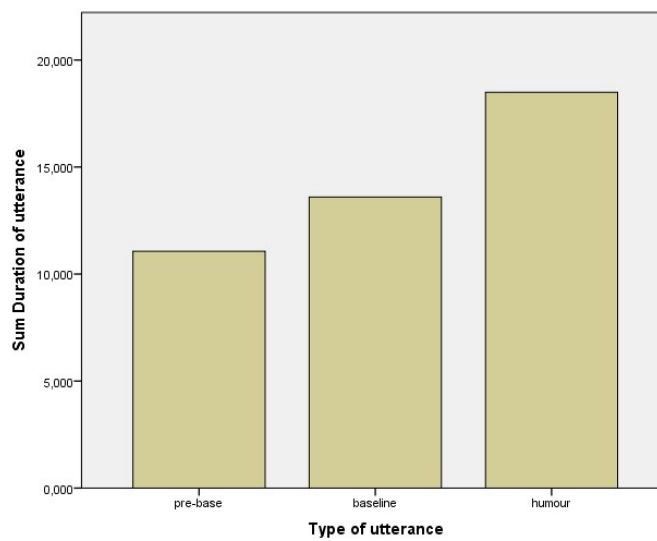
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	8,776	,031	,953	5	,384	34,74250	36,46552	-58,99509	128,48009
	Equal variances not assumed			,799	2,014	,508	34,74250	43,50729	-151,18720	220,67220
Mean intensity	Equal variances assumed	,331	,590	-,215	5	,838	-,55917	2,59631	-7,23320	6,11486
	Equal variances not assumed			-,228	5,000	,829	-,55917	2,45083	-6,85930	5,74096

INTERVIEW 8: VICE PRESIDENT JOSEPH BIDEN

Case Summaries^a

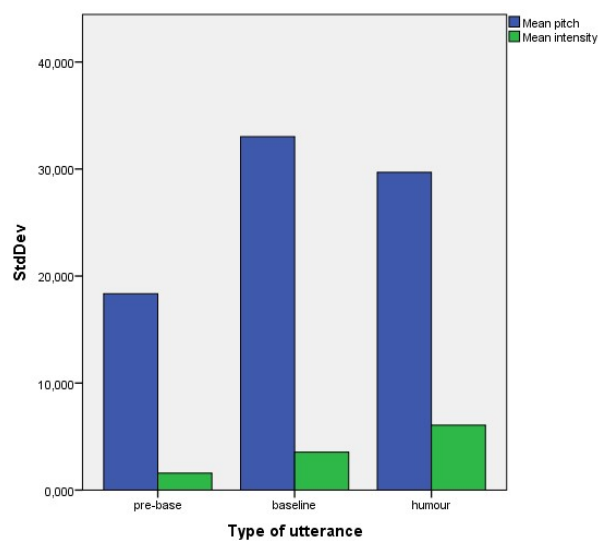
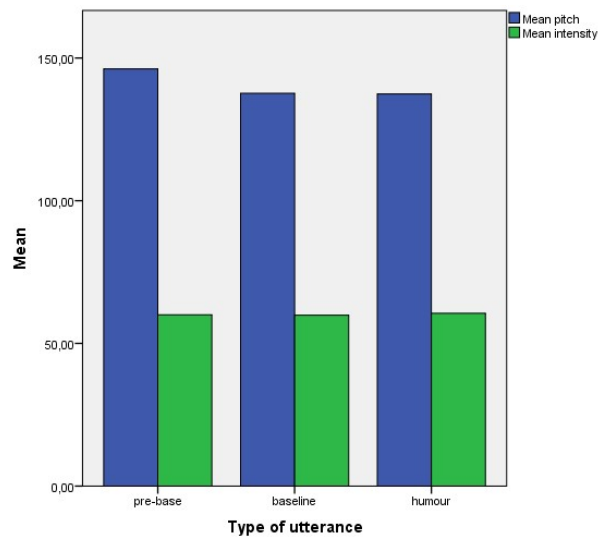
			Duration of utterance	
Type of utterance	pre-base	1	1,780	
		2	2,535	
		3	1,680	
		4	2,330	
		5	2,740	
		Total	N	5
			Sum	11,065
	baseline	1	4,190	
		2	3,915	
		3	1,010	
		4	3,090	
		5	1,390	
Total		N	5	
		Sum	13,595	
humour	1	3,740		
	2	2,700		
	3	2,005		
	4	2,000		
	5	2,950		
	6	2,400		
	7	2,700		
	Total	N	7	
		Sum	18,495	
Total	N		17	
	Sum		43,155	

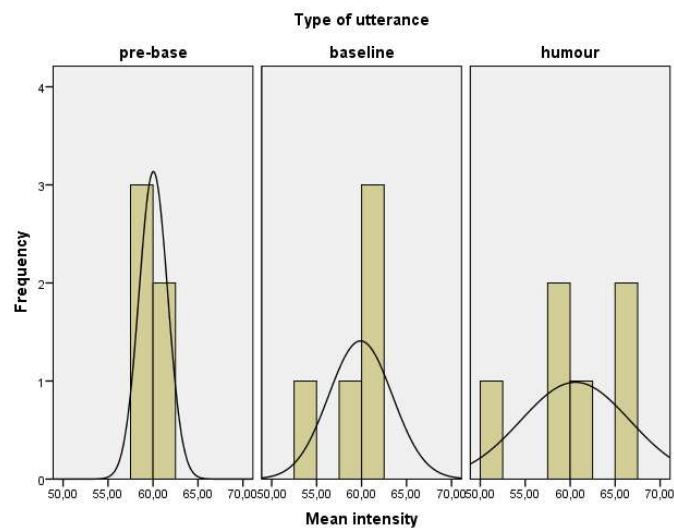
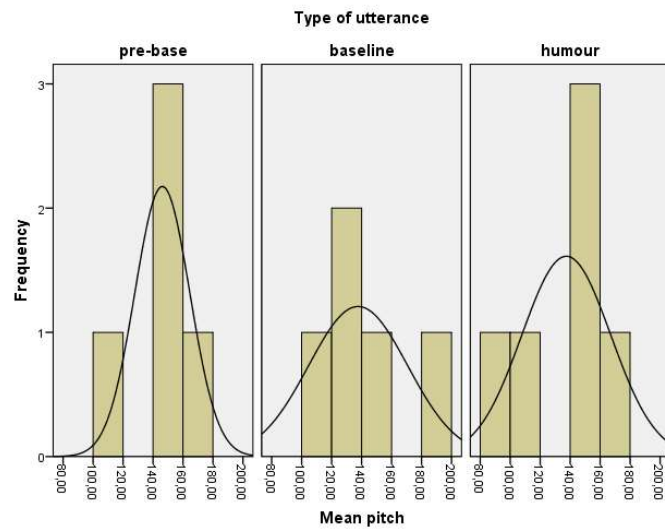
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	146,2080	60,0360
	N	5	5
	Std. Deviation	18,34652	1,59026
baseline	Mean	137,6640	59,8820
	N	5	5
	Std. Deviation	33,02078	3,53944
humour	Mean	137,4000	60,5717
	N	6	6
	Std. Deviation	29,68957	6,06161
Total	Mean	140,2350	60,1888
	N	16	16
	Std. Deviation	26,29940	4,04482





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		16	16
Normal Parameters ^{a,b}	Mean	140,2350	60,1888
	Std. Deviation	26,29940	4,04482
Most Extreme Differences	Absolute	,186	,202
	Positive	,116	,165
	Negative	-,186	-,202
Kolmogorov-Smirnov Z		,744	,809
Asymp. Sig. (2-tailed)		,636	,529

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch pre-base	5	146,2080	18,34652	8,20481
baseline	5	137,6640	33,02078	14,76734
Mean intensity pre-base	5	60,0360	1,59026	,71119
baseline	5	59,8820	3,53944	1,58289

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	1,426	,267	,506	8	,627	8,54400	16,89359	-30,41268	47,50068	
	Equal variances not assumed			,506	6,255	,630	8,54400	16,89359	-32,38847	49,47647	
Mean intensity	Equal variances assumed	1,015	,343	,089	8	,931	,15400	1,73532	-3,84764	4,15564	
	Equal variances not assumed			,089	5,552	,932	,15400	1,73532	-4,17668	4,48468	

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch baseline	5	137,6640	33,02078	14,76734
humour	6	137,4000	29,68957	12,12072
Mean intensity baseline	5	59,8820	3,53944	1,58289
humour	6	60,5717	6,06161	2,47464

Independent Samples Test

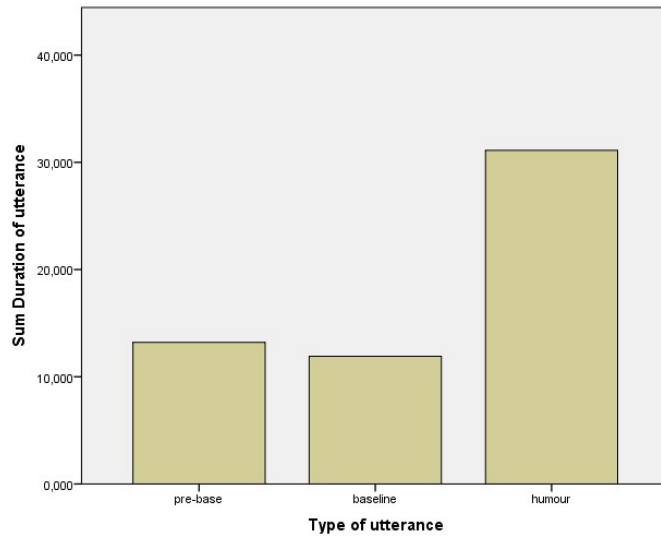
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	,000	,988	,014	9	,989	,26400	18,90103	-42,49311	43,02111	
	Equal variances not assumed			,014	8,220	,989	,26400	19,10461	-43,58670	44,11470	
Mean intensity	Equal variances assumed	,965	,352	-,223	9	,828	-,68967	3,08646	-7,67173	6,29240	
	Equal variances not assumed			-,235	8,210	,820	-,68967	2,93758	-7,43364	6,05431	

INTERVIEW 9: JOHN McWHORTER

Case Summaries^a

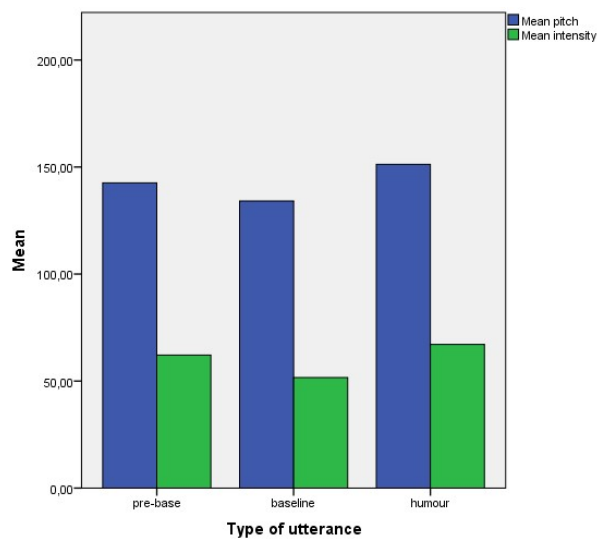
			Duration of utterance	
Type of utterance	pre-base	1	3,450	
		2	3,420	
		3	1,190	
		4	2,890	
		5	2,260	
		Total	N	5
			Sum	13,210
	baseline	1	1,340	
		2	1,750	
		3	1,530	
		4	5,160	
		5	2,130	
		Total	N	5
			Sum	11,910
	humour	1	1,600	
		2	4,440	
		3	2,760	
4		2,310		
5		4,250		
6		7,010		
7		8,750		
	Total	N	7	
		Sum	31,120	
Total	N		17	
	Sum		56,240	

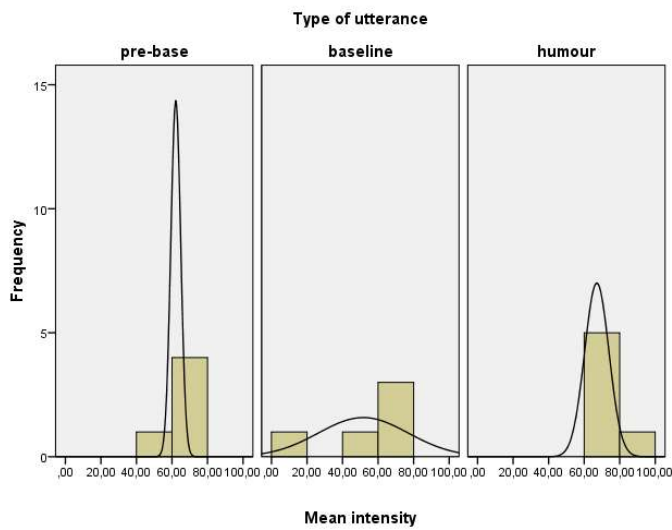
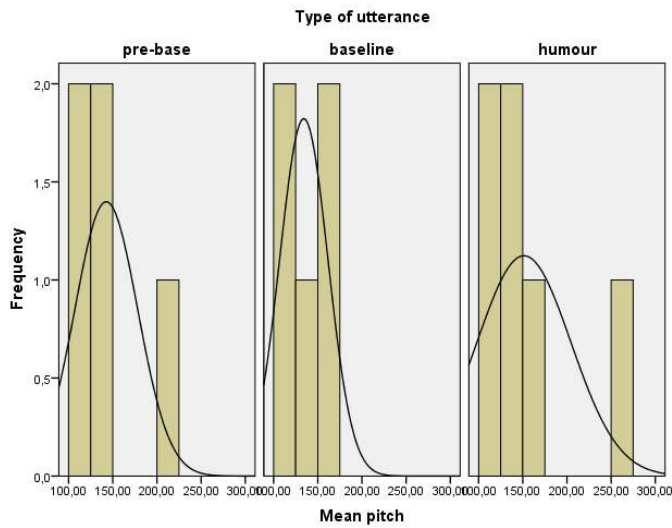
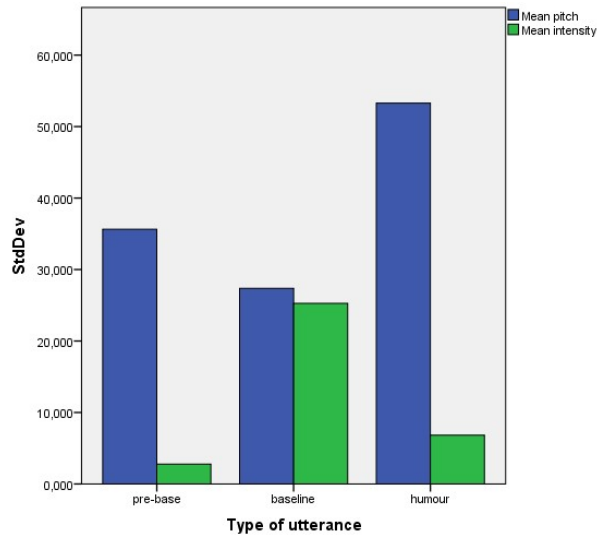
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	142,5920	62,1260
	N	5	5
	Std. Deviation	35,64852	2,77316
baseline	Mean	134,1440	51,6620
	N	5	5
	Std. Deviation	27,36588	25,26174
humour	Mean	151,2717	67,0967
	N	6	6
	Std. Deviation	53,27241	6,83308
Total	Mean	143,2069	60,7200
	N	16	16
	Std. Deviation	39,21853	15,23358





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		16	16
Normal Parameters ^{a,b}	Mean	143,2069	60,7200
	Std. Deviation	39,21853	15,23358
Most Extreme Differences	Absolute	,211	,359
	Positive	,211	,269
	Negative	-,137	-,359
Kolmogorov-Smirnov Z		,845	1,436
Asymp. Sig. (2-tailed)		,474	,032

- a. Test distribution is Normal.
- b. Calculated from data.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
2	The distribution of Mean intensity is the same across categories of Type of utterance.	Independent-Samples Kruskal-Wallis Test	,183	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch pre-base	5	142,5920	35,64852	15,94250
baseline	5	134,1440	27,36588	12,23839

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,005	,947	,420	8	,685	8,44800	20,09830	-37,89876	54,79476
	Equal variances not assumed			,420	7,499	,686	8,44800	20,09830	-38,44305	55,33905

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch baseline	5	134,1440	27,36588	12,23839
humour	6	151,2717	53,27241	21,74837

Independent Samples Test

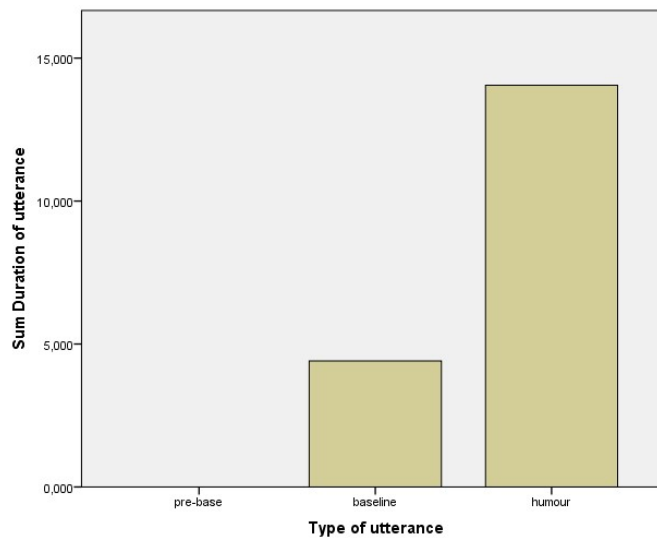
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,364	,561	-,647	9	,534	-17,12767	26,46020	-76,98481	42,72947
	Equal variances not assumed			-,686	7,703	,513	-17,12767	24,95536	-75,06398	40,80865

INTERVIEW 10: GENERAL MICHAEL HAYDEN

Case Summaries^a

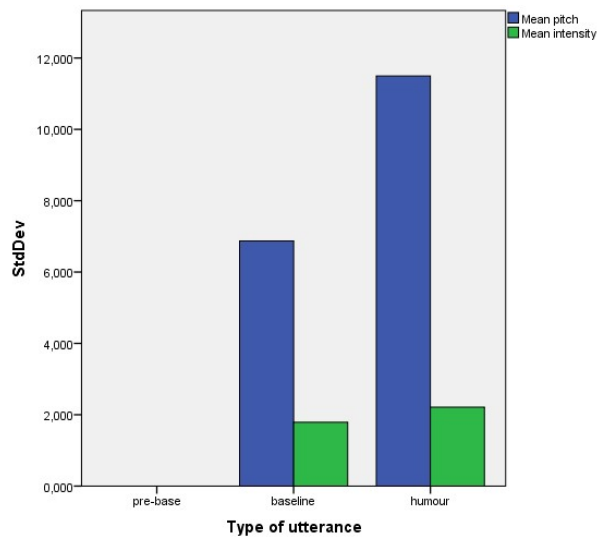
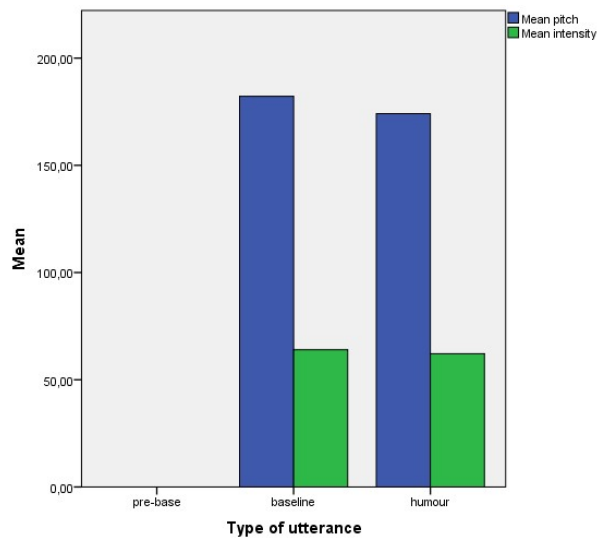
			Duration of utterance
Type of utterance	baseline	1	2,350
		2	,660
		3	1,400
	Total	N	3
	Sum		4,410
humour	1	5,050	
	2	2,960	
	3	1,950	
	4	2,050	
	5	2,040	
Total	N	5	
Sum		14,050	
Total	N	8	
Sum		18,460	

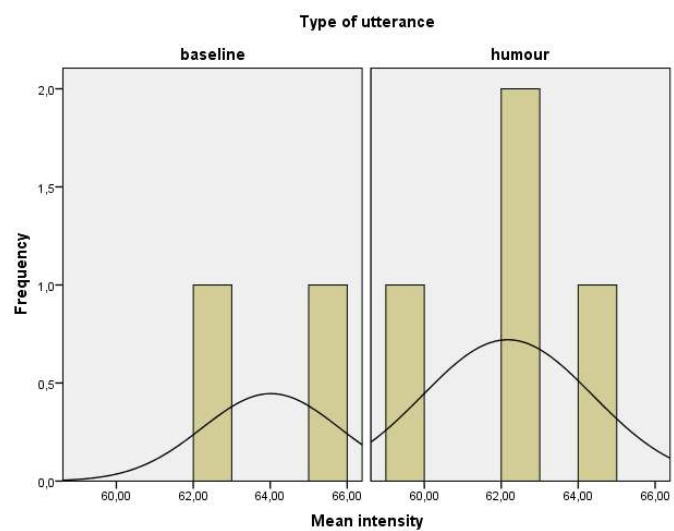
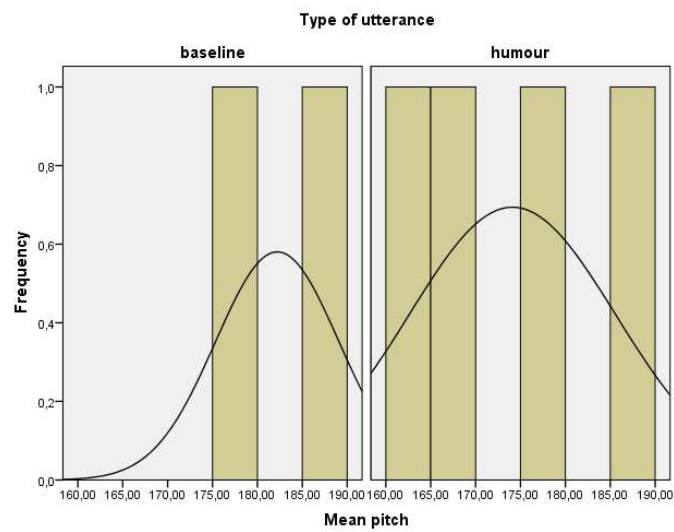
a. Limited to first 100 cases.



Report

Type of utterance	Utterance ID	Mean pitch	Mean intensity
baseline	N	3	2
	Mean	182,2100	64,0150
	Std. Deviation	6,87308	1,78898
humour	N	5	4
	Mean	174,0975	62,1650
	Std. Deviation	11,49548	2,21411
Total	N	8	6
	Mean	176,8017	62,7817
	Std. Deviation	10,30949	2,11994





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		6	6
Normal Parameters ^{a,b}	Mean	176,8017	62,7817
	Std. Deviation	10,30949	2,11994
Most Extreme Differences	Absolute	,188	,225
	Positive	,145	,173
	Negative	-,188	-,225
Kolmogorov-Smirnov Z		,460	,552
Asymp. Sig. (2-tailed)		,984	,921

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean
Mean pitch baseline	2	182,2100	6,87308	4,86000
Mean pitch humour	4	174,0975	11,49548	5,74774
Mean intensity baseline	2	64,0150	1,78898	1,26500
Mean intensity humour	4	62,1650	2,21411	1,10706

Independent Samples Test

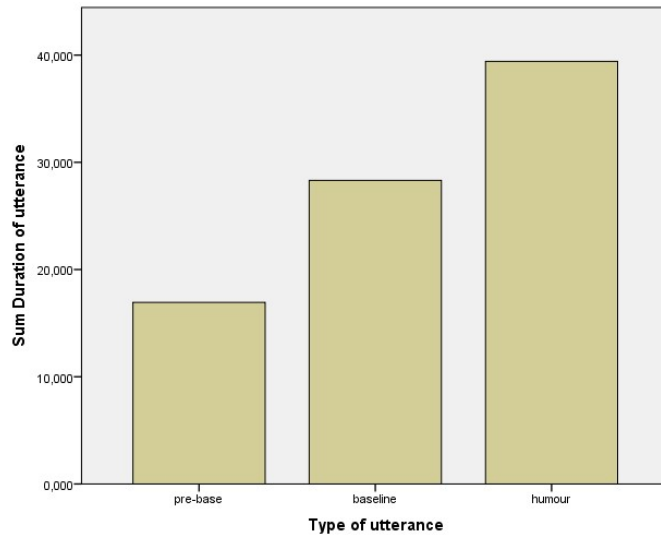
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	1,838	,247	,889	4	,424	8,11250	9,12083	-17,21098	33,43598
	Equal variances not assumed			1,078	3,483	,350	8,11250	7,52703	-14,06906	30,29406
Mean intensity	Equal variances assumed	,011	,920	1,010	4	,370	1,85000	1,83238	-3,23751	6,93751
	Equal variances not assumed			1,101	2,608	,362	1,85000	1,68101	-3,98418	7,68418

INTERVIEW 11: RIZ AHMED

Case Summaries^a

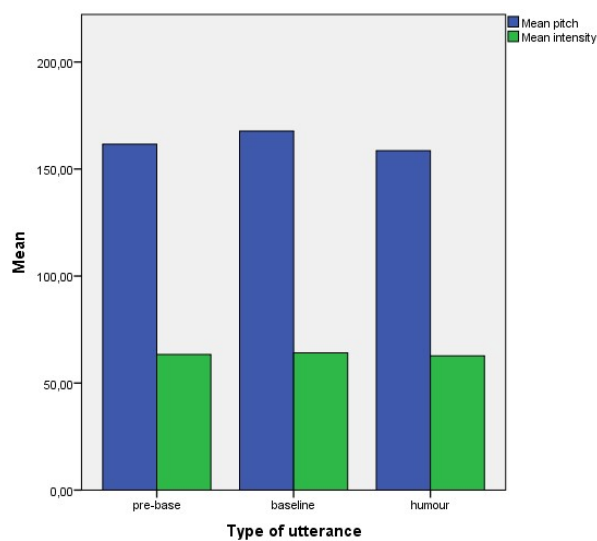
			Duration of utterance
Type of utterance	pre-base	1	2,920
		2	3,350
		3	,830
		4	1,180
		5	3,910
		6	2,320
		7	2,425
	Total	N	7
		Sum	16,935
	baseline	1	3,335
		2	,570
		3	3,170
		4	4,390
		5	1,410
6		3,390	
7		1,470	
8		2,580	
9		5,500	
10		2,500	
Total	N	10	
	Sum	28,315	
humour	1	2,270	
	2	2,930	
	3	1,190	
	4	4,710	
	5	3,990	
	6	2,700	
	7	2,470	
	8	4,430	
	9	4,650	
	10	1,850	
	11	3,330	
	12	1,690	
	13	3,210	
Total	N	13	
	Sum	39,420	
Total	N	30	
	Sum	84,670	

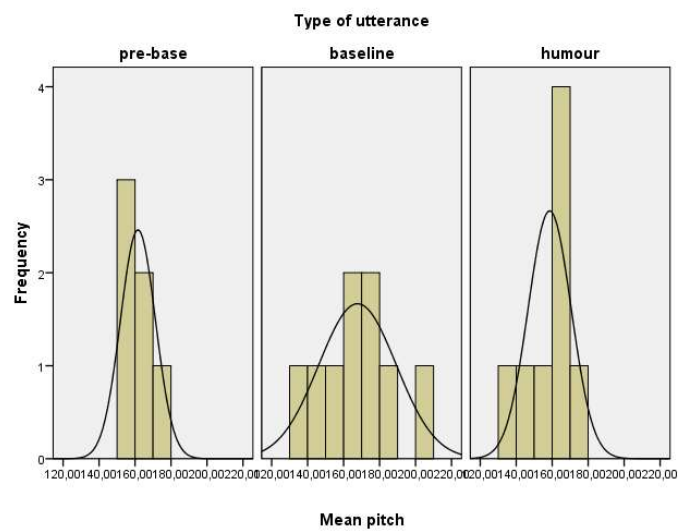
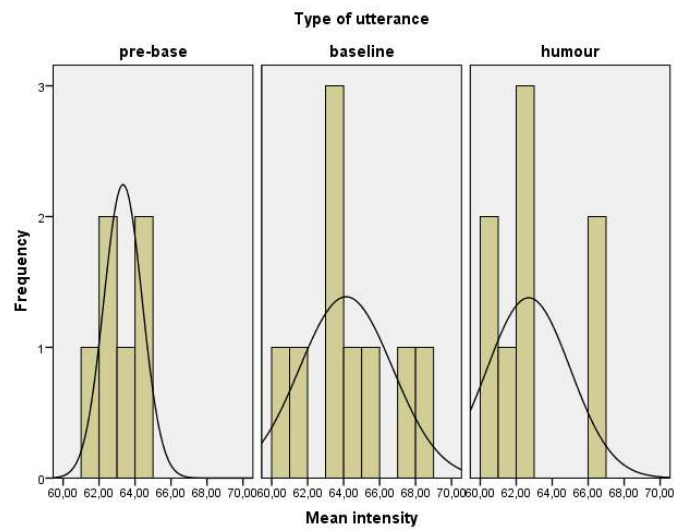
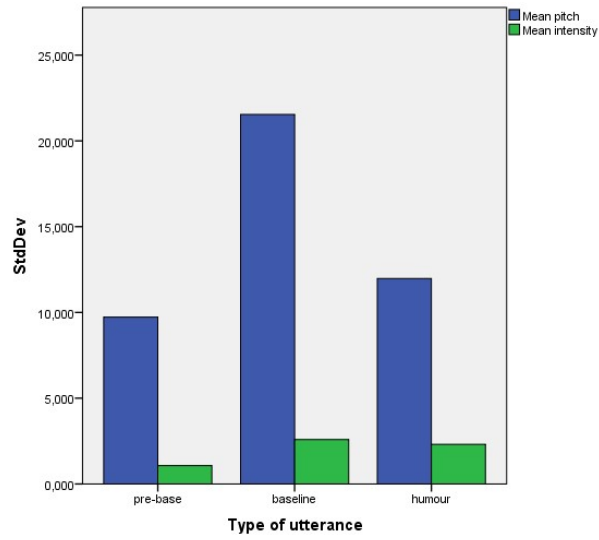
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	161,6117	63,3317
	N	6	6
	Std. Deviation	9,72739	1,06646
baseline	Mean	167,7200	64,1378
	N	9	9
	Std. Deviation	21,54136	2,58913
humour	Mean	158,5888	62,6938
	N	8	8
	Std. Deviation	11,97475	2,31413
Total	Mean	162,9504	63,4252
	N	23	23
	Std. Deviation	15,89283	2,19197





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		23	23
Normal Parameters ^{a,b}	Mean	162,9504	63,4252
	Std. Deviation	15,89283	2,19197
Most Extreme Differences	Absolute	,102	,126
	Positive	,102	,126
	Negative	-,068	-,060
Kolmogorov-Smirnov Z		,490	,604
Asymp. Sig. (2-tailed)		,970	,860

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	pre-base	6	161,6117	9,72739	3,97119
	baseline	9	167,7200	21,54136	7,18045
Mean intensity	pre-base	6	63,3317	1,06646	,43538
	baseline	9	64,1378	2,58913	,86304

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	3,427	,087	-,646	13	,530	-6,10833	9,45678	-26,53846	14,32179	
	Equal variances not assumed			-,744	11,866	,471	-6,10833	8,20544	-24,00885	11,79219	
Mean intensity	Equal variances assumed	2,640	,128	-,716	13	,487	-,80611	1,12580	-3,23825	1,62603	
	Equal variances not assumed			-,834	11,408	,421	-,80611	,96664	-2,92443	1,31221	

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	baseline	9	167,7200	21,54136	7,18045
	humour	8	158,5888	11,97475	4,23371
Mean intensity	baseline	9	64,1378	2,58913	,86304
	humour	8	62,6937	2,31413	,81817

Independent Samples Test

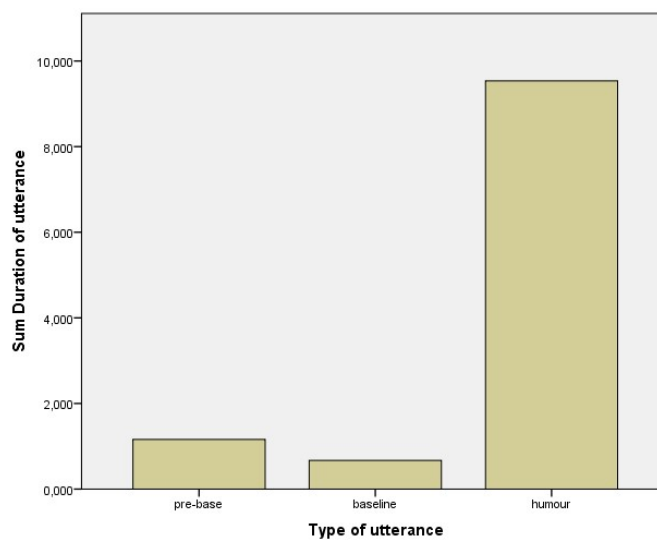
		Levene's Test for Equality of Variances		t-test for Equality of Means							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
										Lower	Upper
Mean pitch	Equal variances assumed	2,894	,110	1,060	15	,306	9,13125	8,61588	-9,23307	27,49557	
	Equal variances not assumed			1,095	12,766	,294	9,13125	8,33566	-8,91047	27,17297	
Mean intensity	Equal variances assumed	,079	,783	1,206	15	,247	1,44403	1,19759	-1,10857	3,99663	
	Equal variances not assumed			1,214	14,997	,243	1,44403	1,18922	-1,09077	3,97883	

INTERVIEW 12: SHERYL CROW

Case Summaries^a

			Duration of utterance
Type of utterance	pre-base	1	1,160
		Total	N
		Sum	1,160
baseline	baseline	1	,670
		Total	N
		Sum	,670
humour	humour	1	3,300
		2	1,900
		3	4,335
		Total	N
		Sum	9,535
		Total	N
Total		Sum	11,365

a. Limited to first 100 cases.

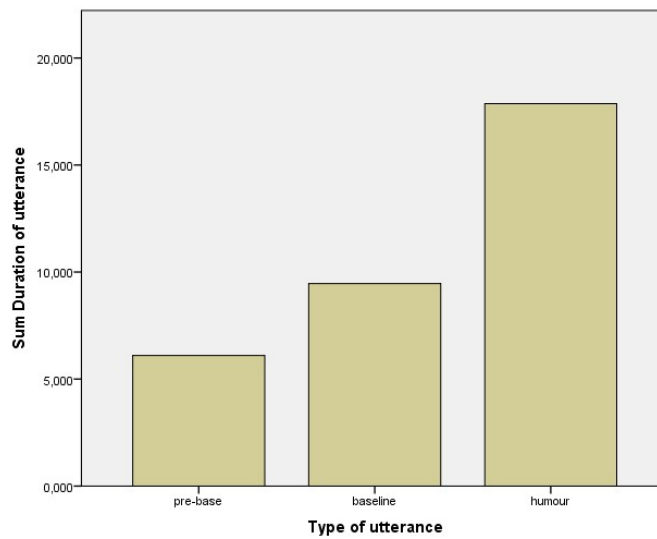


INTERVIEW 13: SIGOURNEY WEAVER

Case Summaries^a

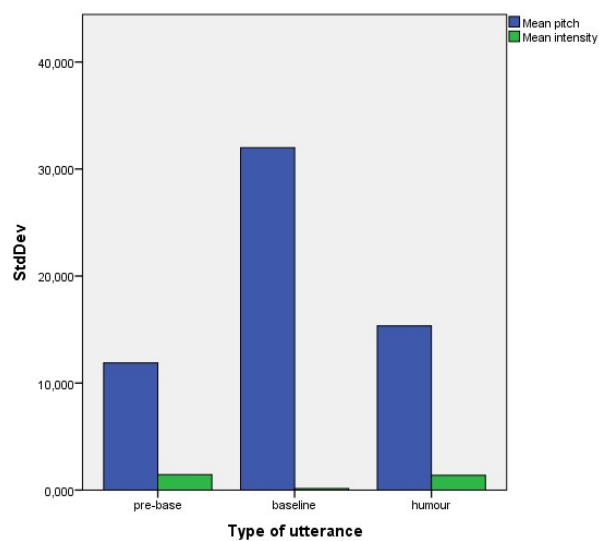
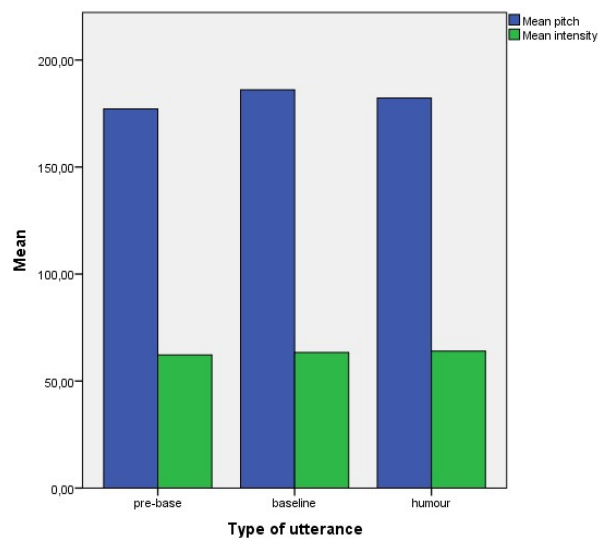
			Duration of utterance	
Type of utterance	pre-base	1	2,230	
		2	,800	
		3	3,070	
		Total	N	3
			Sum	6,100
	baseline	1	4,520	
		2	4,940	
		Total	N	2
			Sum	9,460
	humour	1	3,710	
2		1,330		
3		5,590		
4		3,890		
5		3,350		
Total		N	5	
		Sum	17,870	
Total	N	10		
	Sum		33,430	

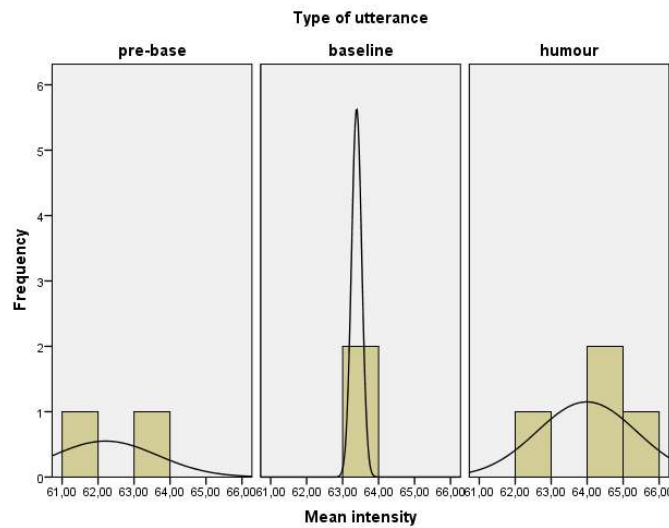
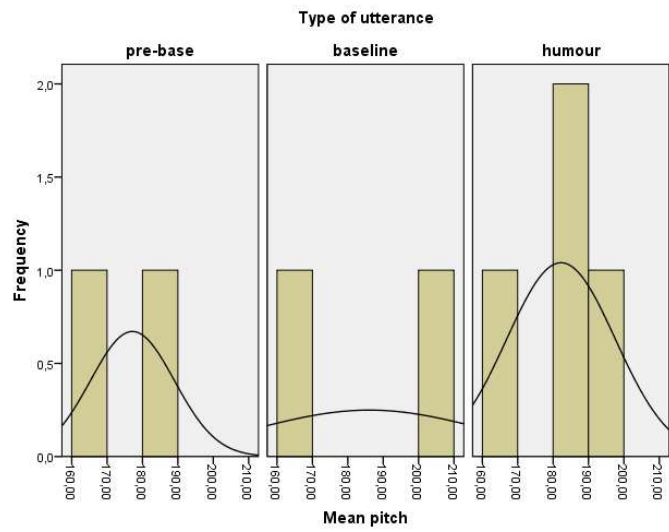
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	177,1450	62,2050
	N	2	2
	Std. Deviation	11,88646	1,44957
baseline	Mean	186,0450	63,3900
	N	2	2
	Std. Deviation	31,99658	,14142
humour	Mean	182,2275	63,9900
	N	4	4
	Std. Deviation	15,33979	1,38740
Total	Mean	181,9113	63,3938
	N	8	8
	Std. Deviation	16,69477	1,31715





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		8	8
Normal Parameters ^{a,b}	Mean	181,9113	63,3938
	Std. Deviation	16,69477	1,31715
Most Extreme Differences	Absolute	,160	,201
	Positive	,160	,096
	Negative	-,117	-,201
Kolmogorov-Smirnov Z		,452	,567
Asymp. Sig. (2-tailed)		,987	,905

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	baseline	2	186,0450	31,99658	22,62500
	humour	4	182,2275	15,33979	7,66990
Mean intensity	baseline	2	63,3900	,14142	,10000
	humour	4	63,9900	1,38740	,69370

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	3,055	,155	,212	4	,842	3,81750	18,00890	-46,18322	53,81822
	Equal variances not assumed			,160	1,238	,895	3,81750	23,88970	-191,41785	199,05285
Mean intensity	Equal variances assumed	1,780	,253	-,576	4	,596	-,60000	1,04235	-3,49402	2,29402
	Equal variances not assumed			-,856	3,122	,453	-,60000	,70087	-2,78199	1,58199

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Mean pitch is the same across categories of Type of utterance.	Independent-Samples Kruskal-Wallis Test	,920	Retain the null hypothesis.
2	The distribution of Mean intensity is the same across categories of Type of utterance.	Independent-Samples Kruskal-Wallis Test	,210	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	pre-base	2	177,1450	11,88646	8,40500
	baseline	2	186,0450	31,99658	22,62500
Mean intensity	pre-base	2	62,2050	1,44957	1,02500
	baseline	2	63,3900	,14142	,10000

Independent Samples Test

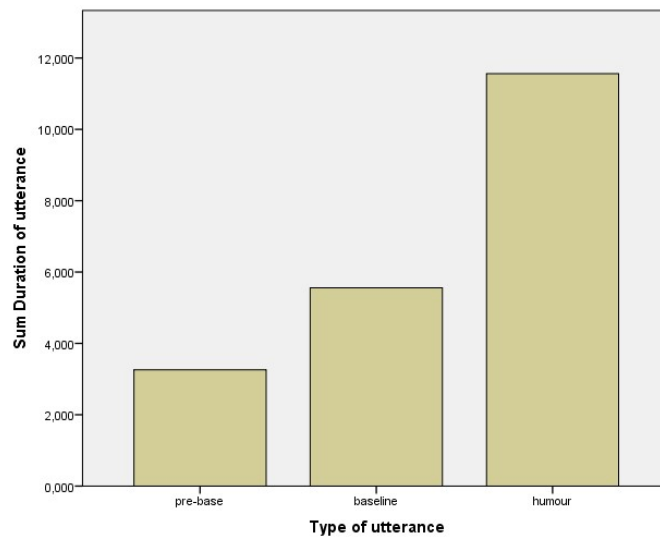
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed			-,369	2	,748	-8,90000	24,13575	-112,74777	94,94777
	Equal variances not assumed			-,369	1,271	,764	-8,90000	24,13575	-197,10545	179,30545
Mean intensity	Equal variances assumed	1,763E+18	,000	-1,151	2	,369	-1,18500	1,02987	-5,61616	3,24616
	Equal variances not assumed			-1,151	1,019	,453	-1,18500	1,02987	-13,70699	11,33699

INTERVIEW 14: SUSAN SARANDON

Case Summaries^a

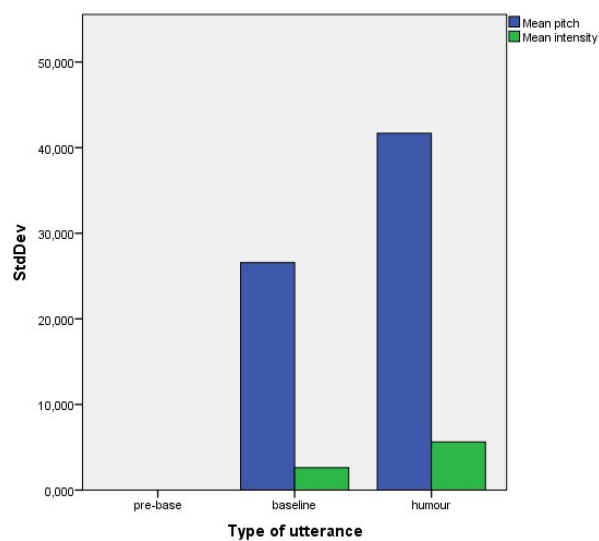
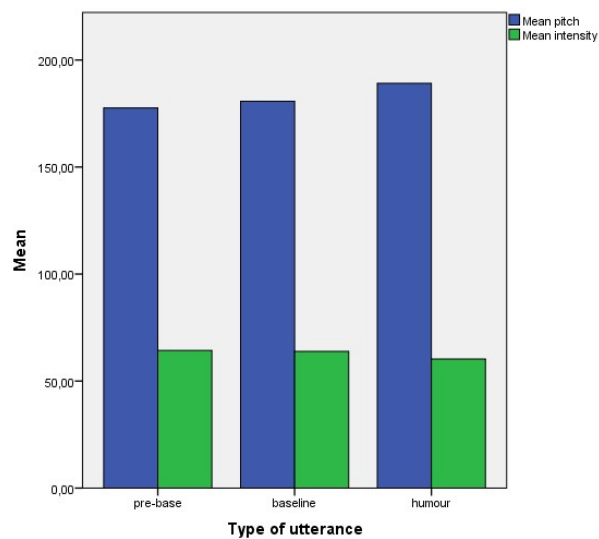
			Duration of utterance
Type of utterance	pre-base	1	2,220
		2	1,040
		Total	N
			2
		Sum	3,260
baseline		1	2,410
		2	1,140
		3	2,010
		Total	N
		Sum	5,560
humour		1	1,240
		2	1,680
		3	3,150
		4	2,310
		5	3,180
		Total	N
		Sum	11,560
Total		N	10
		Sum	20,380

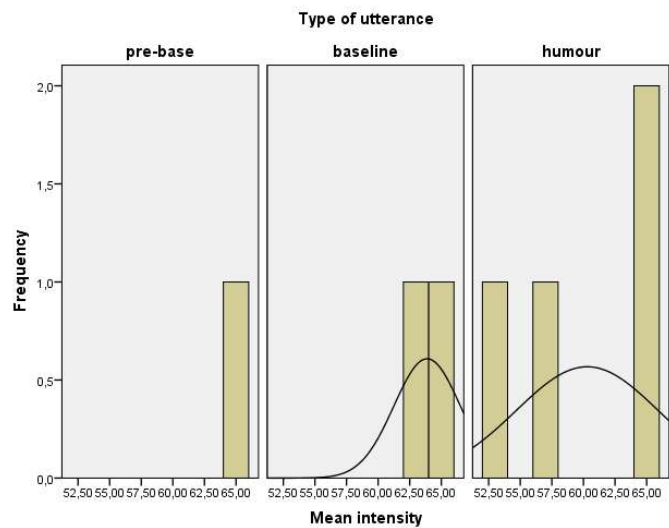
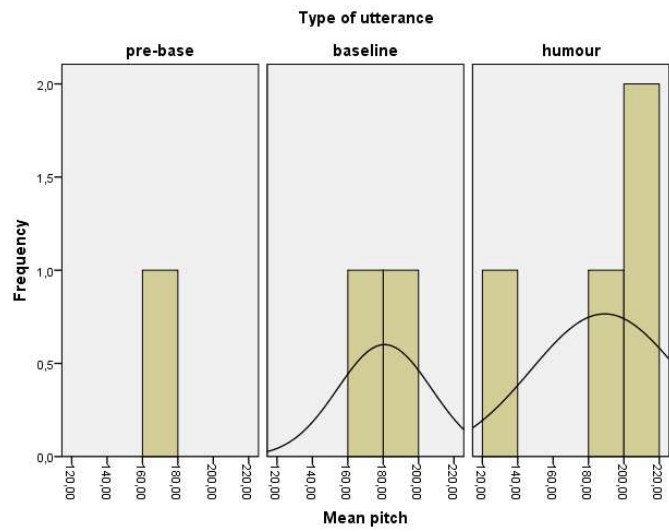
a. Limited to first 100 cases.



Report

Type of utterance		Mean pitch	Mean intensity
pre-base	Mean	177,6300	64,3300
	N	1	1
	Std. Deviation	.	.
baseline	Mean	180,7400	63,8750
	N	2	2
	Std. Deviation	26,55893	2,62337
humour	Mean	189,1150	60,2975
	N	4	4
	Std. Deviation	41,67631	5,62283
Total	Mean	185,0814	61,8957
	N	7	7
	Std. Deviation	31,81828	4,57728





One-Sample Kolmogorov-Smirnov Test

		Mean pitch	Mean intensity
N		7	7
Normal Parameters ^{a,b}	Mean	185,0814	61,8957
	Std. Deviation	31,81828	4,57728
Most Extreme Differences	Absolute	,202	,274
	Positive	,152	,201
	Negative	-,202	-,274
Kolmogorov-Smirnov Z		,534	,725
Asymp. Sig. (2-tailed)		,938	,669

a. Test distribution is Normal.

b. Calculated from data.

Group Statistics

Type of utterance	N	Mean	Std. Deviation	Std. Error Mean	
Mean pitch	baseline	2	180,7400	26,55893	18,78000
	humour	4	189,1150	41,67631	20,83815
Mean intensity	baseline	2	63,8750	2,62337	1,85500
	humour	4	60,2975	5,62283	2,81142

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Mean pitch	Equal variances assumed	,474	,529	-,251	4	,814	-8,37500	33,30574	-100,84657	84,09657
	Equal variances not assumed			-,299	3,307	,783	-8,37500	28,05204	-93,13663	76,38663
Mean intensity	Equal variances assumed	6,992	,057	,819	4	,459	3,57750	4,36744	-8,54845	15,70345
	Equal variances not assumed			1,062	3,940	,349	3,57750	3,36825	-5,83041	12,98541