Who is laughing now?

Laughter-infused dialogue systems

Vladislav Maraev

Department of Philosophy, Linguistics and Theory of Science

UNIVERSITY OF GOTHENBURG

2022

 The author is supported by grant 2014-39 from the Swedish Research Council for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

Cover design: Anastasia Zhikhartseva
Printed in Sweden by Stema Specialtryck AB
Publisher: University of Gothenburg (Dissertations)

Typeset in Bressay and Aller by Dalton Maag Ltd

ISBN 978-91-8009-909-7 (print)
ISBN 978-91-8009-910-3 (PDF)

Distribution:
Department of Philosophy, Linguistics and Theory of Science
Box 100, SE-405 30 Gothenburg

I hereby declare that the work presented in this thesis is my own work and that the research reported here has been conducted by myself unless indicated otherwise in the preface of each chapter.

Sammanfattning

Denna avhandling banar väg för att inkludera skratt i dialogsystem på ett domänoberoende och språkvetenskapligt giltigt sätt, med hjälp av verktyg och metoder från datalingvistiken. Avhandlingen behandlar tre huvudsakliga områden.

Det första området rör placeringen av skratt i relation till tal och andra beteenden. Vi visar att neurala nätverk kan förutsäga skratt från transkriberad dialog, medan människor är sämre på detta. Sådana modeller gör det möjligt för dialogsystem att förutsäga mänskligt skratt och, vid behov, placera skratt från systemet på lämpliga ställen. Vi undersöker också placeringen av skratt i förhållande till ögonrörelser och visar att skrattets funktion är relaterad till olika ögonrörelsemönster. Dessa resultat har viktiga följder för konversationella agenter och sociala robotar med avseende på generering och samordning av multimodalt beteende.

Det andra området rör interaktionen mellan skratt, den kommunikativa avsikten hos användare och system, och den kontext i vilken avsikten realiseras. Vi lägger grunden till den centrala komponenten i ett dialogsystem genom att implementera en teoretiskt motiverad dialoghanterare baserad på linjär logik. Vår dialoghanterare utvidgas sedan till att hantera skratt som utgör eller ackompanjerar återkoppling, och skratt som utgör svar på ja/nej-frågor. Därtill undersöker vi hur skratt kan modifiera eller forma en dialoghandling, samt hur skratt kan förbättra neurala modeller för igenkänning av dialoghandlingar.

Det tredje området är humor. Även om humor inte är en förutsättning för skratt så är de nära sammankopplade. Vi undersöker hur humor är relaterad till resonerande om sociala konventioner och andra inlärda och implicita antaganden. Vi visar hur dessa kan modelleras informellt och påvisar därigenom ett sätt på vilket situationsbunden kreativitet i samtal kan realiseras i konversationella agenter.

Abstract

This thesis paves the way for including laughter in spoken dialogue systems in a domain-general and linguistically valid way using computational linguistics tools and methods. The thesis is concerned with three main areas.

The first area concerns the placement of laughter in relation to speech and other behaviours. We show that convolutional and recurrent neural networks can effectively predict laughs from dialogue transcripts, whereas human perforance in this task is significantly worse. Such models allow dialogue systems to predict user laughter and, if needed, put system laughter in an appropriate place. Further, we look at laughter placement in relation to gaze and show that laughter performing different pragmatic functions is related to different gaze patterns. These findings provide important implications for embodied conversational agents and social robots in regard to multimodal behaviour realisation and coordination.

The second area is concerned with the interaction between laughter and the communicative intent of a user and system, as well as with the context in which the given intent occurs. We lay the groundwork for the central component of a spoken dialogue system by implementing a dialogue manager in a theoretically informed way using a proof-theoretic model based on linear logic. Our dialogue manager is then extended to support laughter functioning as feedback or as a signal accompanying system feedback, and an answer to polar questions. Additionally, we look at how laughter can modify or shape a dialogue act, and how the inclusion of laughter can improve Transformer-based deep learning models in the task of dialogue act recognition.

The third area is humour. Even though humour is not necessary for laughter, they are closely related. We look at how humour is related to reasoning about social conventions and other learned and accommodated implicit assumptions. We show how these can be modelled informally, suggesting one way in which the goal of situational and conversational creativity can be realised in artificial agents.

Contents

Abbreviations x

Acknowledgements xiii

Part I Laughter in spoken dialogue 1

1 Introduction 3
1.1 Motivation 3
1.2 Outline 4
1.3 Contributions 6

2 Ethical considerations 9
2.1 Alexa, laugh! 9
2.2 Dark patterns and deep fakes 10
2.3 Opacity 11
2.4 Sophia's rights 12
2.5 Who is laughing now? 13

3 Laughter 15
3.1 Is laughter important? 15
3.2 The question *why* 16
3.3 What is the laughter about? 18
3.4 What does laughter do? 20
3.5 What form and placement can laughter have? 22
3.6 How does laughter work? 23
3.7 Summary 25

4 Dialogue 27
4.1 Why do we need dialogue to study laughter? 27

4.2 Utterances, turns and dialogue acts 27
4.3 Grounding 29
4.4 Formal accounts for dialogue 30
4.5 Methodological approaches to dialogue 32

5 Spoken dialogue systems 35
5.1 Types of dialogue systems 35
5.2 Dialogue systems architecture 37
5.3 Dialogue management 38
5.4 Laughter in dialogue systems 39

Part II Laughter placement 43

6 Corpus analysis of laughter placement 47
6.1 Laughter placement and relevance 47
6.2 Placement statistics 48
6.3 Examples 51
6.4 Summary 53

7 Predictability of laughter relevance 55
7.1 The aim of this chapter 55
7.2 Data and the task 56
7.3 Baseline models 59
7.4 Deep learning models 62
7.5 Results 64
7.6 Error analysis 65
7.7 Discussion 68

8 Multimodal placement of laughter 71
8.1 Introduction 71
8.2 Gaze in interaction 73
8.3 Gaze and laughter in ECAs 74
8.4 Hypotheses 75
8.5 Methods and materials 76
8.6 Results 81
8.7 Discussion 87

Part III Laughter interpretation 95

9 Laughter which forms a dialogue act 99
9.1 Introduction 99
9.2 Dialogue acts and dialogue act recognition 100
9.3 Corpus study 101
9.4 Laughter for dialogue act recognition 112
9.5 Discussion 122

10 Dialogue management with Linear logic 125
10.1 Yet another dialogue manager? 125
10.2 Proof search as a programming language 127
10.3 Linear logic as a DM framework 128
10.4 Question-answering with metavariables 130
10.5 Linear Dialogue Manager 133
10.6 Summary 142

11 Non-humorous laughter in dialogue management 143
11.1 Introduction 143
11.2 Types of laughter 145
11.3 Formal treatment of non-humorous laughter 150
11.4 Discussion 153

Part IV Laughter and humorous incongruities 159

12 Humour and interaction 163
12.1 Why an interactive perspective on humour? 163
12.2 Theories of humour 164
12.3 Humorous conversation 166
12.4 What's in a laughable? 170

13 The interactive approach to humour 173
13.1 The elements of the theory 173
13.2 Mapping the joke to existing rhetorical resources . . 174
13.3 New resources, their acquisition and decay 177
13.4 The process of integration of new information 181

13.5 Emotional appraisal of the content 189
13.6 Summary 193

14 Case study I. Muffins and bagels: application of the theory 195
14.1 Introducing the joke 195
14.2 Resources 196
14.3 Acquisition and decay 197
14.4 Integration of new information 198
14.5 Appraisal 199
14.6 Summary 201

15 Case study II: The elephant riddle or further analysis of the integration process 203
15.1 The Elephant Riddle 203
15.2 Theory of Types with Records 204
15.3 Formal account of the integration process 205
15.4 Rhetorical reasoning in the Elephant Riddle 206
15.5 Discussion 211

16 Implications for dialogue systems 213
16.1 The aim of this chapter 213
16.2 Computational humour 213
16.3 Humorous laughter in dialogue systems 215
16.4 Rhetorical resources 217
16.5 Extending LDM with rhetorical resources 221

Part V Conclusions 225

17 Concluding remarks 227
17.1 Contributions 227
17.2 Future directions 230

Bibliography 237

Index 269

Abbreviations

AI	artificial intelligence
ALISP	automatic language independent speech processing
AMI	Augmented Multi-party Interaction
AMI-DA	AMI Meeting Corpus
AMT	Amazon Mechanical Turk
ASR	automatic speech recognition
BERT	Bidirectional Encoder Representations from Transformers
BNC	British National Corpus
CA	Conversational Analysis
CNN	convolutional neural network
CRUM	Computational-Representational Understanding of Mind
DA	dialogue act
DAMSL	Dialog Act Markup in Several Layers
DAR	dialogue act recognition
DEC	Directory Enquiries Corpus
DGB	Dialogue GameBoard
DM	dialogue manager
DS	Dynamic Syntax
DS-TTR	Dynamic Syntax and Type Theory with Records
EAC	ELAN Analysis Companion
ECA	embodied conversational agent
FTA	face-threatening act
GHI	Good Housekeeping Institute

GMM	Gaussian mixture model
GTVH	General Theory of Verbal Humour
HCI	human-computer interaction
HMM	Hidden Markov Model
ICM	Interactive Communication Management
ISU	information-state update
LDM	Linear Dialogue Manager
LSTM	long short-term memory
NHS	National Health Service
NLG	natural language generation
NLP	natural language processing
NLU	natural language understanding
NN	neural network
NSU	non-sentential utterance
OSTH	Ontological Semantic Theory of Humour
QUD	question under discussion
RELU	rectified linear unit
RNN	recurrent neural network
SASSI	Subjective Assessment of Speech System Interfaces
SSTH	Semantic-Script Theory of Verbal Humour
SVD	singular value decomposition
SVM	support vector machine
SWDA	Switchboard Dialogue Act Corpus
TRP	transition-relevant place
TTR	Theory of Types with Records
TTS	text-to-speech synthesis
VADER	Valence Aware Dictionary and sEntiment Reasoner
VR	virtual reality
VUI	voice user interface

Acknowledgements

At the end of the distance, two things are extremely difficult to forget: how much fun I had along the way, and how many people have helped me and inspired me. This will be a rather hopeless attempt to express all my gratitude and excitement without using any non-verbal signals.

I would like to start with my wonderful supervisors: Christine Howes, Staffan Larsson and Jonathan Ginzburg. Without your patient support, I would not have made it. Thank you, Chris, for keeping me within the reasonable limit, for making me finish the actual thing, and for the thorough reading of my manuscript. Thank you, Staffan, for many fruitful discussions and for sharing similar attitudes to blending practicality and blue skies research. Thank you, Jonathan, for the great kick-off of my journey, and your encouragement along the way. All of you contributed massively to the actual research reported here.

A very special thank-you goes to Chiara Mazzocconi. We have done so much work together and had so many laughs. Our endless conversations have been endlessly rewarding. I have been learning a lot from you and with you.

Two people were extremely important for the final version of this dissertation: my external examiner, Arash Eshghi, and Eleni Gregoromichelaki. Arash, thank you so much for your comments and the brilliant and flattering presentation of my research at the final seminar. Eleni, thank you for reading the 'very final' version, your comments were extremely thorough and helpful.

Jean-Philippe Bernardy has been most helpful with many parts of my thesis, providing support, inspiration and sharing great attention to detail. We have had countless valuable discussions and pair programming sessions.

Ellen Breitholtz shared the burden of one of the most puzzling parts of my dissertation – the one related to humour. We had many discussions which almost always were ending up with even more questions than we had at the start. Now I can see topoi and elephants everywhere!

Bill Noble has helped me immensely with the parts concerning dialogue acts and laughter. It was so much fun discovering new ideas and new puzzles.

Additionally, I would like to thank Robin Cooper, Julian Hough, Ye Tian, Gregory Mills and Vidya Somashekarappa who have provided very valuable (and documented) input along the way.

I am extremely grateful to CLASP and its present and past members for the best research environment that I ever encountered. I believe I have learned something from each one of you. It is always delightful to come to the university building. And I can't forget to mention Talkamatic for the very entertaining distraction with practical tasks and for being able to think out loud.

It was extremely important for me to be pushed in the right direction which ended up in a very enjoyable time as a PhD student and candidate. In this context, I would like to thank Peter Mitsov, Marina Maraeva and Antonío Branco.

Many thanks to my friends in St. Petersburg, Moscow and those scattered around the globe – you have influenced me and brightened many moments of my life. Masha and Egor, thanks for coping with me and my writing stress at the Veschevo dacha.

I would like to thank my mother, Mila Maraeva and my father, Boris Kantor – I always know that I can turn to you to get as much support and love as I ever need.

Lastly, I would like to thank my girlfriend Nastia Zhikhartseva for boosting the fun and enjoyment of life along with my thesis completeness to 100% (the cover included). I cannot imagine how stressful and boring these final months would have been without you.

Göteborg, August 2022
Vlad Maraev

Part I

Laughter in spoken dialogue

1 Introduction

Смех без причины — признак дурачины.
Laughing for no reason is a sign of stupidity.

Russian proverb

1.1 MOTIVATION

Laughter is ubiquitous in our everyday interactions. In the British National Corpus (BNC) laughter is a frequent non-verbal social signal regardless of gender and age. The spoken dialogue part of the BNC contains approximately one laughter every 14 utterances. In the Switchboard Dialogue Act Corpus (SWDA)[1] non-verbally vocalised dialogue acts (whole utterances that are marked as non-verbal, around 65% of them include laughter) constitute 1.7% of all dialogue acts and laughter tokens make up 0.5% of all the tokens that occur in the corpus.

Laughter is produced in different *dialogical* contexts: from moments of social discomfort to joyful situations, and associated with different emotional states and intentions of interlocutors: from reacting to a humorous remark to accompanying a critique, or even (as we shall see) answering a question. Laughter has an important social function: it creates social bonding between people, which also means that others may be excluded from this group.

[1] Jurafsky et al. (1997a), US English, one-on-one interactions over a phone where participants that are not familiar with each other discuss a potentially controversial subject, such as gun control or the school system

To our knowledge there has been no systematic approach to laughter in dialogue systems, jointly developed with a corresponding dialogue theory. Given that it is not feasible to address all aspects of laughter in relation to dialogue systems in one thesis on a sufficiently deep level, we have identified three crucial areas, around which our research is focussed.

The first is concerned with laughter placement. This is important for a dialogue system which attempts to make sense of laughter in relation to verbal and visual contexts. Where is laughter more likely to occur? Are different types of laughter positioned differently? Does it depend on who is laughing and about what? Is it possible for a computer (a dialogue system) to predict laughter and how easy it is for humans? How does laughter coordinate with visual signals such as gaze?

The second area is concerned with the *change* that laughter brings about. How does laughter relate to the dialogue context, and how does laughter change the context? How does it interact with actions in a dialogue? What would be the proper way to represent laughter for a dialogue system and how can these processes be formally addressed?

Finally, the third area considers humour as one of the contexts for laughter. Previous studies of humour in spoken dialogue systems have either been concerned with canned jokes, or have studied humour independently of its interactive setting and have not been unified with a general conversational model. What does it mean for an element of discourse to be humorous? What aspects of humour can be generalised to, and unified with, established approaches to dialogue?

The three central parts (II–IV) of the thesis aim to fill the knowledge gaps in these areas.

1.2 OUTLINE

The overall structure of the study takes the form of four parts, further divided into chapters:

Part I: Laughter in spoken dialogue

This part begins by laying out the basic concepts of dialogue, such as utterances, dialogue acts and grounding. It continues with a brief overview of approaches to dialogue systems design and architecture. Finally, it introduces the main subject of this thesis – laughter – and outlines the main concerns of previous studies in relation to dialogue and dialogue systems.

Part II. Laughter placement

When considering laughter in natural language understanding (NLU) and natural language generation (NLG) on the level of *syntactic structure and propositional content*, it is important to investigate how laughter is placed in relation to speech and other behaviours. This part analyses laughter placement from two different perspectives.

The first perspective presents a general outlook on laughter placement and predictability. We analyse the statistics of laughter placement in relation to its antecedent – *the laughable* – within different dialogue corpora. Further, we use artificial neural networks to predict laughter from transcripts of natural dialogues, and compare machine performance to human performance.

The second perspective takes laughter placement in relation to non-verbal behaviour, namely, the direction of eye gaze. We look at how different types of laughs are accompanied by distinct gaze patterns, both by the participants who produce laughs and their conversational partners. We also look at the importance of gaze in laughter coordination and reciprocation. This part has additional implications for the behaviour realisation component of embodied conversational agents (ECAS).

Part III: Laughter interpretation

Besides understanding what laughter relates to and how it is placed, it is crucial to see how laughter interacts with the communicative intent of a user and a system, as well as with the context in which the given intent occurs. Hence, this part analyses how laughter in different contexts can be interpreted by a spoken dialogue system.

Our interpretation includes a surface interpretation on the level of communicative intents, and a deeper interpretation on the level of the information states of dialogue participants. We discuss the implemented framework, which allows expressing transformations in information states, and show how it helps to address some non-humorous cases of laughter in the dialogue manager (DM) component of a dialogue system.

Part IV: Laughter and humorous incongruities

The most intuitive case of laughter – in response to something humorous – appears in the final part. We argue that in dialogue systems, we need to look beyond the traditional approach of writing canned jokes for a system and consider more spontaneous and interactive sorts of humour. To do this, one needs to better understand the properties of *humorous laughables*, which can potentially lead to a better understanding of how humour is related to laughter.

We take existing theories of humour as a point of departure and consider humour on a general level, taking a dialogical approach to written jokes. We introduce dialogical rhetorical resources – *enthymemes and topoi* – which we use to formally define the humorous incongruity present in the context of a laughable. We build a theory that accounts for humour from an interactive perspective and apply it to several cases of humour. Finally, we discuss how task-oriented humour reflected in a laughable can be treated in a dialogue system.

1.3 CONTRIBUTIONS

This thesis consolidates three strands of research: i) dialogue management, ii) laughter in dialogue, and iii) humour studies. We attempt to pave the way for including laughter into spoken dialogue systems. This is done in a domain-general and linguistically valid way using computational linguistics tools and methods.

Part II. In Chapter 6 we underline the importance of laughter placement in relation to speech and other behaviours for understanding its meaning. Chapter 7 investigates more specifically how laughter can be predicted with artificial neural models from

dialogue transcripts. In Chapter 8 we study laughter placement in relation to other modalities, specifically the direction of gaze, demonstrating that laughs performing different pragmatic functions are associated with distinct gaze patterns and that gaze is crucial for laughter coordination. These findings provide important implications for embodied conversational agents and social robots in regard to multimodal behaviour realisation and coordination.

Part III. In relation to the change that laughter can bring into dialogue context, in Chapter 9 we look at how laughter can modify or shape a *dialogue act*, and how laughter can improve Transformer-based deep learning models in the task of dialogue act recognition. Chapter 10 prepares the ground for integrating laughter into a DM that reflects the dialogue context, by implementing a DM in a theoretically informed way. The DM is based on the KoS framework (Ginzburg, 2012) and mapped into a proof-theoretic model based on linear logic. In Chapter 11 we extended the DM to support laughter functioning as feedback, as a signal accompanying feedback, or as an answer to polar questions.

Part IV. Humour is intuitively related to laughter, and most instances of laughter are related to pleasant incongruities (Mazzocconi et al., 2020). Moreover, some laughs highlight social norm violations and ironic statements, which bring laughter and humour closer together, even though humour is not necessary for laughter (Chapter 12). In Chapter 13 we investigate how humour is related to reasoning about social conventions and other learned and accommodated implicit assumptions, and how humour can be modelled in relation to creativity. The next two chapters provide two case studies: in Chapter 14 we explain our account in full using a short conversational joke as an example, and Chapter 15 provides a formal account for integrating new information during a humorous dialogue. Chapter 16 presents the implications for situational and conversational humorous creativity in artificial agents.

We believe that this thesis contributes to linguistic research by laying the groundwork for designing dialogue systems that can be used to test theoretical insights about human conversation. It demonstrates how laughter contributes semantic and pragmatic

import to dialogue. The potential of this work includes making future dialogue systems more reactive to laughter. This enables natural and highly responsive behaviour from an artificial agent in interactive settings. Developing a precise account of laughter within the information state of dialogue participants allows scaling up to other non-verbal social signals, including smiling, frowning, sighing and eye-rolling (Ginzburg et al., 2020).

Nevertheless, this thesis does not fully engage with all aspects related to laughter in dialogue systems. It does not engage with laughter in automatic speech recognition (ASR) or text-to-speech synthesis (TTS), nor in laughter behaviour realisation in ECAS. Emotion modelling for artificial agents is another important aspect which is central for laughing behaviour of humans and dialogue systems but which lies beyond the scope of the thesis. Additionally, our models of laughter and humour interpretation are not probabilistic (though potentially adaptable to frameworks for reasoning under uncertainty).

2 Ethical considerations

2.1 ALEXA, LAUGH!

In the beginning of 2018 users of Alexa, Amazon's dialogue system, reported a few instances of unearthly behaviour – she was producing eerie laughter for no particular reason. This was an automatic speech recognition (ASR) issue with misrecognition of background noise as the request to laugh: 'Alexa, laugh'. 'We are changing that phrase to be "Alexa, can you laugh?" which is less likely to have false positives, and we are disabling the short utterance "Alexa, laugh."', Amazon explained in their emailed statement. They also said that instead of laughing when asked whether it can laugh it will first acknowledge the question by saying 'Sure, I can laugh.'[1, 2] In other words, Amazon themselves brought this issue about and found an engineering solution to it. The issue itself and the solution illustrate Alexa's developers' lack of ethnographic awareness – it is evident that one rarely asks human beings whether they can laugh[3] or asks them to laugh. This little example gives us a glimpse of how low the relevance of 'laughter ethics' for current dialogue systems is.

In this thesis we largely observe laughing and humorous behaviours in human-human interaction qualitatively and statistically, and model them in computational systems. This is not drastically

[1] https://www.nytimes.com/2018/03/08/business/alexa-laugh-amazon-echo.html

[2] Notably, a human will usually interpret such 'questions' as an indirect requests, not queries.

[3] We searched in the dialogue part of the British National Corpus (BNC) for 'Can you laugh' using SCoRE search tool (Purver, 2001) and found no matches whatsoever. Such a question posed to a dialogue system does not assume the answer to be as obvious as for humans, but we are not aware of any empirical evidence for this.

different from the methodology which constituted the famous Turing test for machine intelligence (Turing, 1950) – we can attribute intelligence only by making inferences from observable behaviours.

We are aware that giving a dialogue system new capabilities as a result of studying human (linguistic) behaviours might raise new ethical questions in the future. Our central argument here is that potential harm arises not from the advancements of our studies but from the misuse of research outcomes, and from political and commercial decisions which are either improvident or malicious.

In this chapter we go slightly beyond the scope of dialogue systems and focus on three potential problems with future artificial intelligence (AI) systems: manipulation, opacity and competition with humans.

2.2 DARK PATTERNS AND DEEP FAKES

There have been many cases of using AI to manipulate human behaviour. Many industries have been using cutting-edge technology to maximise their profits. Prime examples of this are online gaming and gambling industries, but the scope of manipulations is not limited to them – it includes social media and different kinds of online marketing.

Even without any help of AI, many websites use so-called *dark patterns* – deceptive design patterns that benefit the platform and steer the user into making unintended (and potentially harmful) decisions (Mathur et al., 2019; Gray et al., 2018). Some of the *dark patterns* are exploiting cognitive biases, such as sunk cost fallacy (or Concorde fallacy) bias (Arkes and Ayton, 1999) – if someone has invested resources into some action, they tend to continue the endeavour otherwise they would think that the resource has been wasted. Mathur et al. (2019) discuss this pattern, along with a taxonomy of other cognitively informed dark patterns, in relation to manipulations used by online shopping websites. This raises the question of the awareness among the behavioural and cognitive scientists in regard to how their research outcomes might inform the designers of deceptive user experiences.

In recent years, a vast increase in the amount of data and computational power has caused the rise of AI systems based on deep learning techniques. Basing their predictions on past user behaviours, they can predict future user actions. The rise of deep learning has resulted in various 'faking' techniques, including creation of fake news articles, photos and videos – 'deepfakes' (a blend of 'deep learning' and 'fake'). Müller (2021) claims that 'Soon, sophisticated real-time interaction with persons over text, phone, or video will be faked, too.' We are doubtful about this happening in the near future. For instance, Google Duplex[4], an AI assistant used to make phone reservations (e.g. at restaurants), does not seem to deliver on its ambitious promises.[5]

In summary, manipulation and 'faking' techniques raise many concerns about trust in digital interactions.

2.3 OPACITY

Another significant aspect of AI research in the age of big data is opacity. Deep learning algorithms trained on vast amount of data tend to make generalisations which are susceptible to biases. For instance, a system trained on historical employment records (in order to get more data) might exploit gender, racial and other forms of discrimination that were common in the past, even if now such practices are less prominent. A daunting aspect of this is that such information is often opaque, both for end-users (e.g. the employment candidate and HR officer) and system developers. In Maraev, Breitholtz, Howes and Bernardy (2021) we proposed a distinction between *passive explainability* where system behaviour is explained by its developer through analysing its internal state, and *active explainability* where the system explains its own decision in a way humans can understand (i.e. by having a conversation). In

[4]https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

[5]https://www.theverge.com/2021/4/1/22361729/google-duplex-ai-reservation-availability-49-us-states

the latter case we hypothesise that active explainability can increase trust in a dialogue system.

Questions of bias and opacity resonate with the topic of humour in an interesting way. In Part IV of this thesis we claim that humour uses rhetorical resources, some of which are opaque (or covert) by nature – humour often relies on something which is implicitly assumed. Humour often plays with biases and stereotypes, and sometimes this play is aggressive and taboo-breaking. Even though the impact of jokes on society is in some respects questionable (MacHale et al., 1997), jokes (supposedly successfully) created with the help of biases learned from data might be harmful and offensive for some people. The question of offence inflicted by AI is tightly connected to its moral status, which we touch upon in the next section.

2.4 SOPHIA'S RIGHTS

In October 2017, Sophia – the human-like robot from Hanson Robotics – was awarded an honorary Saudi Arabian citizenship. This opened a huge debate in regard to robot rights. Such an act can be seen as undermining the rights of human beings, especially women and migrant workers in Saudi Arabia.[6] Bryson (2018) draws attention to the problem of AI competing with humans for resources and a higher social status, and argues that it would not be possible to create coherent ethics if we grant subjectivity to AI artifacts. Importantly, this media hype and subsequent debate has taken place despite Sophia's very limited abilities, especially in its conversational skills.

This poses a question whether advancements in behavioural research and dialogue systems would contribute to a higher moral status of AI. Danaher (2020) argues that even 'rough performative equivalence' with humans can be sufficient for granting AI a significant moral status, and that the threshold for reaching it does

[6] https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia

not have to be high, and might even have been already reached. Bryson (2018) points out that it is possible for us to create AI with high moral status but we should not do it because of negative costs (moral equilibrium in society, moral and legal responsibilities for robots' actions etc.). In light of Danaher's and Bryson's arguments it seems that the core issue here concerns ascribing agency to artificial entities, rather than raising the bar of their performative capabilities. On the other hand, better performance will make ascription of moral agency more justifiable and likely.

It remains an open empirical question whether the ability to be humorous and to use laughter appropriately would contribute to a higher likelihood an artificial entity being ascribed agency.

2.5 WHO IS LAUGHING NOW?

Alexa's laughter was a good example of contextual unawareness, which is why her laughter was called 'creepy' and 'eerie' by some of the users who encountered it. In this thesis we lay the groundwork for making future dialogue systems more sensitive to context in relation to humour and laughter. In light of the concerns that we raised in this chapter, it is also important to counterbalance them with some of the potential benefits of adding laughter to dialogue systems and making the interactions with them more natural.

Even minimal contextual awareness can benefit the interaction with a dialogue system. Hu et al. (2021) showed that laughter which is inserted as an interjection echoing a positive emotion detected in a user's speech can improve the perceived emotional intelligence ratings of the dialogue system. Maintaining trust and empathy in dialogue systems is important for systems aimed at learning and companionship. For instance, learning a new language is important for an immigrant to be included in a society. Therefore a language-tutoring dialogue system, which could teach a new language through natural interaction with the help of humour and laughter, could be a very helpful tool.

We do not intend to give any strong recommendations at this point, but the practical implications of this thesis have to be care-

fully considered together with the issues that we have discussed in this chapter.

3 Laughter

3.1 IS LAUGHTER IMPORTANT?

Laughter is crucial for our daily dialogues and relationships, it is omnipresent in nearly all dialogue corpora regardless of gender, age and the seriousness of the topics discussed. Laughter has long been of great interest in a wide range of fields. However, there is no agreement on the causes of laughter, with, for example, some research which focusses on humour (Hempelmann and Attardo, 2011; Raskin, 1985), other research which highlights the social functions of laughter, such as affiliation and agreement (Chapman, 1983; Scott et al., 2014) and qualitative analysis of the roles of laughter in interaction and its coordination with speech (e.g., Glenn, 2003).

A much debated question is to what extent laughter is under voluntary control. Despite a very particular bodily reaction (laughter causes tensions and relaxations of our bodies), it is believed that we laugh in a very different sense from sneezing or hiccuping (Ginzburg et al., 2020; Prusak, 2005; Plessner, 1970). Many scholars agree that we laugh for a reason, *about* something (see Section 3.2).

One of the most prominent arguments against involuntary laughter is its social function, which is well-documented (e.g., Mehu, 2011): laughter is associated with a sense of closeness and affiliation, establishing social bonding and smoothing away discomfort. According to Provine (1993), humans are 30 times more likely to laugh in the presence of others. Even the 'primitive' case of tickling not only requires the presence of another (self-tickling is almost impossible), but tickling stimulation is also more likely to elicit laughter if the subjects have a close relationship (Harris, 1999). Therefore, it is hard to claim that this highly socially dependent behaviour is involuntary.

A question about laughter, 'Why are you laughing?' is very different from, for instance, 'Why are you sneezing?'. The latter only addresses bodily function[1] (we can explain it by referring the flu or an allergy, for example), while the former is explained in relation to either the content ('because *it* is funny') or, one's perception of self in the situation ('because *I* am in a great mood'). Except on rare occasions, for example, of drug use (e.g. nitrous oxide), laughter is unlikely to be explained in relation only to the conditions of the body. In general terms, this means that laughter is nearly always voluntary. Nevertheless, it is important to distinguish an actual event of producing laughter with post-hoc explanations by a laugher or an observer.

Since laughter is very natural for humans it ought to be implemented in dialogue systems to make them more natural and responsive, for instance, it can be desirable for companion robots for the elderly. Recent research has focussed on creating more human-like spoken dialogue systems by means of adding the capability to produce (Ding et al., 2014), or recognise laughter (Truong and Leeuwen, 2007; Tahon and Devillers, 2015; Kaushik et al., 2015) react appropriately, (Niewiadomski et al., 2013; Haddad et al., 2016), recognise sarcasm (Tepperman et al., 2006) and be humorous (Katevas et al., 2014; Nijholt et al., 2017).

Here we set up the background to our investigations of laughter in dialogue around the question *why*, with the aim of addressing it from different perspectives.

3.2 THE QUESTION *WHY*

Let us consider the following example of *speech-laughter* (laughter which co-occurs with the same person's speech):

[1] Although sneezing can sometimes be psychosomatic (Gopalan and Browning, 2002), it is believed to be not volitional. The situation with coughing and sighing is different. Although we can sigh and cough deliberately (e.g. to express displeasure), unlike laughter (except in rare medical conditions) it can be a reflex: we cough to free our airways and breath deeply from time to time in order to normalise respiration.

(3a) from the British National Corpus (BNC):

18	*Interviewer.*	…[cough] Right, you seem pretty well qualified.
19	*John.*	I hope so <laughing: yes> erm

BNC JNV

Why did John laugh? It is a puzzling question, if one wants to be as specific as possible. Even on a general level there are several different types of causal explanations of John's laughter.

Given the context of the interview, John was probably flattered by the interviewer's comment about his qualification, and laughter was a way to deal with the compliment. Alternatively, he might have been accompanying his reaction ('I hope so') or trying to disguise his direct commitment to the interviewer's claim out of politeness. Choosing between these possibilities is hard, but could be facilitated by giving more context, listening to the recording and watching the video of this interaction episode.[2]

For current purposes, let us focus on the latter explanation – we can say that John's laughter has something to do with him uttering 'yes'. We can also say that John laughed because he wanted to mitigate the effect of uttering a straight 'yes' (the hedge 'I hope so' also smooths the straight answer).

We can also speculate about the form of the laughter, why it was placed this way, produced with a certain level of arousal and blended with speech. Additionally we should consider its placement in regard to 'yes' – the referent of the laughter.

We can also consider the semantic and pragmatic mechanisms which would lead to an interpretation of the laugh that was desired by John. These mechanisms potentially should account for misinterpretation of laughter or being unable to understand it. For example, in case of misunderstanding, the interviewer might have uttered a clarification request like 'why are you laughing?'.

[2] Not for the given case though, as only audio recordings are available in the BNC.

These types of explanations can be formulated as the following questions:[3]

1. What is the laughter about? (Section 3.3)
2. What does the laughter do? (Section 3.4)
3. What form and placement can the laughter have? (Section 3.5)
4. How does the laughter work? (Section 3.6)

3.3 WHAT IS THE LAUGHTER ABOUT?

The first cause relates to the circumstances under which we can produce laughter. This cause includes the perceived context and expectations of the interlocutors regarding possible changes of context. We can highlight at least three intertwined dimensions of context: social, linguistic and emotional. The social context includes interlocutors' impressions of the social situation, including the social hierarchy and social situation with respect to tension and discomfort, social face (Goffman, 1955; Brown and Levinson, 1987) and personae (Eckert, 2008; Burnett, 2019, among others). Linguistic context includes the current state of the dialogue judged from the speakers' perspectives. Emotional context includes the emotional states of participants, for instance their mood and levels of stress and anxiety.

The circumstances of laughter can be the subject of post-hoc explanations: for example, 'I laughed because I am nervous' and 'I laughed because it is very funny'. Nevertheless, addressing such explanations for investigating the nature of laughter would not be sufficient. It is impossible to recover from them the actual state

[3]Our analysis vaguely follows Aristotle's account of causality. In *Metaphysics V 2* and *Physics II 3* he describes the four causes: the efficient cause, the final cause, the material cause and the form cause (Falcon, 2019). This description of causality is also inspired by James Pustojevsky in his description of qualia structure of Generative lexicon (Pustejovsky, 1991).

of affairs at the time of laughter, moreover, Vettin and Todt (2004) provide anecdotal evidence that people are not aware of how frequently they use laughter.

We follow Ginzburg et al. (2015) and Mazzocconi et al. (2020) in their borrowing of the term *laughable* from Conversational Analysis (CA) (Glenn, 2003) as denoting something that laughter *relates to* while being agnostic about whether it is humorous or not. The closest linguistic analogue for a laughable is an anaphoric antecedent. In example (3a) above, the laughable is the word 'yes'. The laughter and laughable can be positioned quite freely, according to Tian et al. (2016) and the laughable can follow, precede or overlap with the laughter. In Chapter 6 we take closer look at how the laughable is placed depending on who produces the laughter and the form of laughter.

Another important component is *incongruity* as a factor that can describe an idiosyncratic relation between the situation and the laughable. According to the definition of incongruity provided by Mazzocconi et al. (2020)[4], it '[...] involves a clash between a general inference rule (a *topos*) and a localized inference (an *enthymeme*)' (in Chapter 12 we further explain these notions). Therefore, in our example (3a) the incongruity arises from the mismatch between the intended meaning of 'yes' (as in 'yes, I am competent, but I am not cocky') to the one communicated by pure 'yes'. Even though it is challenging to provide an analytic definition of incongruity (it can be vaguely defined as something not ordinary, e.g. violating expectations), we can define it by giving instances of incongruities of various sorts: a mother walking like a penguin, an interviewer's impolite question, an unusual use of a word.

The term 'incongruity' can be found in numerous humour studies (see Chapter 12 for the review). The relation between humour and laughter is such that not all humour leads to laughter and not all laughs are preconditioned by humour. Nevertheless, humour draws on knowledge resources and common-sense reasoning which

[4]The definition of incongruity in terms of topoi and enthymemes was first proposed by Ginzburg et al. (2015).

are also engaged (by their potential violation, for instance) in other occasions which result in laughter. In case of laughter in conversation, it can be used to smooth a criticism (direct criticisms are usually considered impolite). The same norm can be violated in a humorous way, often covertly. Political humour is often driven by the purpose of criticising a certain individual or group. For instance, consider a joke: 'If you are not part of the solution, you're probably running for President'. The covert critique here is directed to a person running for President, and hints at them being a problem.

Although laughter is closely associated with humour, and humorous and joyful remarks can be thought of as a prerequisite for laughter, this is not necessarily the case: especially in dialogue laughter can display surprise, nervousness, embarrassment, disagreement etc. (Poyatos, 1993). This suggests that laughter is not exclusively associated with positive emotions (happiness, joy, pleasure and more) – other emotional dimensions and their (perhaps contradictory) combinations should also be considered. In Part IV we aim to bridge the gap between humour and laughter by formalising their common mechanisms.

3.4 WHAT DOES LAUGHTER DO?

In our example John smoothed his utterance 'yes' – he modified its content, if we want to be more specific. In this way he has done something with his contribution using laughter as a 'tool'. In CA laughter is viewed as a resource which is available for the participants in a conversation. Along with other resources participants use it jointly and locally, indicating to each other what they are doing and thus maintaining intersubjectivity (Petitjean and Morel, 2017). From this perspective, laughter functions in coordination with interaction. For example, many laughs can serve as backchannels, which, either alone or together with some of the other functions, can signal understanding of the previous utterances. CA views our interactions as organised and systematic, and where laughter performs a variety of actions: signalling non-seriousness (which is preferred to the notion of humour), managing turn-taking, and dealing with

misalignment and misunderstanding. Actions of a longer reach are also considered, such as identity construction. For an overview, see Glenn (2003) and Glenn and Holt (2013).

Mazzocconi et al. (2020) present a taxonomy of laughter functions from the perspective of the 'effect that the laugher intends her own laughter to have'. They present a decision-tree based framework and use it to annotate laughs in various corpora. The decision tree takes into account four types of incongruity that characterise a laughable and groups the functions accordingly:[5]

Pleasant incongruity '[…] any cases in which a clash between the laughable and certain background information is perceived as witty, rewarding and/or somehow pleasant' (Mazzocconi, 2019). Laughter functions associated with pleasant incongruity include showing enjoyment of a pleasant incongruity (e.g., laughing at a joke), and marking or recognising incongruity (e.g., to indicate absurdity of the situation which is otherwise not salient to the others)

Social incongruity '[…] a clash between social norms and/or comfort and the laughable.' (ibid.) Laughter functions associated with social incongruity include softening and trouble-telling (e.g., to accompany a criticism), benevolence induction (e.g., to accompany self-criticism), smoothing (e.g., example (3a) above), and showing sympathy (e.g., to sympathise with a weakness of an interlocutor)

Pragmatic incongruity '[…] incongruity that arises when there is a clash between what is said and what is intended.' (ibid.) Laughter functions associated with it include marking irony (e.g., 'I <laughter> like Alan's style'), scare quoting or lexical enrichment (e.g., 'I like Alan's <laughter> style'), and lexical uncertainty or editing phrase (e.g. it can be used for repairing previous input)

[5]As Mazzocconi (2019) reports, Krippendorff's α inter-annotator agreement for each of the four groups range from 0.37–0.72 for BNC (3 coders) and 0.65–0.71 for the French part of DUEL (Hough et al., 2016) (2 coders).

No incongruity 'In these cases what is associated with the laughable is a sense of closeness that is either felt or displayed towards the interlocutor, e.g., while thanking or receiving a pat on the shoulder.' (ibid.). Laughter functions that are not related to any incongruity include demonstration of friendliness and affiliation, for instance 'Right, thanks Fred. You're on holiday after today? – Lovely. <laughter>'.[6]

From the functional perspective it is also necessary to mention that laughter has intrinsically important social effects, being crucial for bonding and managing relationships, while also being immensely influenced by social context (Fridlund, 2014; Devereux and Ginsburg, 2001; Provine, 2004).

3.5 WHAT FORM AND PLACEMENT CAN LAUGHTER HAVE?

The resulting laughter will then have a particular form, which can be characterised objectively, from the point of view of producing the corresponding sound and moving the body in a particular way. In example (3a) we should note that laughter was produced simultaneously with speech (so called *speech-laughter*). The alternative is to produce laughter in isolation when it is not blended with the laugher's speech (so called *laughter bouts*). Below we show the contrast between them in the transcription convention used throughout the thesis.

(3b) (speech-laughter) I hope so <laughing: yes> erm

(3c) (laughter bout) I like Alan's <laughter> style.

Laughter bouts can be classified depending on their level of intensity (Mazzocconi et al. (2020) coded in their corpus study three levels of arousal: Low, Medium and High), but there can be much more fine-grained descriptions, down to observing the spectrogram of a laughter and determining the exact phonetic and prosodic

[6] from the BNC, example from the conversation in a bar (BNC KDP)

realisations of it. Laughter is characterised by high variability of acoustic and phonetic characteristics which exhibits individual patterns (Urbain and Dutoit, 2011). Mazzocconi (2019) reports significant main effects of acoustic features (F0 Mean, F0 Max and Harmonic to Noise Ration) on type of incongruity, which in their analysis only included social and pleasant incongruities.

Laughter can also be viewed not only individually but in relation to the behaviour of other dialogue participants. This can include whether people tend to laugh together, whether their laughter can sometimes overlap, or whether one person finishes laughing when the other starts speaking.

Another aspect is how the position of a laughter selects a specific laughable. Let us consider two examples:

(3d) I <laughter> <u>like Alan's style</u>.

(3e) I like Alan's <laughter> <u>style</u>.

In (3d) the laughable (underlined) is the whole *verbal phrase* which can be negated (given that laughter marks irony here) or modified in a some other way similar to adverbial modification. In contrast, in (3e) the laughable is just the word 'style', where laughter will add a modification of meaning of this *word* (something similar to scare quotes, meaning something like 'weird style').

In Part II we will take a closer look at how a laughter and laughable can be placed in relation to each other, and in Part III we see how their joint interpretation can produce new meanings.

3.6 HOW DOES LAUGHTER WORK?

The mechanisms of laughter should serve as a perspective which integrates all the other elements. One of the key ideas we draw on here is a multi-level outlook on laughter: from its form and placement in relation with the laughable to its meaning. The meaning itself also has several levels. We distinguish the basic core meaning, and the integrated meaning in the context of the current state of affairs in a dialogue.

Ginzburg et al. (2020) claim that laughter conveys a propositional content and they view the core meaning of laughter as predication $P(l)$, where the predicate P can either relate to incongruity or pleasantness. Oatley and Johnson-Laird (2014) have shown in their survey that emotions not only reflect the physical states of the agents but also function as judgements, depending on the current state of affairs. For instance, the emotion evoked can depend on the significance or urgency of the event for a certain person in a given moment. Such an evaluation is called *appraisal*. Ginzburg et al. (2020) state that laughter arises from the appraisal process of the laughable, following the literature on cognitive theories of emotions (e.g,. Marsella and Gratch, 2009; Oatley and Johnson-Laird, 2014; Scherer, 2009). We return to appraisal in relation our dialogical approach to humorous incongruities in Chapter 13.

In order to lead to laughter as a response, certain judgements about the situation are required, including i) judging the relation between the situation and the laughable as incongruous, ii) judging the appropriateness of the laughter that is about to be produced, and iii) judging the possible consequences of laughter. These judgements can have different degrees of automatism. In example (3a) above, to produce his laughter, John will need to judge the situation as appropriate in a certain way. A more obvious example would be laughing (or not laughing) at a sexist joke,[7] which may contain an incongruity, but may not be considered to be appropriate to laugh at. Billig (2005) uses the term *unlaughter* to describe situations of deliberately withholding laughter and shows a typical example of it as reaction of European parliamentary members to Silvio Berlusconi's failed attempts to produce humour during his address to the European Parliament.

The term *affordance* was coined by Gibson (1979) and it defines how the environment can offer a possibility for an action. For example a flower for a human and a honeybee would provide two

[7]Another example is how the 'black man's cock' joke is told (or restrained) by David Brent character in BBC's 'The Office' in different circumstances (Season 2, Episode 1).

different affordances: a human can admire it, and a honeybee can collect pollen from it.[8] Thus, the situation might create an affordance for laughter, but then this affordance can be appraised – considering the limited access to knowledge and resources that human cognition has – for appropriateness.

In this thesis we formalise the aforementioned laughter mechanisms as updates of information states of dialogue participants.

3.7 SUMMARY

In this chapter we have highlighted the importance of laughter for everyday interactions and established a background for investigating it around four different perspectives. In this thesis we explore laughter from all these perspectives.

[8]This example is borrowed from Cooper (2022) who uses it to illustrate the attunement of different organisms to the variable significance for them of the same object in the environment.

4 Dialogue

4.1 WHY DO WE NEED DIALOGUE TO STUDY LAUGHTER?

In Chapter 3 we started right off with an example of laughter in dialogue. In contrast to many approaches in neuroscience and neuropsychology we do not consider laughter in isolation, for instance, as some sort of stimulus. In this thesis we treat laughter as a dialogical phenomenon. The meaning of laughter can only be understood in relation to the situational, emotional and conversational context. In this chapter we draw readers' attention to the elements of dialogue and corresponding theories of dialogue that are used in this thesis.

Despite being disregarded by traditional linguistic theories (e.g., Chomsky, 1965, p. 58) due to being inherently related to linguistic performance, dialogue plays a primary role in human linguistic activities. It is not as tidy as written texts but it is structured and systematic (Sacks et al., 1974; Clark, 1996; Pickering and Garrod, 2004, among many others). 'Untidy' elements, such as false starts, long pauses, phrasal breaks, split utterances, repairs, disfluencies, and – for the sake of this study – laughs provide data for analysing the structure and coherence of a dialogue.

4.2 UTTERANCES, TURNS AND DIALOGUE ACTS

Our first aspect of dialogue is the unit of analysis. People typically do not utter full sentences. For instance, they often use non-sentential utterances (NSUS) (Fernández and Ginzburg, 2002), like 'yes' or 'on Tuesday' (for instance, as a reply to a question 'Will you come to meet me?'). Utterances are sometimes incomplete, and can be completed by the same or speaker or their partner. Units constructed from the utterances, also known as collaborative

constructions, might be 'ungrammatical' (in a traditional sense, see Gregoromichelaki et al., 2011) and their attribution to a single speaker might be put under question.

People typically exchange turns when they speak, but their speech and non-verbal contributions can also overlap. In backchannel responses (e.g. 'Mm' or laughter as a backchannel, see Chapters 9 and 11) it is an open question of annotation convention whether the turn is taken by the listener and then quickly returned back to the speaker to continue their speech or if a single turn is constituted from continuations by the speaker. In (4a) in accord with annotation guidelines (Jurafsky et al., 1997a) the contributions of the speakers constitute independent turns.

(4a) from Switchboard Dialogue Act Corpus (SWDA), (discussing air pollution):

1 *A.* We eat and sleep the stuff, everything we do over here.
2 *B.* <laughter>
3 *A.* and, uh, it's an interesting job.

sw2006

An utterance can play one or several functions in the context of a dialogue. Bunt (1981) introduces the term *dialogue act* which extends the notion of speech act established by Searle (1969) and Austin (1975). Speech acts assume perfect communication and an individualistic view of a speaker which is not applicable to dialogue. One utterance may correspond to several dialogue acts depending on the local context and on the domain of a dialogue. Another difference is that dialogue presumes more functions than the ones pointed to by Austin (1975). These functions are specific to dialogue, such as those used for *interaction management* (for example, to signal understanding). Dialogue acts underline the structure of dialogue and some combinations of dialogue acts more frequently co-occur, such as greeting and counter-greeting, offer and acceptance, question and answer, apology and downplayer (see Chapter 9 for the role of laughter in this pair). Moreover, one utterance can perform several dialogue acts at the same time. 'An utterance "Bill

will be there" can simultaneously function as an information act and as a promise (or a threat)' (Fernández, 2022, p. 182).

These are examples of challenges to be taken into consideration while annotating and analysing dialogue data, due to the fact that different corpora have adopted different standards of annotation. Therefore, the results reported in this thesis ought to be taken with a pinch of salt. We cannot always eliminate the possibility that certain observations are artefacts of annotation conventions, even if the inter-annotator agreement indicates good reliability.

Additionally, not only are the annotations domain-dependent, but the contexts of dialogues themselves vary hugely. There can be different genres, situations, or levels of familiarity between participants, for example. For instance, the Switchboard corpus, which has been widely used in computational linguistics (e.g., in automatic speech recognition (ASR) research), contains conversations of participants that do not know each other and have to discuss specific subjects over the telephone. This might put the generalisability of the results under scrutiny, because they might be not applicable to other domains, such as ones that involve face-to-face interactions. In Chapter 9 we discuss this issue in relation to the role that laughter plays in dialogue act recognition across two different domains.

4.3 GROUNDING

While communicating, especially over unreliable communication channels, humans give each other evidence that their contributions are understood to a certain extent, sufficient for current purposes. Clark (1996) and Allwood (1995) distinguish four *levels of action* related to different degrees of grounding. Here we list them according to the *action ladder* (Clark, 1996), from the hearer's perspective.

1. *Acceptance* level determines whether the content of the utterance was accepted or rejected by the hearer.

2. *Understanding* level specifies whether the utterance was understood by the hearer

3. *Perception* level determines whether the utterance was perceived by the hearer.

4. *Contact* level determines whether interlocutors have established a channel of communication.

According to the principle of *downward evidence*, the action ladder assumes that if the level above is complete, then all levels below are complete. For instance, if Bob asks 'Do you like Paris' and Mary replies 'Yes', then Bob's utterance is accepted (and also understood, perceived, and their contact has been established). If she asks 'Paris?' then it signals that Bob's utterance was perceived but not understood (and thus not accepted).

In a dialogue with a machine which involves uncertainty of ASR and natural language understanding (NLU) components, we cannot assume perfect communication, and grounding therefore becomes a major issue. Larsson (2002) accounts for different levels of action within the IBiS2 dialogue management framework using a set of rules to update the common ground represented in the information state of the system. He uses Interactive Communication Management (ICM) moves (Allwood, 1995) as explicit signals concerned with communicating the updates to the common ground, and sequencing moves, e.g. restarting a dialogue. In Chapter 10 we will propose our implementation of Larsson's account for ICM moves and then in Chapter 11 will show how laughter can be interpreted in relation to interaction management.

This is an example of formalisation of one of the dialogical phenomena within a theoretical and computationally tractable framework. In the next section we discuss the frameworks which we use throughout the thesis.

4.4 FORMAL ACCOUNTS FOR DIALOGUE

4.4.1 Information-state update

In this thesis we employ an information-state update (ISU) approach, following several authors, including Larsson (2002) and

Ginzburg (2012). In this view we present the information available to each participant of the dialogue (either a human or an artificial agent) in a rich information state. Being rich entails that the information state contains a hierarchy of facts, including the ones that are thought to be shared and the ones that have not been yet publicised.

The main benefit of using a rich representation of the information state with underspecified components is to be able to address a wide range of clarifications from both parties. This is especially beneficial for dialogue systems, because of the need to recover from automatic speech recognition or natural language understanding errors. Moreover, ISU can take care of topically relevant follow-up questions by the system, or contributions when the user provides more information than they were asked (over-answering, see Larsson, 2002).

4.4.2 KoS

KoS (not an acronym but loosely corresponds to Conversation Oriented Semantics, Ginzburg, 2012) provides one of the most detailed theoretical treatments of domain-general conversational relevance, especially for query responses – see the work of Purver (2006) on clarification requests, and Łupkowski and Ginzburg (2017) for a general account – which ties into the KoS treatment of NSUS which are crucial for modelling naturalistic dialogue. KoS provides one of the most detailed analyses of NSUS (Fernández et al., 2007; Ginzburg, 2012) and we follow these approaches in our implementation in Chapter 10.

Following the seminal work of Lewis (1979), KoS models dialogue as a game, containing players (interlocutors), goals and rules. KoS represents language interaction by a dynamically changing context. The meaning of an utterance is then how it changes the context. Compared to most approaches, which represent a single context for both dialogue participants (e.g. Roberts, 2012), KoS keeps separate representations for each participant. Consequently, the information-states of the participants comprise a private part and the public part called Dialogue GameBoard (DGB). The DGB

represents information which arises from publicised interactions. The DGB tracks, at the very least, the assumptions that are assumed to be shared (including the visual field), moves (which are comprised of utterances, their form and content), and questions under discussion.

KoS is and ISU approach, formalised using the Theory of Types with Records (TTR) (Cooper, 2022). It can thus leverage a wide range of work based on TTR, including the modelling of intentionality and mental attitudes (Cooper, 2005), generalised quantifiers (Cooper, 2013), co-predication and dot types in lexical innovation, frame semantics for temporal reasoning, reasoning in hypothetical contexts (Cooper, 2011), spatial reasoning (Dobnik and Cooper, 2017), enthymematic reasoning (Breitholtz, 2014; Breitholtz, 2020), clarification requests (Purver, 2006; Ginzburg, 2012), negation (Cooper and Ginzburg, 2012), non-sentential utterance resolution (Fernández et al., 2007; Ginzburg, 2012) and iconic gesture (Lücking, 2016). We discuss TTR in further detail in Chapter 15 in order to employ it for detailed analysis of humorous interactions.

4.5 METHODOLOGICAL APPROACHES TO DIALOGUE

There exist several very distinct schools with their own established methodological approaches to dialogue. Conversational Analysis (CA) relies on naturalistic data and radically ignores all other sources of data including interviews about actual behaviour, structured observations, constructed examples of dialogues and experimental methods of influencing subjects' behaviours. CA researchers, with rare exceptions, have avoided quantification of their findings. In contrast, experimental methods rely on statistical significance testing new hypotheses against a null-hypothesis, with hypotheses typically inspired by existing literature or other fields of study. In recent decades the crisis in replication of experimental results in psychology has drawn the attention of many scholars (e.g., Pashler and Harris, 2012).

More recent studies in CA, originating from Jefferson (1989) have adopted count-based measures in their research. Coding

schemes inspired by CA quantify the phenomena of interest and also allow hypotheses testing (Chapter 8 is an example using this methodology). Ideally, observational CA studies can be a source which identify a hypothesis concerned with interactive phenomena, then corpus studies can help quantifying it, and psycholinguistic experiments allow testing the hypothesis, typically, for a specific task (Ruiter and Albert, 2017). Experimental approaches which are surgical in their intervention-based methodology are also possible, including ones looking at laughter (Mills et al., 2021; Maraev, Mazzocconi, Mills et al., 2020, beyond the scope of this thesis).

Formal theories of dialogue, such as Dynamic Syntax (DS) (Kempson et al., 2001) and KoS (Ginzburg, 2012), are rooted in the idea of bringing context with the remit of the grammar (Stalnaker, 1978; Barwise and Perry, 1981, among others). These approaches are aimed at formally defining dynamic and incremental changes in the context. Ginzburg (2012) draws attention to the role of miscommunications and repair for semantic theory claiming that the adequacy of semantic theory is 'the ability to characterize for any utterance type the contextual update that emerges in the aftermath of successful exchange and the range of possible clarification requests otherwise' (ibid.).

Computational approaches to dialogue are either rooted in connectionist models or in symbolic approaches (see Chapter 5). According to Computational-Representational Understanding of Mind (CRUM) (Thagard, 2005), the currently dominant approach in cognitive science, computational procedures which operate on representational structures are a viable way of understanding the human mind. This hypothesis encompasses experimental, formal and computational approaches, and has been fruitful in different fields of study. Computational models can be informed by real data and theoretical findings. The quality of implementation of such models and their closeness to reality can also improve the quality of human-computer interaction (HCI). Conversely, CRUM additionally highlights the importance of dialogue systems as a way to evaluate linguistic theories which aim to model human cognition – modelling conversational competence computationally is a key chal-

lenge in cognitive science research (McTear, 2020). One of the ways to evaluate such competence is to assess the ability of a computer to converse in a natural and coherent way, including understanding of emotional climate and making sense of non-verbal social signals, including laughter.

In this thesis we start from the position that human laughter is a dialogical phenomenon and its investigation can be informed by all these approaches to dialogue. Corpus studies are represented by Chapters 6, 8 and a fragment of Chapter 9. Chapters 7 and 9 model dialogue phenomena using connectionist models (artificial neural networks), whereas Part IV and Chapters 10 and 11 adopt formal models of dialogue and symbolic computational approaches.

5 Spoken dialogue systems

5.1 TYPES OF DIALOGUE SYSTEMS

Throughout this thesis, the term 'dialogue systems' will be used to refer to a wide range of computer programs which are able to maintain spoken, written and multimodal dialogues with humans. There are many synonyms that are used to describe dialogue systems, each indicating a specific set of capabilities or a purpose: embodied conversational agents (ECAS), chatbots, personal digital assistants, virtual personal assistants, conversational user interfaces, and voice user interfaces.

Generally speaking, dialogue systems serve two types of purpose: they can be either task-oriented or non-task-oriented (chit-chat). Different methods have been proposed to classify dialogue systems. For the purposes of this short overview we find the classification proposed by McTear (2020) very useful. According to McTear, dialogue systems can be categorised by i) research and commercial tradition, ii) supported types of interactions, and iii) the approaches to dialogue systems' design and development.

Dialogue systems have been developed in different research and industrial communities. The main tradition that we inherit here is a set of academic approaches to dialogue systems, often voice-enabled, typically aimed to reproduce properties of human dialogue (Chapter 4). Commercial spoken dialogue systems are often known as voice user interface (VUI), they are often used to provide customer support over the telephone. Another strand of dialogue systems are text chatbots, aimed to mimic human conversation and pass Turing's Imitation Game (Turing, 1950). Chatbots have been around for several decades, but became popular with advances of end-to-end neural models (see Section 5.3) trained on

very large datasets (Roller et al., 2021, among others), although such datasets often involve not strictly dialogical sources and can't account for many dialogue phenomena (Shalyminov et al., 2017; Noble and Maraev, 2021). The applicability of end-to-end dialogue systems to voice-enabled interactions is also questionable, due to the need to recover from automatic speech recognition (ASR) errors or deal with ASR uncertainty (Wen et al., 2017).

ECAS, such as social robots and situated agents are types of dialogue systems that crucially depend on the ability to understand multimodal input and produce appropriate multimodal behaviour. Such systems often include models of emotions and annotation schemes for behavior realisation, including facial expressions (see McTear et al., 2016, Chapters 13–16). We believe that this strand of research will benefit from the findings and discussions in this thesis.

In regard to supported types of interactions there are several dimensions of classification, such as types of initiative and dialogue genres covered. Dialogue systems commonly fall into three categories with respect to supported initiative in dialogue. The first case, *user-initiated dialogues*, is currently widespread in virtual assistants and smart speakers and typically consists of one-shot requests to a system. In *system-directed dialogues* the control is given to a system: the system asks questions to collect user requirements, or give instructions, etc. The third case, in our classification (cf. McTear, 2020), would not only contain open-domain dialogues which imply *mixed-initiatve* which is the main topic of interest of end-to-end neural models, but also task-oriented cases where mixed-initiative is allowed. This entails that the user is given an opportunity to play an active role in a dialogue flow in a natural way: e.g., issuing clarification requests, signalling misunderstandings in accord with the model of grounding (see Chapters 4 and 10), asking for explanations (Chapter 16), changing topics (Larsson, 2002). In Chapter 11 we will give suggestions regarding how laughter can be treated in task-oriented mixed-initiative dialogue as a meaningful signal that interacts with the dialogue system's information state and model of grounding.

In Figure 5.1 we present a simplified architecture of a typical spoken dialogue system. Typically the following components are identified.

Automatic speech recognition (ASR) interprets input speech signal into a ranked list of N speech recognition hypotheses $u_{U_i}(1 \leqslant i \leqslant N)$ about user utterances accompanied with confidence scores c_i.

Natural language understanding (NLU) interprets ASR hypotheses into a ranked list of user moves $m_{U_j}(1 \leqslant i \leqslant N)$, in the form of dialogue acts or intents, accompanied with confidence scores c_j.

Dialogue manager (DM) is a central component of spoken dialogue systems, which maintains the information state of a dialogue and decides about the system's next move(s) m_S in response to interpreted user's move(s) m_{U_j}.

Natural language generation (NLG) converts the system's moves m_S into utterance(s) u_S, typically, strings of words.

Text-to-speech synthesis (TTS) converts systems utterances u_S into an acoustic signal.

In embodied conversational agents (ECAS), ASR and TTS components are typically extended or complemented to support multimodal behaviour realisation and recognition.

This thesis provides some insights into enabling all the components of spoken dialogue systems with laughter, but the component which is involved the most is the DM. The DM often also interacts not only with NLU and NLG components (as we show in Figure 5.1), but with other components too, providing, for instance, a dialogue state for the ASR to contextualise, thereby improving the quality of speech recognition.

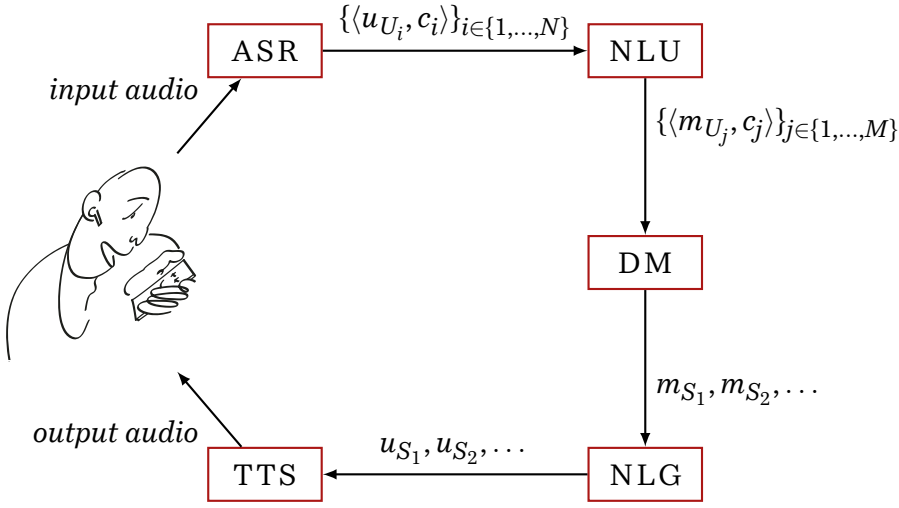


Figure 5.1: Simplified architecture of a typical spoken dialogue system

5.3 DIALOGUE MANAGEMENT

One of the most challenging tasks in the design of dialogue systems concerns their capability to support dialogue strategies that are similar to ones that occur in human dialogue. The key component of a dialogue system in this respect is the dialogue manager, which selects appropriate system actions depending on the current state and the external context.

Two families of approaches to dialogue management are typically considered: hand-crafted dialogue strategies (e.g., Larsson, 2002; Jokinen, 2009) and statistical modelling of dialogue (e.g., Rieser and Lemon, 2011; Young et al., 2010; Lison, 2015). Frameworks for hand-crafted strategies range from finite-state machines and form-filling to more complex dialogue planning and logical inference systems, such as information-state update (ISU) (Larsson, 2002) which we employ in this thesis. Statistical models help to contend with the uncertainty that arises in human interaction; from

noisy signals from speech recognition and other sensors to pragmatic ambiguities. Statistical models also allow joint training of several components. For instance, Lemon (2011) has shown that jointly optimised DM and NLG outperformed the combination of the same components optimised separately.

End-to-end neural dialogue systems that do not specify a dialogue manager as an explicit component have gained a lot of attention recently (Huang et al., 2020), and are now the central focus of research in dialogue systems (McTear, 2020). Although such systems are mostly focused on chit-chat dialogues (with exception of, for instance, Li et al., 2017; Liu and Lane, 2017; Wen et al., 2017), coherence plays a crucial role there too. Typically the main issues associated with such systems are related to memory limitations which cause repetition, self-contradiction and forgetfulness. These issues are inherently different from the issues that cause miscommunication in human dialogue, for which there are well-studied repair strategies (e.g., see Purver, 2006; Maraev, Bernardy and Ginzburg, 2020).

Although there has been a lot of development in dialogue systems in recent years, only a few approaches reflect advancements in *dialogue theory*. Our aim is to closely integrate dialogue systems with work in theoretical semantics and pragmatics of dialogue. This field has provided accounts for linguistic phenomena intrinsic to dialogue such as non-sentential utterances (Schlangen, 2003; Fernández et al., 2007; Ginzburg, 2012), clarification requests (Purver, 2006; Ginzburg, 2012) and self-repair (Ginzburg et al., 2014; Hough and Purver, 2012), where the resolution is intuitively tied to the coherence of what is being said. In Chapter 10 we develop our own DM in spirit of KoS, as discussed in Section 4.4.2.

5.4 LAUGHTER IN DIALOGUE SYSTEMS

A considerable amount of literature has been published on laughter in dialogue systems and ECAS.

Ding et al. (2014) developed one of the first approaches to laughter animation synthesis (lip, jaw, head, eyebrow, torso and shoulder

animation) for a virtual character based on the input which is represented as pseudo-phonemes of laughter. Regarding the speech signal, Aucouturier et al. (2016) created an audio platform that is able to add emotional tone to the voice of a speaker, including sadness, fear and happiness. Matsumoto et al. (2020) address laughter from a similar perspective, assuming it to be a realisation of emotions. To date, several studies have investigated laughter synthesis. Nagata and Mori (2020) trained their laughter synthesis model from a corpus of spontaneous speech and synthesised laughter using a Hidden Markov Model (HMM)-based speech synthesis model. Tits et al. (2020) proposed to use TTS based on sequence-to-sequence principles and claim that it reaches a higher degree of naturalness compared to HMMS, although they were only concerned with laughs related to amusement. Mansouri and Lachiri (2020) compares autoencoder and variational autoencoder models for laughter synthesis. Luong and Yamagishi (2021) use the abstract representation of the waveform to control laughter synthesis and a generative adversarial network as a vocoder as the basis for their system. Overall, these studies highlight that the issue of laughter synthesis in different contexts serving different functions still remains to be addressed. For instance, one unexplored question for future research is how can laughter which accompanies an apology be synthesised together with speech in a naturalistic way (see Chapter 11).

Laughter detection and classification is a more developed topic compared to synthesis. State of the art laughter detection is based on machine learning techniques: from support vector machines (SVMS), Gaussian mixture models (GMMS) (Tahon and Devillers, 2015) and automatic language independent speech processing (ALISP) (Pammi et al., 2012) to deep learning approaches, including convolutional neural networks (CNNS) (Kaushik et al., 2015). A more recent study of Gillick et al. (2021) accounts for laughter in noisy real-life environments with fine granularity. Gillick et al. (2021) also provide a comprehensive overview of laughter detection techniques proposed in recent years.

Further, laughter can be detected not only from speech signals

but from other modalities, such as facial expressions, although most of the studies are concerned with smiles, not distinguishing laughter as a separate category (e.g., Tsujita and Rekimoto, 2011; Matsumoto et al., 2020).

In regard to laughter prediction, which we explore in Chapter 7, previous work is generally only concerned with humorous laughter and either stage performances, such as TED talks (Chen and Lee, 2017), or sitcoms (Purandare and Litman, 2006; Bertero and Fung, 2016; Patro et al., 2021) (see Chapter 16 for discussion of humour in dialogue systems). Conversational laughter was recently addressed by Xu et al. (2021) who used an attention mechanism to investigate specific contexts of conversations in which laughs occur, highlighting co-occurrences of affirmative laughs (in their own classification) with positive words and embarrassment laughs with words like 'um' and pauses longer than 0.5 seconds (such pauses are claimed to indicate topic shifts). A broader human-computer interaction (HCI) perspective which is more relevant for dialogue systems was adopted by Lala et al. (2020) who used a corpus of human-robot dialogue with the ERICA robot (Glas et al., 2016) in a Wizard-of-Oz operated speed dating scenario to predict whether laughs are shared between the human and system. Another Wizard-of-Oz study, highlighting the function of laughter in the context of shame is planned by Hladky et al. (2021).

In HCI and affective computing studies laughter has also been studied in relation to child-robot interaction (Batliner et al., 2019) and in studies of user engagement (Soury and Devillers, 2014). Kantharaju et al. (2020) looked at laughter in multi-party interactions and showed that laughter occurs more frequently where group cohesion is higher. In regard to multi-party dialogue and similarly to Xu et al. (2021), Gilmartin et al. (2013) show that laughter is significantly more likely to occur in the 15 seconds before a topic change, following Conversational Analysis (CA) studies of laughter and topic closure (Jefferson, 1979; Holt, 2010). In Chapter 8 we additionally provide a review of previous work on laughter in ECAS in the context of gaze.

Overall, existing integration of smiling and laughter in ECAS is

not based on the dialogue state and is typically triggered by a joke told by a user or an agent. This means that a variety of laughter functions that exist in spontaneous human-human interaction cannot be accommodated. Up to now, far too little attention has been paid to the role of laughter in NLU, NLG and DM components.

In regard to NLU and NLG, this dissertation seeks to explain the relevance and predictability of laughter (Chapter 7), its role in relation to dialogue acts (Chapter 9) and to the direction of gaze (Chapter 8). The mutual impact of laughter and dialogue context, which is assumed to be represented by a DM module, is studied in Chapters 11 and 16.

Part II

Laughter placement

Introduction to Part II

...Когда все это случилось, то знай, что смехотворная точка найдена. После этого можешь приступать к своей юмористической программе, и, будь спокоен, успех тебе обеспечен.
...When all this has taken place, then the point at which laughter can be induced has been reached. After this you may proceed to your programme of humour and, rest assured, success is guaranteed.

Daniil Kharms, *On Laughter* (25 September 1933)

In this part we are going to explore the placement of laughter within speech and in relation to other non-verbal behaviours. The question of laughter placement is an important concern of at least the natural language understanding (NLU) and natural language generation (NLG) modules of a dialogue system. The importance of laughter placement for these modules is plain to see: a model of laughter placement helps the NLG module put laughter in the right place and makes its prediction (for NLU and automatic speech recognition, ASR) easier. The question of the inter-relation between gaze and laughter has important implications for behaviour recognition and realisation in embodied conversational agents (ECAS).

We will start with a brief background discussing what we know about laughter placement (Chapter 6), and then present two studies which look at laughter placement from different perspectives.

In Chapter 7 we will look at how laughter relevance at a certain point in a conversation can be predicted by neural network models and compare their performance with human performance on the same task.

Chapter 8 is dedicated to the placement of laughter in a wider perspective, i.e. in relation to other non-verbal signals, in this case, the direction of gaze.

6 Corpus analysis of laughter placement

6.1 LAUGHTER PLACEMENT AND RELEVANCE

A primary concern of the linguistic role of laughter is its placement. Following the terminology from conversational analysis (Glenn, 2003), we employ the term *laughable* to refer to what the laughter is pointing at, without making any claims about its possible humorous content (as discussed in Section 3.3). As is shown in the examples below, laughter can be placed after (6a), before (6b), or in overlap with the laughable (6c) (the laughable is underlined).

(6a) 1 *Lecturer.* The other announcement erm is er Dr ** has asked me to address some delinquents, no that's not fair, some er <u>hard working but misguided students</u>
2 *Audience.* <laughter>

BNC JSM

(6b) Do you have, uh, have any <laughter> <u>criminal records</u>?

DUEL Chinese 2_3

(6c) 228 *A.* I don't know if you can help that man or not.
229 *B.* I'll have <laughing: <u>a word</u>> with him Terry
230 *A.* [...] Because Mr had represented him

BNC F7X

With respect to laughter placement, Provine (1993) claims that laughter always follows the laughable and punctuates speech exclusively on phrase boundaries. Provine's claim, based on impressionistic observation of conversations, has recently been strongly

challenged by detailed corpus-based studies that demonstrate that there is a relatively free time alignment between laughter and the laughable (Tian et al., 2016; Mazzocconi et al., 2020, and Section 6.2 of this chapter).

We introduce the term *laughter relevance spaces* analogous to transition-relevant places (TRPS) (Sacks et al., 1974), backchannel relevance spaces (Heldner et al., 2013) and feedback relevance spaces (Howes and Eshghi, 2021). We define a *laughter relevance space* as a position within the interaction where an interlocutor can appropriately produce a laughter (either during their own or someone else's speech). Following the approach of Heldner et al. (2013) to backhannels, we distinguish *actual laughs* from *potential laughs*. By definition, the number of potential spaces for laughter is larger than the number of actually produced laughs. In Chapter 7 we investigate the predictability of laughter relevance spaces either by humans or machine learning methods.

Having defined what is meant by *laughter placement*, we will now move on to discuss statistics of laughter placement in the British National Corpus (BNC). The data that we discuss in this chapter was annotated as part of the study reported by Mazzocconi et al. (2020).

6.2 PLACEMENT STATISTICS

Laughter can precede (forward-looking laughter), follow (backward-looking laughter) or overlap with the laughable, with the time alignment between the laughter and laughable dependent on who produces the laughable and the form of the laughter (Tian et al., 2016).[1] The task of laughable identification is analogous to the task of finding anaphoric antecedents (also known as coreference resolution), which is considered a hard natural language processing (NLP) problem and has been the subject of many studies (e.g., Clark and Manning, 2016). Laughable identification seems to be an even

[1] In this chapter we only make a distinction between *laughter bouts* and *speech-laughs*, for definitions see Section 3.5.

Producer of laughable	Type of laughter	Position of laughter w.r.t. laughable: after #	after %	during #	during %	before #	before %	Total #	Total %
Self	bout	55	19.03	2	0.69	7	2.42	64	22.15
	speech	11	3.81	53	18.34	6	2.08	70	24.22
	total	66	22.84	55	19.03	13	4.50	134	46.37
Other	bout	118	40.83	15	5.19	3	1.04	136	47.06
	speech	13	4.50	3	1.04	0	-	16	5.54
	total	131	45.33	18	6.23	3	1.04	152	52.60
Unknown		1	0.35	2	0.69	0	-	3	1.04
Total		198	68.51	75	25.95	16	5.54	289	100.00

Table 6.1: Number of laughables of each type with % of total laughables

more challenging task, since it is not constrained by grammatical features such as gender and number agreement. There have been some preliminary investigations of laughter by Tian et al. (2016), but as yet without any systematic understanding of how laughs and their corresponding laughables are distributed in dialogue.

According to Mazzocconi (2019), in the BNC speech-laughs make up 29.7% of all laughs and the majority of these point at laughables produced by the laugher (81.5%). With laughter bouts the situation is the opposite; two thirds of these point at a laughable produced by their partner. The distributions of laughs in the BNC according to their placement are provided in Table 6.1, and Figures 6.1 and 6.2 depict these distributions graphically, for all laughs and depending on the form of laughter (whether it is a laughter bout or a speech-laugh) correspondingly.[2]

[2]The statistics are reported by Mazzocconi (2019) for their corpus study are the following: Duration of annotated dialogues; 603 minutes. Total number of laughs; 289 (5 laughs every 10 minutes). Inter-annotator agreement (Krippendorff's α); 0.65 for backward-looking and overlapping laughs, 1 for forward-looking laughs, 0.73 for whether it is a laughter bout or speech-laughter.

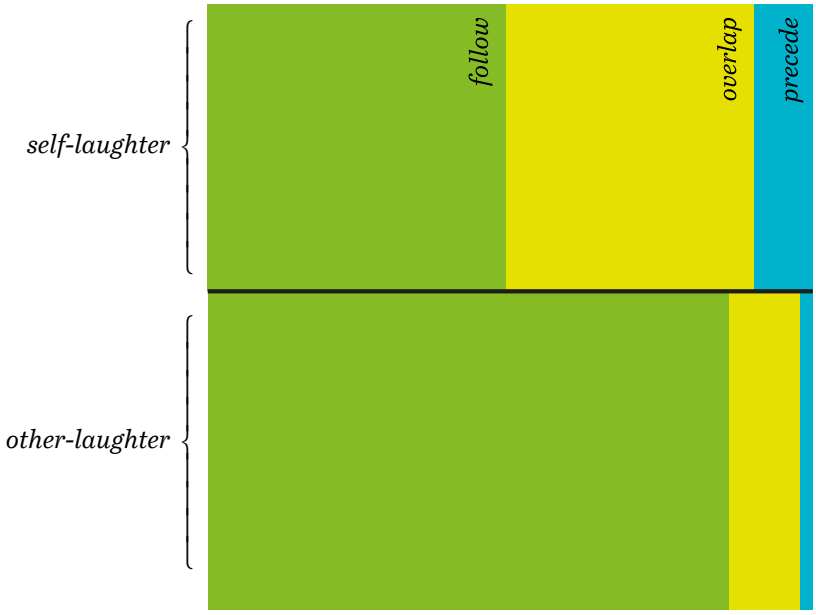
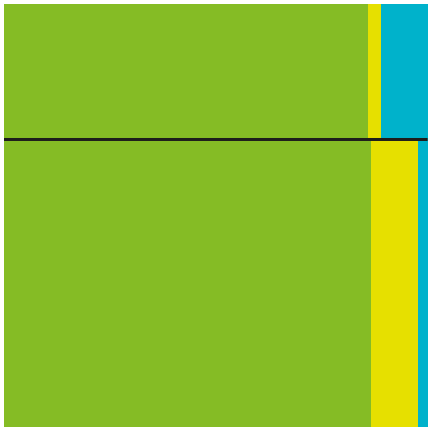
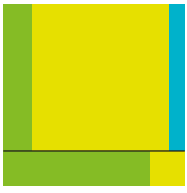


Figure 6.1: Distribution of laughs that follow (green), overlap (yellow) or precede (blue) the laughable. Top part: laughable is produced by the laugher, bottom part: laughable is produced by the partner.



(a) for laughter bouts (70.3% of all laughs)



(b) for speech-laughs (29.7% of all laughs).

Figure 6.2: Same as Figure 6.1 but depending on a laughter form.

6.3 EXAMPLES

We next present a number of examples of laughter from the BNC that vary along these parameters and cover all the cases of placement (laughables are underlined).

6.3.1 Self-laughter

Backward-looking See (4a).

Forward-looking See (6b), (3e) and (3d). In another example (6d) Michael Heseltine laughs along with many filled pauses and his laughter can be considered a sort of a filled pause with some emotional tint, related to his upcoming humorous and embarrassing contribution.

(6d) *Michael Heseltine (talking about his experience in the hospital).* Erm e e they're all individual and mine I think, looking back, was erm erm er a light er experience and I mean as I've said, the <laugh> the erm er w <u>this telescope they shove up you</u>, er I mean, Chris Patton had the same thing.

BNC K6A

Overlapping See (6c). In another example (6e) the scope of the laughable is debatable; it could be considered to be the *punchline* ('I actually made some notes') or to be the complete humorous situation. In either case, however, there is an overlap between the laughter and laughable.

(6e) *Michael Heseltine.* Well I suppose erm well I- <laughter> I- I- I mean I remember er ludicrous really, sort of thinking of what I would say at the Tory party conference while I was lying in *<laughing: that hospital in Venice. And <u>I actually made some notes</u>>*

BNC K6A

6.3.2 *Other-laughter*

In (6d) John laughs after the full unit was produced by Ian, however, the laughable comes earlier in the dialogue, but, due to overlap with Ian (marked by square brackets), John did not produce his laughter immediately after the laughable because he was still speaking. We hypothesise that the laughter is placed where it is, because that is the next available laughter relevance space.

(6d) 372 *Ian.* have to be away and we don't hear from you we think, [Oh you naughty boy, yes].
373 *John.* [You might get worried, yes.]
374 *Ian.* We'll take him off the scheme
375 *John.* <laughter>
376 *Ian.* because he hasn't replied to any of our correspondence.
377 *John.* Yes.
378 *Ian.* So er

BNC JNW

The partner's laughter in (6d) is backward-looking. The partner's laughter can also be forward-looking or overlapping, anticipating the upcoming laughable by prediction. For instance, if the listener of the joke anticipates the punchline (the laughable), they may start laughing before or in overlap with the punchline.

6.3.3 *Ambiguous placement of laughter*

In example (6e), the ambiguity lies in what the first laughter (shown in bold) relates to. It can either be an ironic appreciation of the preceding utterance 'Have a nice day!' or relate to Nicola's own upcoming remark on 'Doing the hoovering'. The precise timing of the laughter and its acoustic form can perhaps help with this kind of disambiguation. Note that the function is also ambiguous here as the laughter could be just expressing gratitude without irony.

(6e) 543 *Nicola.* This is a lovely surprise. Thank you! It's very nice nice of you!

544 *Linda.* Have a nice day!
545 *Nicola.* **<laughter>** Doing the hoovering <laughter>

BNC KDE

Given the ambiguity in identifying a laughable it is likely that different aspects participate in the process, such as pragmatic inference, cultural and world knowledge, and perhaps even the common ground which is specific to the interlocutors. This is an argument in favour of having an account in which different ways to integrate laughter are available in the given situation (see Part III).

6.4 SUMMARY

This chapter began by describing what we mean by laughter placement and laughter relevance. It went on to present some corpus-based background on the distributions of different placements of laughter and how this relates to the laughable. Despite arguing that pragmatic inference is important for identifying laughables (this is the focus of Part IV), in the following chapters we will focus on laughter coordination while being agnostic about its meaning. In next chapter we present a study of laughter predictability from dialogue transcripts which puts the emphasis on the concept of laughter relevance spaces.

7 Predictability of laughter relevance

The content of this chapter was previously published in Maraev, Howes et al. (2019).[1] *It has been substantially revised.*

7.1 THE AIM OF THIS CHAPTER

In Chapter 6 we introduced the notion of *laughter relevance spaces*, as the position in interaction where it is appropriate to laugh. This notion subsumes both *actual* laughs – the ones that were actually produced, but acknowledges other *potential* places for an appropriate laugh.

The specific objective of this chapter is to introduce the new task of predicting actual laughs in dialogue and address it with various deep learning models, namely recurrent neural networks (RNNS), convolutional neural networks (CNNS) and combinations of these. We also attempt to evaluate human performance for this task.

In this work we are guided by the following research questions:

RQ1 Can laughs be predicted from text transcrips of human-human dialogues by humans and by deep learning models?

[1] Vladislav Maraev, Christine Howes and Jean-Philippe Bernardy (2019). 'Predicting Laughter Relevance Spaces in Dialogue'. In: *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction - 10th International Workshop on Spoken Dialogue Systems, IWSDS 2019, Syracuse, Sicily, Italy, 24-26 April 2019*. Ed. by Erik Marchi, Sabato Marco Siniscalchi, Sandro Cumani, Valerio Mario Salerno and Haizhou Li. Vol. 714. Lecture Notes in Electrical Engineering. Springer, pp. 41–51. Section 7.1 was added and other sections have gone through minor edits. Maraev planned the study and carried out the experiments. The original manuscript was written by Maraev with input from Bernardy and Howes.

RQ2 How does the performance of human prediction compare to deep learning models?

In an attempt to address these questions we present:

- The dataset and the task of predicting laughter from dialogue transcripts.
- Human annotations of potential laughs from dialogue transcripts.
- Automatic methods for predicting actual laughs with deep learning models.

7.2 DATA AND THE TASK

The Switchboard Dialogue Act Corpus (SWDA) (Jurafsky et al., 1997a) consists of 1155 dyadic telephone conversations (221,616 utterances) between participants who were unfamiliar to each other. For the purposes of our study we make use of the disfluency annotations (Meteer et al., 1995) in the corpus. In regard to laughter, SWDA only contains laughter bouts (no speech-laughs) annotated as a single '<laughter>' token – such tokens make up 0.5% of all the tokens that occur in the corpus.

For our experiments we split utterances into tokens using the Python library *swda*[2] and combine consecutive laughs within a turn into a single laughter token. The laughter tokens are then removed from the text and replaced by laughter annotations. That is, the data is a sequence of tuples $\langle t_i, l_i \rangle$ such that:

- $t_i \in \mathbb{N}$ is the i-th speech (typically representing a word) or speaker token (For either A or B).
- $l_i \in \{0, 1\}$ is a laughter marker, which indicates whether laughter follows immediately after the token t_i.

[2] https://github.com/cgpotts/swda

(7a) Example of a sequence which ends with a laughter (disfluency annotation is shown in curly braces):

original	`sp_A It's, {F uh, } <laughter>`
t_i	`sp_A It's, {F uh, }`
l_i	`0  0  0  0  1`

The goal of this study is to determine whether l_i can be predicted, that is, does laughter occur after a given sequence of tokens $(t_0 \dots t_i)$.

7.2.1 *Exploratory task*

The obvious way to tackle the goal is to predict the probability of laughter for each token. To do so we split the corpus on turn boundaries, with no overlap (7b) and train an RNN model (Section 7.4.1) on 80% of the corpus (total number of samples range from 17k examples for 10 turns-per-sample split to 73k for 3 turns-per-sample split).

The prediction function for this task is the following:

$$\tilde{Y_1} : \vartheta_1 \rightarrow \mathbf{t} \rightarrow \mathbf{l} \tag{7.1}$$

where:

- $\mathbf{t}$ is a vector of tokens t_j ($\mathbf{t}^+$ is extended with laughter tokens), $1 \leqslant j \leqslant n$
- $\mathbf{l}$ is a vector of laughter relevance markers l_j, $1 \leqslant j \leqslant n$
- ϑ_1 is a parameter set for the model
- n is the length of the sequence including zero-padding from the front of the sequence.

In Table 7.1 we report the results depending on a number of turns per sample (in other words, the size of a context window) and threshold for converting the predicted probability of laughter into a binary value. We observed that adding more context leads to better

turns per sample	threshold	laughs to predict	precision	recall	F_1
3	0.50	1128	0.733	0.010	0.007
5	0.50	1116	0.786	0.010	0.005
10	0.50	1127	0.630	0.015	0.018
10	0.45	1127	0.407	0.020	0.132
10	0.40	1127	0.400	0.039	0.036
10	0.35	1127	0.255	0.060	0.049

Table 7.1: Predicted laughs depending on the number of turns per sample and the threshold. The number of laughs to predict vary due to different splits of the data.

predictions even if it leads to decreasing the size of the training data. In all our experiments we keep 80%-10%-10% training-validation-test split.

(7b) Example of a dialogue which is split into two samples of 3 turns each.

```
1 sp_A {F Oh, } I know. /
1 sp_A It's really amazing. /
1 sp_B Yeah. /
2 sp_A It's, {F uh, } <LAUGHTER> -/
2 sp_B Beautiful, beautiful machine. /
2 sp_A Absolutely, /
```

Yet, even in the case of 10 turns per sample, recall was only 1.5%. A direct attempt to increase the recall by increasing the threshold to report a laugh lowered the precision to unacceptable levels.

7.2.2 *Balanced task*

The above experiment indicates to us that this task is difficult to tackle using deep learning models. We attribute this difficulty to the corpus being unbalanced towards negative predictions, due to the sparsity of laughs. Indeed, the proportion of actual laughter

tokens is around 0.5% in the whole corpus. Additionally, it is also a hard and unrealistic task for humans because annotating every token is tedious.

We therefore, instead, create a balanced data set, including all context windows of a given span size which end with a laughter token and collecting (at random) a sample of the same size where the context window doesn't end with a laughter token. To do so, we run a sliding window through all the examples.

The size of the sliding window is fixed to a given number of tokens (not turns), in our case, 50 or 100 tokens. All the laughs (except the final one for the sequence) are represented as a special token and the final laughter is removed and represented as a positive label. The resulting training set (80% of all data) contained around 17k samples and remaining 20% were left out for validation and testing. This amended task is the focus of the rest of the chapter.

The prediction function for this task is the following:

$$\tilde{Y_2} : \vartheta_2 \rightarrow \mathbf{t}^+ \rightarrow l \tag{7.2}$$

where:

- $\mathbf{t}$ is a vector of tokens t_j ($\mathbf{t}^+$ is extended with laugther tokens), $1 \leqslant j \leqslant n$
- $l \in \{0, 1\}$ is a single laughter relevance marker
- ϑ_2 is a parameter set for each model
- n is the length of the sequence including zero-padding from the front of the sequence.

7.3 BASELINE MODELS

7.3.1 *Sentiment analysis baseline*

Even though laughter can be associated with a variety of sentiments, it is often naively associated with positive sentiment. Therefore, as a baseline, we employed the Valence Aware Dictionary and sEntiment

Reasoner (VADER) sentiment analyser (Hutto and Gilbert, 2014) to check whether its prediction of positive sentiment correlates with laughter. VADER is designed to classify sentiment along the positive or negative scale and mainly used for sentiment classification in social media. It is not specifically designed for dialogue but arguably should perform relatively well on 'noisy' texts such as those found in the Switchboard corpus (Howes et al., 2014). VADER is built in the Python NLTK library (Bird and Loper, 2004).

The sentiment analysis baseline gives rather low scores (see Table 7.3), although demonstrating higher recall (0.749) than precision (0.511).

7.3.2 Human performance baseline

In order to understand how well humans perform at this task, we conducted an experiment with non-expert annotators located in the USA through Amazon Mechanical Turk (AMT) platform. They were given a task description with the following salient points:

1. An invitation to complete the task with a notice that native level of English is required.

2. A sound sample of an arbitrary dialogue containing laughs (in order to help coders understand that laughter can occur in non-humorous conditions).

3. Three excerpts from the test set were the laughter tokens at the end were removed. Each of the excerpts has to be annotated regarding the potential to elicit laughter as:

    - very unlikely
    - not very likely
    - quite likely
    - very likely

The subset of 399 excerpts – in exactly the same format as the one that was used for model training and evaluation – was annotated by

selection principle	accuracy	precision	recall	F_1
average of 4-class annotations	0.51	0.50	0.92	0.65
average of binary annotations	0.51	0.49	0.67	0.57
agreed judgements w.r.t. the valence (in 271 cases out of 399)	0.51	0.49	0.98	0.66

Table 7.2: Human annotations as compared with the test set. Scores are computed based on the valence. For all the cases examples labelled as 'quite likely' or 'very likely' valence is positive, and the rest – as negative.

at least two annotators per sample. We computed Cohen's κ chance-adjusted inter-annotator agreement both for four-class predictions and for predictions converted into binary (judgements 'quite likely' or 'very likely' are counted as positive and 'very unlikely' or 'not very likely' – as negative). The resulting κ was very low (below chance level: $\kappa = -0.125$ for four-class predictions and $\kappa = -0.071$ for binary predictions), which indicates either that quality of AMT annotations are very low or that human judgements about laughter are very subjective.

Subjects also showed a disposition towards laughter: 66% of excerpts were annotated as 'quite likely' or 'very likely', and only 2% were annotated as 'very unlikely' or 'not very likely' by both annotators. After comparison with the distribution of actual laughs in the corpus we observe that AMT respondents are not very good in predicting whether there was actually a laughter at the end of the sequence, but they might instead be predicting *potential* laughter, which is suggested by the predominance of such predictions. These results are similar to sentiment analysis baseline, but for humans they are even more pronounced. In Table 7.2 we show accuracy and F_1score for human predictions of actual laughs.

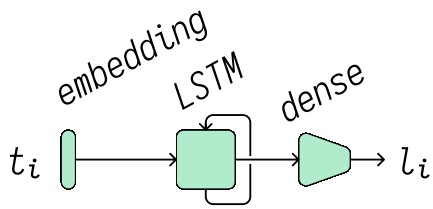


Figure 7.1: Architecture of the RNN model ('rolled' view). For the main task only the final prediction (l_n) is considered.

7.4 DEEP LEARNING MODELS

We present several deep learning models: RNN, CNN and combinations of them.

These models are implemented using our own high-level deep-learning library, which uses TensorFlow as a backend.[3]

7.4.1 *Recurrent neural network*

Our RNN-based model architecture is shown in Figure 7.1 and consists of three layers:

1. *An embedding layer* which is characterised by the size of token embeddings (d).

2. *A long short-term memory (LSTM) recurrent layer* (Hochreiter and Schmidhuber, 1997) characterised by state size n. Each LSTM cell additionally includes dropout (on its inputs, outputs and hidden state inputs) of a probability ε.

3. *A dense layer* which predicts laughter relevance for each token. We have exactly two classes: relevant (1) or irrelevant (0). For the main task we only output the final prediction of the dense layer.

[3]TypedFlow: https://github.com/GU-CLASP/TypedFlow

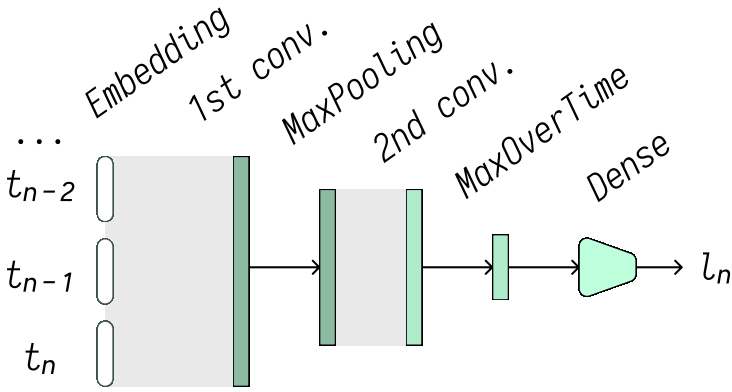


Figure 7.2: Architecture of the CNN model.

7.4.2 *CNN model*

The convolution neural network model includes the following parts:

1. *An embedding layer* which is characterised by size of token embeddings (d).

2. *A first 1D convolution layer* characterised by filter size h_1 and number of filters k_1. The layer is followed by a rectified linear unit (RELU).

3. *A first max-pooling layer* with a stride $s = 2$.

4. *A second 1D convolution layer* characterised by filter size h_2 and number of filters k_2. The layer is followed by RELU.

5. *A max-over-time pooling layer* which computes element-wise maximum along all the features of the second convolution layer.

6. *A dense layer* that predicts laughter relevance for the sequence.

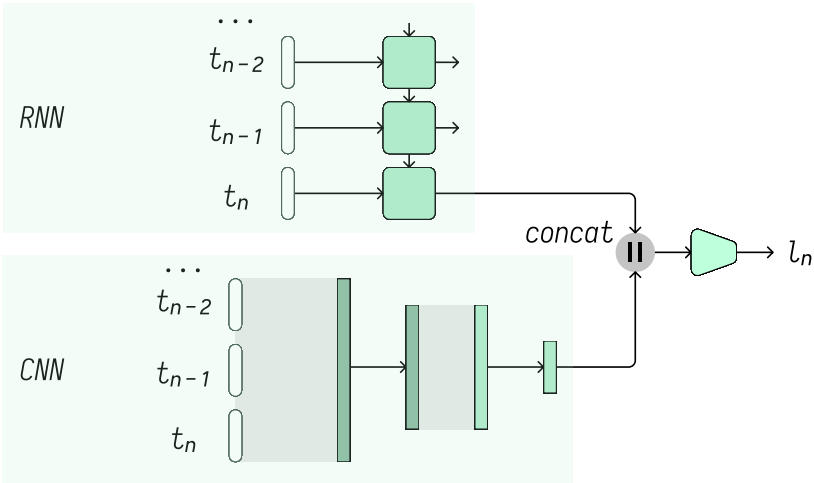


Figure 7.3: Architecture of the fusion model. Outputs of the RNN's last cell and CNN's max-over-time pooling layers are concatenated and then dense layer is applied.

Combinations of RNN and CNN

In order to estimate whether RNN and CNN models pick up on either the same or different features, we also tested two combinations of the above models:

A fusion model (Figure 7.3) where outputs of an RNN and a CNN model (both without a dense layer) are concatenated, and a dense layer operates on this concatenation.

A hybrid model (Figure 7.4) similar to the fusion model, but when token embeddings are shared between RNN and CNN.

7.5 RESULTS

We present results for the different models trained and tested on the balanced dataset in Table 7.3. Given the results of the AMT experiment, we posited that the task of predicting actual laughters

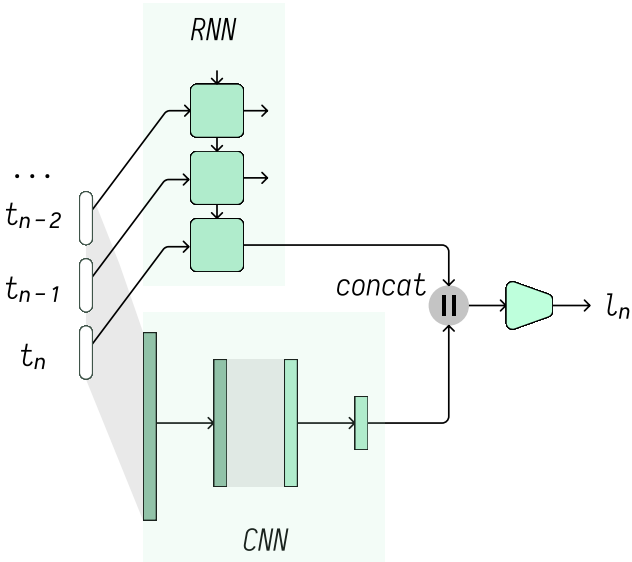


Figure 7.4: Architecture of the hybrid model. Token embeddings are shared between RNN and CNN.

in dialogue is hard for untrained humans to perform. We also saw that the task is difficult to tackle by simple sentiment analysis. Thus we expect the task to be difficult for deep learning models as well. However, the deep learning models perform considerably better than our baselines, especially in terms of accuracy. Additionally, the CNN model consistently outperforms the RNN model. Further, combining RNN with CNN provides no significant benefit. This suggests that the RNN model does not detect more laughter-inducing patterns than the CNN model.

7.6 ERROR ANALYSIS

After analysing the results, we noted that there were a large number of examples where laughter occurs at a transition-relevant place (TRP). In this case the last token of the sample is a next utterance token 'sp_A' or 'sp_B'. A concern was that this would significantly affect the results. In order to measure this effect, we removed these

model	val. acc.	test acc.	test prec.	test recall	test F_1
majority class	–	0.500	0.500	0.500	0.500
AMT	–	0.510	0.500	0.920	0.650
VADER	–	0.518	0.511	0.749	0.607
RNN (span=50)[a]	0.762	0.743	0.732	0.763	0.747
RNN (span=100)[a]	0.756	0.770	0.761	0.777	0.769
CNN (span=50)[b]	0.789	0.765	0.761	0.771	0.766
CNN (span=100)[b]	0.783	0.787	0.777	0.794	0.785
fusion (span=50)[c]	0.794	0.766	0.760	0.778	0.768
hybrid (span=50)[d]	0.793	0.776	0.775	0.774	0.774

Hyperparameters:
[a] $d = 50;\ n = 40;\ \varepsilon = 0.1$
[b] $d = 100;\ k_{1,2} = 40;\ h_{1,2} = 7$
[c] see RNN and CNN
[d] see RNN and CNN, shared embedding layer: $d = 100$

Table 7.3: Summary of the prediction results.

results from the test set and observed the accuracy and F_1-score shown in Table 7.4. We observe a drop of F_1-score (around 6 percentage points) but accuracy is almost unchanged. This indicates that the model relies on utterance boundaries (possibly combined with other features – and consequently captured by neural networks) as an important predictor for laughter. Examples where laughter is predicted to occur immediately after a next utterance token are shown in (7c) and (7d).[4]

(7c) 573 *A.* let me ask you this.
574 *A.* How, how old are you?

[4] Disfluency annotations are removed for the sake of readibility.

model	accuracy	precision	recall	F_1
RNN	0.743	0.732	0.762	0.747
RNN*	0.738	0.673	0.705	0.689
CNN	0.765	0.761	0.771	0.766
CNN*	0.761	0.715	0.694	0.705

Table 7.4: Performance of the models before and after removing the examples where turn change token is the last token (marked with a *). As a result, the dataset is 22% smaller and it is missing 36% of positive examples. All deep learning models use the dataset with the span of 50 tokens.

575 *B.* I'm, uh, thirty-three.
576 *A.* Thirty-three?
577 *B.* Thirty-two,
578 *B.* excuse me.
579 *A.* Okay.
580 *B.* ((correct prediction: <laughter>))

sw3821

(7d) 104 *B.* when I was a freshman in college
105 *A.* Uh-huh.
106 *B.* uh, my degree was in computer, uh, technology originally
107 *B.* and it seemed like it would,
108 *B.* ((wrong prediction: <laughter>))

sw2835

In conversational analysis studies many laughs are considered to form adjacency pairs with prior laughs (Jefferson et al., 1987), and preceding laughter by another speaker seems to be a relevant feature for our models (7e). However, in excerpts where there are a lot of laughs the model sometimes gets it wrong (7f) which doesn't necessarily mean that this is a place where the laughter was irrelevant.

(7e) 414 *A.* I'm not really sure what the <laughter>

415 *B.* Yeah,
416 *B.* really,
417 *B.* it's one of those things that you read once,
418 *B.* and then, if you're not worried about it, you just forget about it <laughter>
419 *A.* ((correct prediction: <laughter>))

sw4048

(7f) 492 *A.* (…) don't get a hot tub and
493 *B.* <laughter> Yes.
494 *A.* shave my legs, I'm going to die <laughter>
495 *A.* And I had <laughter>
496 *B.* Yes
497 *B.* I understand that <laughter>
498 *A.* I got enough of it right ((wrong prediction: <laughter>))

sw3293

7.7 DISCUSSION

The main conclusion of our experiments is that for the given task of predicting *actual laughs* deep learning approaches perform significantly better than untrained humans. This suggests that participants are judging whether a laugh could appropriately be produced at that point in the dialogue, not whether the dialogue participants themselves actually did produce one. We conjecture that this result extends to the general population, if asked to make laughter judgements in the same conditions (i.e. when little effort is spent for each judgement). The expert prediction of laughter, that is by subjects trained in the relevant psycholinguistic aspects, is beyond the scope of the present thesis.

We observe certain parallels between our study and studies of turn-taking in human-computer interaction (HCI). Just like the TRPS, laughter relevance spaces cannot be directly observed in data (Skantze, 2021), and there is no clear evidence that false predictions

of turn-taking prediction models indicate potential places for a turn-change.

There is a certain continuity between TRPS and feedback-relevance spaces (Howes and Eshghi, 2021) (subsuming both backchannels and clarification requests), because feedback is one possible variant of a move which takes place after the turn is taken. Laughter is another option (which can also be a type of feedback, as we will discuss in Chapters 9 and 11). However, the models that we present here predict laughs that do not necessarily occur at TRPS which, as discussed in Section 7.6 is often the case. Therefore, one further investigation would be to consider models that predict turn-taking (e.g., Ekstedt and Skantze, 2020) and use them as the basis for predicting laughter at turn changes.

The big difference between precision and recall (low precision and high recall) in the case of non-expert annotators and sentiment analysis baselines might suggest that these baselines attempt to predict potential laughs, whereas deep learning models rather predict actual laughs (it is especially evident in the case of our exploratory model being very cautious in laughter prediction). The question of whether false predictions of the models correspond with potential laughs is a matter of empirical study in the same vein as intervention-based studies of laughter (Mills et al., 2021; Maraev, Mazzocconi, Mills et al., 2020) which used only very simple heuristics to induce additional laughs in real time and investigated whether this disturbs the flow of conversation and influence task completion. Embodied conversational agent (ECA) studies analogous to Poppe et al. (2011), who systematically vary the time, placement and quantity of backchannels, might be practical.

We are optimistic that the introduced task and the approaches that we have developed constitute a big step towards inferring appropriate spaces for laughter from textual data. Together with approaches based on audio components (e.g., El Haddad et al., 2017) this should enable dialogue systems to understand when is it appropriate to laugh. Nevertheless, we are aware of the fact that this requires understanding laughter on a deeper level, including its various semantic roles and pragmatic functions.

It will be useful to extend our Amazon Mechanical Turk experiments by introducing more annotators in order to get more reliable results. Probabilistic annotations in our future crowdsourced experiments following Passonneau and Carpenter (2014) may yield results that are more consistent with the concept of laughter relevance spaces.

Regarding the task itself, it might be useful to address it in a more 'dialogical' way. Training data can be considered not as one input to a neural network which contains speaker tokens but as two possibly overlapping streams. This would introduce the notion of coordination between speakers into the prediction model. Two streams can be also extended by additional information provided in separate inputs, such as information about disfluencies, discourse markers, along with fundamental frequency and other acoustic features. It is compelling to see which features will make a more robust contribution to the task of predicting relevant laughs. In the case of TRPS, the contribution of linguistic information is not as big when combined with prosody (Skantze, 2017), therefore it would be interesting to see how different would it be for laughter. Additionally, as we will show in Chapter 9, there is a clear connection between laughter and adjacent dialogue acts in terms of co-location and functions performed by laughter. Therefore it can be hypothesised that dialogue act information can be a valuable cue for laughter relevance spaces prediction.

Another interesting perspective on laughter relevance spaces are the cues which elicit laughter. A laughable can be such a cue in some cases, but not necessarily, given that there are forward-looking laughs, as discussed in Chapter 6. One of the widely discussed cues for yielding a turn or a backchannel response is gaze (Kendon, 1967; Duncan and Fiske, 1979; Bavelas et al., 2002). The role of gaze as a laughter-yielding cue is one the main focal points of Chapter 8.

8 Multimodal placement of laughter

The content of this chapter was previously published in Mazzocconi et al. (2021).[1] *It has been substantially revised.*

8.1 INTRODUCTION

It is clear that both gaze and laughter are crucial elements to be taken into account when implementing algorithms for embodied conversational agents (ECAS) (Ochs and Pelachaud, 2013; Becker-Asano and Ishiguro, 2009), both for what concerns the interpretation of the users' dialogue acts and for what concerns their own behaviour, in order to make ECAS more competent from a pragmatic perspective and also more human-like in terms of emotional displays, where this is desirable.

Our conversations are highly coordinated, with synchronisation occurring even across modalities (Fusaroli and Tylén, 2012; Dale et al., 2013). Both laughter and gaze have been the object of in-depth independent analyses, and their crucial role in managing and coordinating interaction is not in doubt. Both gaze and laughter

[1] Chiara Mazzocconi, Vladislav Maraev, Vidya Somashekarappa and Christine Howes (2021). 'Looking for Laughs: Gaze Interaction with Laughter Pragmatics and Coordination'. In: *Proceedings of the 2021 International Conference on Multimodal Interaction*. ICMI '21. Montréal, QC, Canada: Association for Computing Machinery, pp. 636–644. Maraev and Mazzocconi planned the study and annotated the laughs. Somashekarappa annotated the gaze. Mazzocconi carried out the statistical analysis. Maraev put together the visualisations. Analysis, writing and editing of the paper were carried out in collaboration between Mazzocconi and Maraev with input from Howes and Somashekarappa.

are perceivable actions (termed visible/audible acts of meaning in Bavelas et al., 2002) which affect the unfolding of the upcoming dialogue (Mazzocconi et al., 2020). While there is some work on the interaction of smiles, laughter and gaze in relation to humour (Gironzetti, 2017; Brône, 2020), less is known about the relation of laughter and gaze when this is not related to humour, but rather to what we call *social incongruity* (as discussed in Section 3.4). The only exception we are aware of is Romaniuk (2009), who take a Conversational Analysis (CA) approach on the use of gaze to decline a laughter.

An example of the fine coordination between laughter and gaze is presented in (8a), where we see the onset of gaze at the partner from A shortly before the onset of A's laughter. The onset of A's laughter is then shortly followed by B gazing at A, just before joining B's laugh with her own.

(8a) Gaze at partner is marked with 'x', gaze anywhere else with '-'. Participant A gives a too general comment about hummus:

```
----------------------------------------xxxxxxxxx---------
A: It's "like slightly"?..   I like hummus <laughter>

B:                        yeah ((shrugs))   yeah<laughter>
-----------------------------------------------xxxxxx------
```

GHI corpus, Pair03 (00:02:17)

We aim at filling this gap by investigating the following, to our knowledge, as yet unexplored questions. The answers are to provide insights into how meaning is constructed in interaction across modalities, as well as provide empirical data for the implementation of human-like ECAS:

RQ1 Is the laugher's gaze different in terms of probability and timing, depending on the pragmatic function performed by the laughter?

RQ2 Is the gaze at the laugher influenced by the type of laughter produced by their dialogue partner, in terms of probability and timing?

RQ3 Does gaze play a significant role in laughter coordination and alignment between participants?

8.2 GAZE IN INTERACTION

The role of gaze in maintaining the conversational flow and coordinating dialogue acts is not in doubt.

While many works have argued for the importance of individual gaze for the fine regulation of turn-taking (Duncan, 1972; Goodwin, 1980), some scholars actually highlighted a lack of systematic relationship between gaze and turn-taking (Beattie, 1978; Torres et al., 1997; De Ruiter, 2005), proposing rather that gaze might function to elicit a response (Goodwin and Goodwin, 1986; Bavelas et al., 2002), which is not necessarily a speech turn (Rossano, 2013). More specifically, it has been argued that turn-taking is only a partial explanation for gaze behaviour in conversation. Our study of gaze therefore has to take into account both turn-taking and the informational structure of dialogue (Torres et al., 1997; Bonin et al., 2012).

Despite turn-taking not being the only function performed by gaze, and the fact that not all turn shifts are accompanied by gaze towards the listener, it has been consistently observed that there is a tendency for listeners to display more gaze at the speaker during the course of dyadic interaction, while the speaker tends to direct their gaze at the listener mainly towards the end of their speaking turn (Kendon, 1967; Duncan and Fiske, 1979). In this way, when a speaker gazes at the listener mutual gaze is attained (Goodwin, 1980), a brief mutual *gaze-window* (Bavelas et al., 2002) is established, and a change of floor may occur, where the previous listener looks away as they begin their speaking turn (Somashekarappa et al., 2020; Kendon, 1967; Rossano, 2013).

Gaze patterns to the interlocutor have also been found to differ depending on the speech act they accompany and on their pragmatic function (Sandgren et al., 2012; Mirenda et al., 1984; Rossano et al., 2009).

Of interest in the study of gaze in interaction is not only gaze directed at one's partner, but also its absence or avoidance (Rossano, 2013). For example, using CA, Romaniuk (2009) observed how gaze aversion can be used to decline laughter and terminate its relevance; while Kendrick and Holler (2017) report that most preferred responses are produced with gaze toward the questioner, whilst most dispreferred responses are produced with gaze averted. Moreover, it has been proposed that gaze aversion could also be explained (or influenced) by social stress (Stanley and Martin, 1968), with evidence from patients with social disorders (Schneier et al., 2011). Conversely, results from other studies (Doherty-Sneddon and Phelps, 2005) suggest that cognitive load has the most impact on gaze aversion (Glenberg et al., 1998). The latter hypothesis is based on the fact that visual cues are an important source of information and facilitate conversation, but cause higher cognitive load. This explanation seems to be supported by results observing more gaze aversion in the initial phase of request formulations (Sandgren et al., 2012; Kendon, 1967), and by speakers showing less fluency when forced to constantly look at their listener (Beattie, 1981), even though these results could also be explained by the social stress factor.

8.3 GAZE AND LAUGHTER IN ECAS

Recently, there has been a growing research interest both on gaze and other non-verbal behaviours, especially in the Affective Computing community, for the implementation of ECAS which are more competent from a pragmatic perspective and able to process and produce appropriate emotional responses (Stevens et al., 2016; Bailly et al., 2010; Lee and Marsella, 2006; Niewiadomski et al., 2009). Virtual agents benefit from a detailed analysis of multimodal input and output patterns observed during human-human inter-

actions. Bailly et al. (2010) established a basis for a context-aware eye-gaze generator for an ECA. In order to develop an improved gaze generator we should isolate the significant events detected in the multimodal scene that impact the closed-loop control of gaze. Lee and Marsella (2006) discuss the interpersonal role of gaze in interaction to signal feedback and direct conversation flow which current ECAS still lack. Simultaneously, in a dynamic environment, even the state-of-the-art ECAS struggle to direct gaze attention to peripheral movements. An embodied conversational agent should therefore employ social gaze not only for interpersonal interaction but also to possess human attention attributes so that its eyes and facial expression portray and convey appropriate distraction and engagement behaviours.

Non-verbal behaviours also can help create a stronger relationship between the ECA and user as well as allow applications to have richer, more expressive characters. Overall, appropriate non-verbal behaviours should provide users with a more immersive experience while interacting with ECAS, whether they are characters in video games, intelligent tutoring systems, or customer service applications. The current state-of-the-art of laughter in ECAS was discussed in Section 5.4.

Becker-Asano and Ishiguro (2009) evaluated the role of laughter in the perception of social robots and indicated that the situational context, determined by linguistic and non-verbal cues (such as gaze) played an important role. In particular, in their experiments, the human-like robot's direct gaze at the participant while laughing led to the perception of the robot's laughter as 'laughing at someone' rather than 'laughing with someone'.

Our aim is to provide empirical data useful for the implementation of systems able to engage in multimodal interaction, profiting from the availability of cross-modal cues, such as gaze and laughter.

8.4 HYPOTHESES

Based on the literature reviewed above, our predictions in relation to the three main questions motivating our work are the following:

Hypothesis 1 Based on the social stress hypothesis of gaze aversion (Stanley and Martin, 1968; Schneier et al., 2011), and on research showing that gaze aversion is more likely when subjects are offering a dispreferred answer (Kendrick and Holler, 2017), we expect a laugher's gaze to be less likely to be directed at their partner if the laughter is related to a social incongruity rather than to a pleasant incongruity (both around the onset and offset of the laugh).

Hypothesis 2 On the basis of studies indicating that laughter can function as an attention getting device (Stevenson et al., 1986; Reddy et al., 2002; Pinheiro et al., 2017) we hypothesise that after laughter is produced interlocutors will direct their gaze at the laugher.

Hypothesis 3 Given the role of gaze in eliciting a response (Rossano, 2013; Bavelas et al., 2002), we expect laughs where one participant joins in with another's laugh (*joining in* laughs) to be preceded by an 'inviting' gaze from their partner, as in (8a). Similarly we expect the person joining the laugh to gaze at the partner, in order to instantiate the 'gaze window' which may enable a turn shift (Bavelas et al., 2002).

8.5 METHODS AND MATERIALS

Our data consist of 23 minutes taken from three female-female dyadic interactions from the Good Housekeeping Institute (GHI) corpus (Lavia et al., 2018).

The GHI corpus contains video and audio of pairs of participants discussing and rating different kinds of hummus on a paper questionnaire (see Figure 8.1). We annotated the interactions for laughter and gaze as described in the following sections.

8.5.1 Laughter annotation

Our annotations have been conducted using ELAN (Brugman and Russel, 2004). Analogously to Mazzocconi et al. (2020) we annotated the form of laughter, its temporal sequence in relation to speech

Figure 8.1: Data collection setting (Somashekarappa et al., 2020)

		Pair 03	Pair 07	Pair 15	Total
incongruity type	Pleasant	27	11	2	40
	Social	19	10	5	34
	Pragmatic	1	3	0	4
	Friendly	5	1	0	6
laughter coordination	Isolated	28	15	5	48
	Antiphonal	20	4	2	26
	Coactive	4	6	0	10
	Total	104	50	14	168
	Minutes	10	10	3	23

Table 8.1: Distribution of different laughter annotations across dyads and minutes of interaction analysed.

and other laughs, context of occurrence, laughable (as discussed in Section 3.3), and pragmatic function. In the current study we focus on two of these features:

incongruity type – the type of incongruity present in the laughable,

laughter coordination – the coordination of laughter in relation to laughter of the other participant in the dialogue.

We assessed the agreement on laughter identification and segmentation (start-time and end-time boundaries) using the Staccato algorithm implemented in ELAN (Lücking et al., 2011), having two

annotators marking 70% of the data. We run the analysis with 1000 Monte Carlo Simulations, a granularity for annotation length of 10, and $\alpha = 0.05$. The degree of organisation is 0.8386.

Following Mazzocconi et al. (2020) we consider laughter as an event predicate, the meaning of which is constituted by two dimensions: the laughable and arousal, which we do not consider in the current work. Different kinds of laughable can be distinguished based on whether they contain an incongruity or not, and if so, which kind of incongruity (see Ginzburg et al., 2020 for a formal definition of incongruity and Section 3.3 for definitions of incongruity types). The annotation categories are as follows:

- Pleasant incongruity
- Social incongruity
- Pragmatic incongruity
- Friendliness

In the current work we focus on the observation of gaze patterns accompanying laughs related to *pleasant incongruity* compared to *social incongruity*, as these are the most frequent kinds of laughable across contexts of interaction and languages (Mazzocconi et al., 2020) (see also Table 8.1). These categories are also the furthest apart in terms of pragmatic function, since *pleasant incongruities* are related to something pleasant and rewarding, whilst *social incongruities* are related to potential discomfort and unpleasantness.[2]

In our annotation we distinguish 3 classes pertinent to the sequential distribution of the laughter in relation to laughs produced by the partner:[3]

[2]Our dataset is therefore constituted of 74 laughs: 40 related to *pleasant* and 34 to *social incongruity*. 60% of the data (50 laughs) were annotated by two of the authors. The inter-annotator percentage agreement was 82%, with Krippendorff's $\alpha = 0.69$.

[3]Inter-annotator agreement for this variable conducted over 60% of the data (50 laughs) reached 85.7%; Krippendorff's $\alpha = 0.76$.

Isolated laughter a laugh not preceded by or co-occurring with another laughter;

Antiphonal laughter a laugh shortly following a laugh from the partner, starting during the partner's laugh, or within one second after its offset;

Coactive laughter a laugh with the same onset time as a laughter from the interlocutor. We did not give an exact time definition for shared laughter onset, rather we relied on annotators' intuitions. We tested whether this intuitive notion was appropriate by calculating inter annotator agreement, which was high. Laughs which were considered to be coactive had a relative onset time of less than 100ms.

8.5.2 Gaze annotation

Following Somashekarappa et al. (2020), the gaze annotation was coded for four aspects:

Participant1 and Participant2 (P1, P2) The gaze of each participant to an object, for example, Hummus (H), Questionnaire (Q), Breadstick (B) etc.;

Joint attention (JA) looking at the same object, obtained by temporal and object overlap in P1 and P2;

Gaze1 and Gaze2 (G1, G2) for each participant, these encoded whether they were looking at their partner;

Mutual Attention (MA) looking at each other, obtained by temporal overlap in G1 and G2.

In the current work we explore only gaze at each other (G1, G2), leaving a more fine-grained analysis of gaze reciprocity (MA), and questions about gaze to objects including joint attention (JA), for future work.

8.5.3 Data extraction

In order to perform our analysis we made use of the ELAN Analysis Companion (EAC) software (Andersson and Sandgren, 2016) to conduct event-related analysis. Our dependent variables is *Gaze at partner* (G1 and G2), both from the laugher and from their partner. To address questions RQ1 and RQ2 we used *incongruity type* as a predictor to determine whether a laughter was related to a pleasant or social incongruity); while in order to address RQ3 the predictor is *laughter coordination* (isolated, antiphonal, or coactive).

Following Andersson and Sandgren (2016) we considered a time window of 3000 milliseconds (i.e. 1500 seconds before and after the laughter onset/offset). We selected 10ms resolution, using a 'first come first served' overlap handling and binned the data at intervals of 100ms, rounding up any fractions to 1.

Given the type of gaze G (laugher's or their partner's) and the *incongruity type* X (or the *laughter coordination* for RQ3), the probability of gaze before or after the event for a given time window ('bin' b) is calculated as follows:

$$P_b(G|X) = \frac{\sum_{i=1}^{N} P(g_b|x_i)}{N} \tag{8.1}$$

where N is the total number of laugh events, and $P(g_b|x_i) \in \{0, 1\}$ is the probability of gaze for a single bin b for a given predictor x_i. Figure 8.2 shows the example of calculation in a simplified case.

For each of our models, reported below, we ran a mixed-effect logistic regression in R, using the *glmer* function from the *lme4* package (Bates et al., 2015), with subjects as a random factor.[4] The dependent variable, *gaze* was treated as a dichotomous dependent variable (present or not present) for each 100ms bin of the time window of interest (3000ms centered around the value to one.]

Together with *incongruity type* (RQ1 and RQ2), and *laughter coordination* (RQ3), we considered the binary variable *time* as a predictor, contrasting the time-window preceding the laugh onset

[4] Including dyads as a random effect did not improve the models.

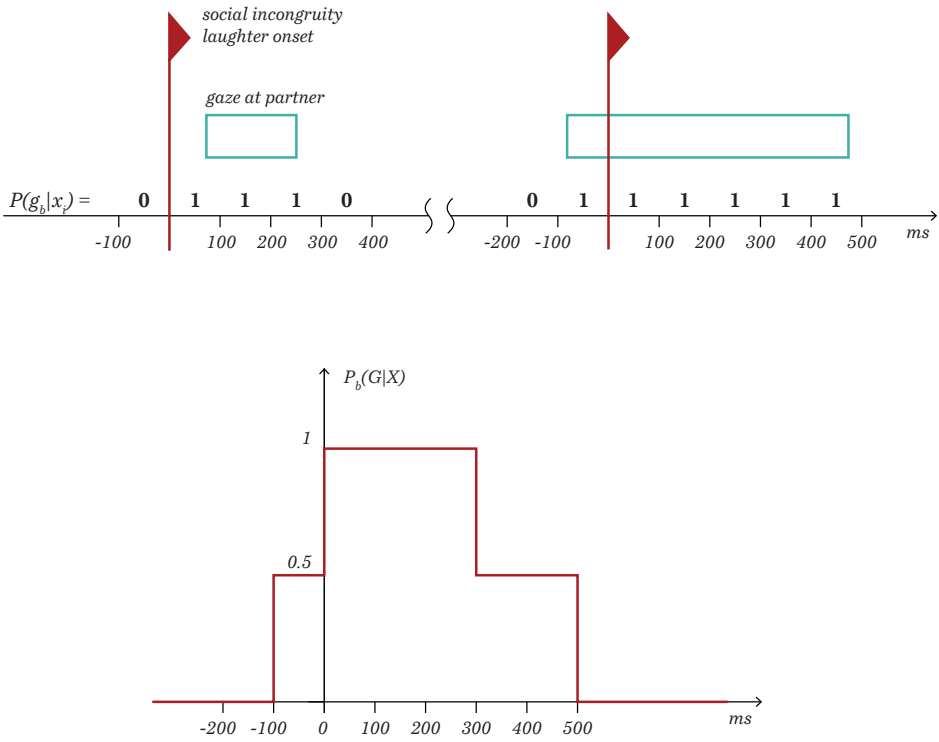


Figure 8.2: Example of calculating $P_b(G|X)$, where X is *incongruity type*. In this example only one conversation is considered (shown above). It has only two laughter events associated with social incongruity and two accompanying gazes at partner. Below the resulting graph of the probability of gaze is shown; it is a simplified version of the figures that will follow in this chapter.

or offset (1500ms, *before*) to the time-window following it (1500ms, *after*).

8.6 RESULTS

8.6.1 *Depending on* incongruity type

At laughter onset

Figure 8.3 (solid line) shows the probability of **the laugher** to direct their gaze at the partner depending on whether the laughter is related to a pleasant or a social incongruity. We observe a contrasting

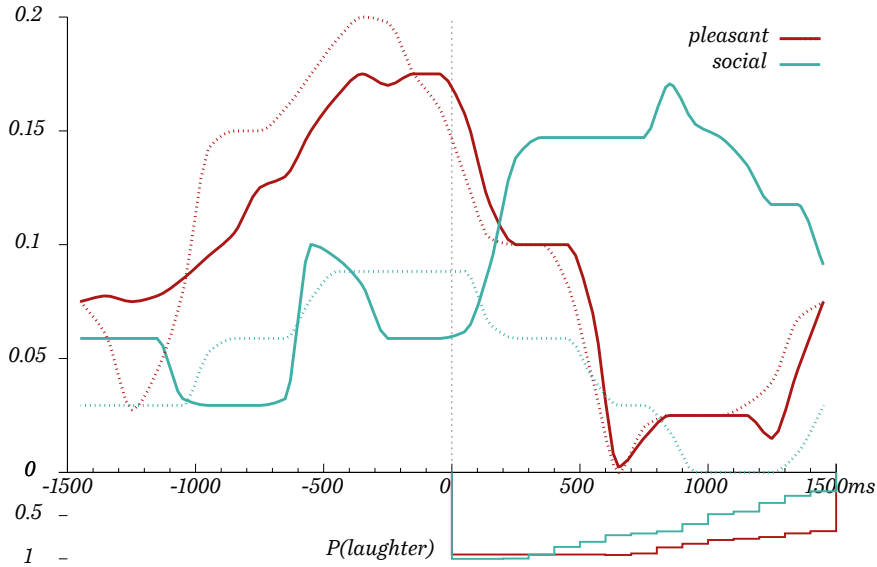


Figure 8.3: Probability of gaze at the onset of laughter depending on incongruity type. Solid line – laugher. Dashed line – partner. The probability of laughter duration is shown at the bottom of the figure.

pattern of gaze, especially after the onset of laughter: the laugher is more likely to look at their partner in the case of social incongruity.

We observe main effects of *incongruity type* ($CE = -0.81, SE = 0.22, z = -3.56, p < .001$) and *time* in relation to the onset of the laughter ($CE = -0.70, SE = 0.20, z = -3.51, p < .001$) and a significant interaction between the two factors ($CE = 1.60, SE = 0.30, z = 5.31, p < .001$).

The laugher is more likely to gaze at their partner before the onset of the laughter when it is related to a pleasant incongruity, as in (8b). After the onset of the laughter the opposite is true: the laugher is more likely to gaze at their partner when the laughter is related to a social incongruity, as in (8c).

(8b) Pleasant incongruity (by participant A):

```
------------------------------------------------------------>
A: I quite like it 'cause I like it when there's the little
```

```
B:
------------------------------------------------------------>

->-----------xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx----------------
A: chickpeas on top   'cause it's quite posh <laughter>

B:                 yeah                        yeah <laughter>
->------------------xxxx-------------------------------------
```

GHI corpus, Pair15 (00:01:25)

(8c) Social incongruity (by participant A):

```
----------------------------xxxxxxxxxxxxxxxxxxx----
A: Shall we say... No.  Ta- tasty <laughs>

B:                                            <laughs>
--------------------------------------------------
```

GHI corpus, Pair03 (00:02:59)

Figure 8.3 (dashed line) shows the probability of **partner's gaze** at the onset of the laugh depending on *incongruity type*. We observe a significant main effect of *time* ($CE = -0.83, SE = 0.2, z = -4.13, p < .001$) and *incongruity type* ($CE = -0.99, SE = 0.22, z = -4.43, p < .001$), and no significant interaction ($CE = 0.31, SE = 0.35, z = 0.87, p = 0.38$). The partner is more likely to direct their gaze at the laugher in the case of a pleasant incongruity and also more likely to gaze at the laugher prior to the laughter onset.

At laughter offset

Figure 8.4 (solid line) shows the probability of **the laugher's** gaze around the laughter offset depending on *incongruity type*. We observe a significant main effect of *time* ($CE = -0.45, SE = 0.09, z = -5.09, p < .001$), but no significant main effect of *incongruity type* ($CE = 0.21, SE = 0.17, z = 1.18, p = 0.23$), nor a significant interaction. This suggests that regardless of the *incongruity type*, the laugher is more likely to look at their partner before the laughter offset (i.e. before the laughter ends) rather than after it.

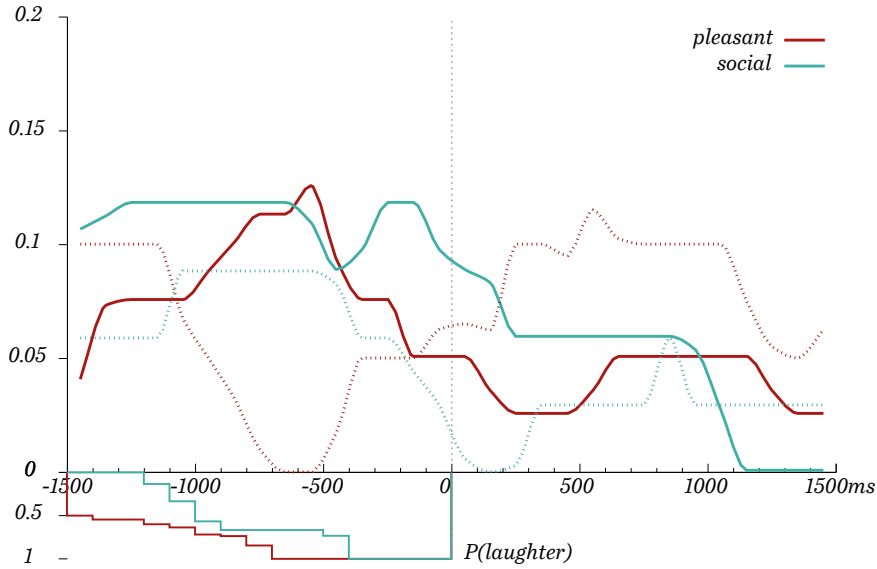


Figure 8.4: Probability of gaze at the offset of laughter depending on incongruity type. Solid line – laugher. Dashed line – partner. The probability of laughter duration is shown at the bottom of the figure.

Figure 8.4 (dashed line) shows the probability of **the partner** gazing at the laugher around the laughter offset depending on *incongruity type*. We observe a significant effect of *time* ($CE = 0.51, SE = 0.22, z = 2.27, p = 0.02$), while the main effect of *incongruity type* is not significant ($CE = 0.27, SE = 0.25, z = 1.07, p = 0.28$). Of particular interest is the significant interaction ($CE = -1.43, SE = 0.38, z = -3.71, p < .001$). After the laughter offset the partner is much less likely to be looking at the laugher if the laughter was related to a social incongruity rather than to a pleasant one, while the opposite pattern is observed before the offset.

8.6.2 *Depending on* laughter coordination

At laughter onset

Figure 8.5 (solid line) shows the probability of **the laugher** gazing at their partner around the laughter onset depending on *laughter coordination*, having isolated laughter as a reference level. We do

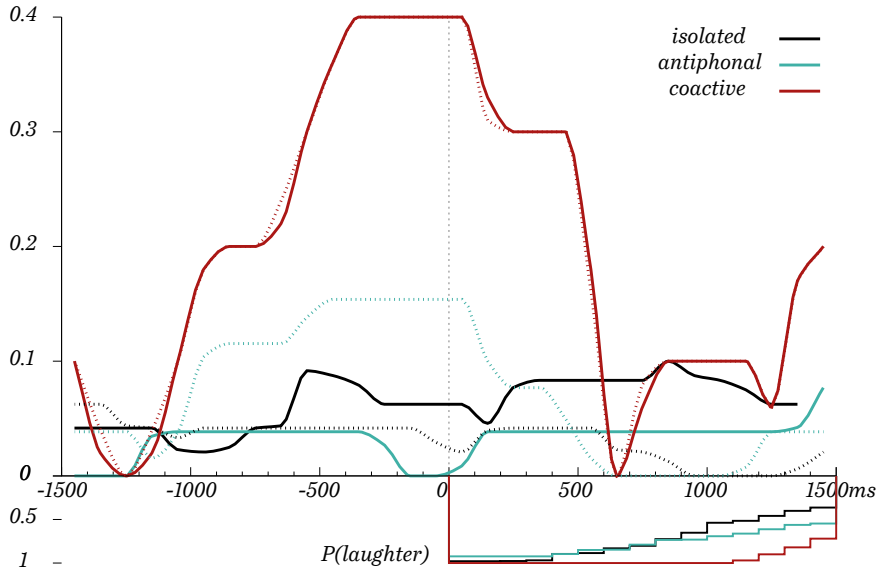


Figure 8.5: Probability of gaze at the onset of laughter depending on laughter coordination. Solid line – laugher. Dashed line – partner. The probability of laughter duration is shown at the bottom of the figure.

not observe any significant difference in the probability of laugher's gaze depending on *time* ($CE = -0.37, SE = 0.21, z = -1.78, p = 0.07$), while we do observe a main effect of *laughter coordination* (Antiphonal-Isolated: $CE = -0.76, SE = 0.31, z = -2.46, p = 0.01$; Coactive-Isolated: $CE = 1.24, SE = 0.25, z = 4.79, p < .001$). Laughers gaze at their partners more in coactive laughter, followed by isolated laughter with the least gaze at partners for antiphonal laughter. No interaction was significant (*time* $\times$ Antiphonal-Isolated: $CE = -0.24, SE = 0.45, z = -0.54, p = 0.58$; *time* $\times$ Coactive-Isolated: $CE = 0.57, SE = 0.35, z = 1.63, p = 0.10$).

Figure 8.5 (solid line) shows the probability of **the partner** gazing at the laugher around the laughter onset depending on *laughter coordination*, having isolated laughter as a reference level. We observed significant main effects of all the predictors included in the model, but no significant interactions: partner's gaze is more likely before the onset of the laugh (*time*: $CE = -0.69, SE =$

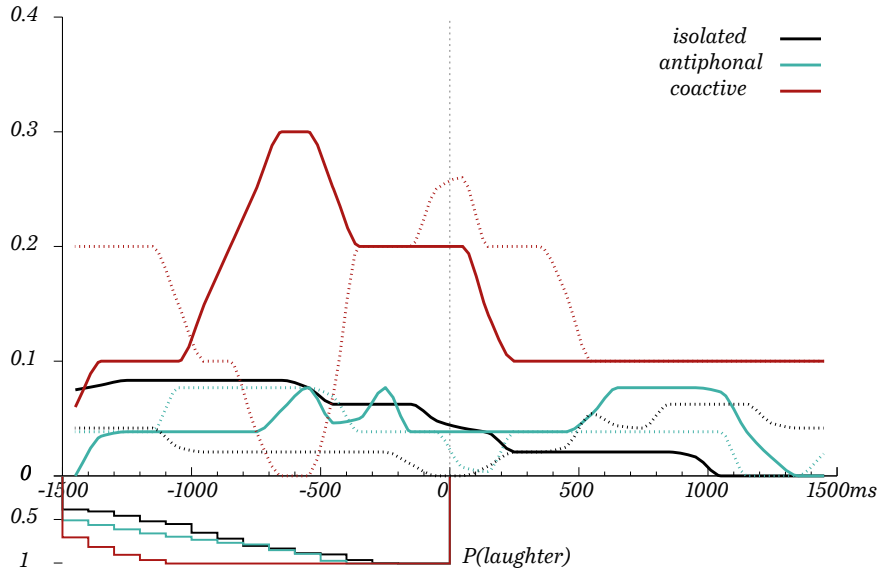


Figure 8.6: Probability of gaze at the offset of laughter depending on laughter coordination. Solid line – laugher. Dashed line – partner. The probability of laughter duration is shown at the bottom of the figure.

$0.31, z = -2.28, p = 0.02$), significantly more likely around the onset of an antiphonal laughter than an isolated one (Antiphonal-Isolated: $CE = 1.06, SE = 0.26, z = 3.81, p < .001$), and even more likely before the onset of a coactive laughter (Coactive-Isolated: $CE = 2.22, SE = 0.29, z = 7.53, p < .001$).

At laughter offset

Figure 8.6 (solid line) shows the probability of **the laugher** gazing at their partner around the laughter offset depending on *laughter coordination*, having isolated laughter as a reference level. The significant effect of *time* shows that the laugher is more likely to be gazing at her partner before rather than after the offset of the laughter ($CE = 1.62, SE = 0.32, z = 5.04, p < .001$). We observed laugher's gaze to be significantly more likely when the laugher is producing an antiphonal laughter in comparison to an isolated one (Antiphonal-Isolated: $CE = 1.6, SE = 0.39, z = 4.08, p < .001$),

and even more likely when producing a coactive laughter (Coactive-Isolated: $CE = 2.49, SE = 0.4, z = 6.19, p < .001$). We observe significant interactions of *time* and *laughter coordination* (*time* $\times$ Antiphonal-Isolated: $CE = -1.62, SE = 0.46, z = -3.51, p < .001$; *time* $\times$ Coactive-Isolated: $CE = -0.94, SE = 0.47, z = -2.0, p = 0.04$). Laughers gaze at their partners more before the offset of a antiphonal laughter that an isolated one, and more before the offset of coactive laughter than an isolated one. This probably can be explained by the steep decline of the isolated laughter line (solid, black) on Figure 8.6.

Figure 8.6 (dashed line) shows the probability of **the partner** gazing at laugher around the laughter offset depending on *laughter coordination*, having isolated laughter as a reference level. We observed a significant main effect of *time* ($CE = -0.72, SE = 0.28, z = -2.55, p = .01$), no significant difference between antiphonal and isolated laughter (Antiphonal-Isolated: $CE = 0.14, SE = 0.27, z = 0.51, p = 0.6$), but a significant difference between coactive and isolated laughter (Coactive-Isolated: $CE = 1.19, SE = 0.31, z = 3.81, p < .001$). We also observe a significant interaction between *time* and *laughter coordination* (*time* $\times$ Antiphonal-Isolated: $CE = 1.59, SE = 0.38, z = 4.14, p < .001$; *time* $\times$ Coactive-Isolated: $CE = 0.89, SE = 0.43, z = 2.04, p = 0.04$). Figure 8.6 (dashed lines) shows these interactions: whereas the curve for isolated laughter gets up, the antiphonal one gets down and the one for coactive peaks around laughter offset.

8.7 DISCUSSION

Our data show that laughter related to different types of laughables, performing different pragmatic functions in interaction, is characterised by different accompanying gaze patterns both from the laugher and from their partner. These observations confirm that both laughter and gaze play a crucial pragmatic role in the unfolding of dialogue. We thus provide further evidence for the stance that to model gaze behaviour one needs to consider not only turns, but also propositional content and dialogue acts performed

(see Part III and Torres et al., 1997). Further, our results validate the laughter taxonomy proposed in Mazzocconi et al. (2020), with respect to two types of incongruity, showing that laughs belonging to different classes are produced and perceived as performing different pragmatic functions, eliciting different multimodal behaviours from the interactants. Below we discuss our results concerning laugher's and partner's gaze in relation to the incongruity type, and in relation to laughter coordination between interlocutors.

8.7.1 Laugher's gaze and incongruity type

Hypothesis 1 (short version). We expect a laugher's gaze to be less likely to be directed at their partner if the laughter is related to a social incongruity rather than to a pleasant incongruity (both around the onset and offset of the laugh).

We observe that the laugher is more likely to gaze at the participant *before* the onset of a laughter related to *pleasant incongruity* rather than to *social incongruity*, while the opposite is true *after* the onset. This result clashes with our Hypothesis 1, according to which we expected an absence of gaze to the partner both *before* and *after* the onset of a laughter related to *social incongruity*. Our hypothesis was based on Stanley and Martin (1968) and Schneier et al. (2011) proposing that social stress makes gaze aversion more likely. Kendrick and Holler (2017) also suggest that gaze aversion is more likely while producing a dispreferred answer, which is a dialogue act that belongs to the *social incongruity* laughable class in Mazzocconi et al. (2020)'s taxonomy.

However, we can explain this data considering that we are analysing gaze during the *laughter* rather than during the *laughable* production (for instance, a dispreferred answer can constitute a laughable). Most of laughs follow the production of the laughable (see Chapter 6), hence our data might be consistent with Kendrick and Holler (2017). Indeed, we observe a lower probability of gaze at the partner *before* the onset of laughter – a time that often coincides with laughable production. We might therefore imagine a scenario where the expression of a dispreferred answer is not accompanied

by gaze at the partner, and only while laughing the laugher looks at the partner to monitor that the laughter has smoothed the disagreement or the unmet expectation, having the desired positive effect. Our data cannot therefore neither confirm nor disconfirm the social stress explanation of gaze aversion. It is possible that the 'social stress' component is what influences the laugher to not look at her partner *before* the onset of a laughter related to *social incongruity*, while at the same time being the motivation to check, during the laughter production, how the laughter (produced to ease the situation) is being perceived by the partner.

The lower probability of gaze *after* the onset of laughter related to *pleasant incongruity* matches the results reported by Gironzetti (2017), who observed a lower probability of attention directed towards smiling facial expressions (including laughter) when it occurred in the context of humorous exchanges in comparison to non-humorous ones. These observations can be interpreted in the light of Becker-Asano and Ishiguro (2009): they observed that when their robot was directing its gaze at the partner while laughing, the laughter was interpreted negatively as being directed *at* the participant, rather than being produced cooperatively. We can therefore speculate that the tendency to avoid looking at the partner while producing a laugh related to *pleasant incongruity* is a way to disambiguate the laugher's intention and social attitude towards the partner.

The opposite pattern observed *before* the laughter onset, i.e. the higher probability of laugher gazing at the partner when about to produce a laughter related to *pleasant incongruity*, might be a result of the fact that the laugher is 'careful' to assess whether laughter is an appropriate contribution, before producing it. There are indeed judgemental, moral, and cognitive aspects related to laughter production (e.g., not everything can be the subject of laughter, it is silly to laugh at some things, and some laughter can be offensive to some people, as discussed in Section 3.6).

The consistency of our results with Gironzetti (2017) is interesting considering the differences between corpora in terms of physical arrangement, data considered, and the task. In Gironzetti (2017),

participants were opposite each other, conversing freely (without the goal-directed task of our study), and they also considered smiling, suggesting that similar dynamics are at play in the multimodal integration of smiling and laughter pragmatic processing. In our corpus, by contrast, participants are engaged in a task which requires them to pay (and share) attention on objects on the table in front of them, and they are seated at a 90° angle (a setting similar to the referential communication task in Sandgren et al., 2012). This means that gaze at the partner and mutual gaze are rarer than in other corpora (engagement in competing activities allows to look away from their partner more frequently, Rossano, 2013), which allows us to speculate that when gaze at the partner instead does occur it is specifically motivated by pragmatic functions.

8.7.2 Partner's gaze

Partner's gaze at the laugher is significantly more likely to occur *before* the laughter onset rather than *after*. This is compatible with the idea of gaze being a cue for eliciting a response (Bavelas et al., 2002; Goodwin and Goodwin, 1986), and that such a response does not have to be a verbal speech turn (Rossano, 2013).

The main effect of *incongruity type* (i.e. that gaze at the laugher is more likely before the onset of laughter related to *pleasant incongruity*) has to be considered together with the data represented in Figure 8.5 about gaze and laughter coordination. We report that antiphonal and coactive laughs are significantly more likely to be preceded by gaze from the partner, meaning that gaze can be interpreted as an invitation to join in. The distribution of antiphonal laughs is though skewed towards *pleasant incongruity* (a pattern replicated in several corpora e.g., Mazzocconi et al., 2020), which therefore constitutes a confounding variable. Due to the small data set we cannot consider such a factor in our statistical modelling. We leave this exploration to further work when a larger dataset will be available.

We also observe that *after* the offset of the laughter, the partner is less likely to look at the laugher if the laugh was related to a *social incongruity*. We interpret this as a 'choice' from the partner to

avoid direct gaze in order to not put extra pressure on the laugher (who has appraised some situation or dialogue act as potentially discomforting) and maybe choose to give feedback (reassuring the laugher) in another modality, signalling that the issue should be dismissed as not disturbing (Romaniuk, 2009).

The higher probability of gaze at the laugher *after* the offset of laughter related to a *pleasant incongruity*, may be explained as a partner's strategy to check whether it would be appropriate to join the laughter. This is consistent with the results reported in Figure 8.3, i.e. laughers being more likely to gaze at the partner *before* the production of laughter related to a *pleasant incongruity*. Laughing can indeed be 'no laughing matter': not all laughs should be reciprocated (Jefferson, 1984).

Hypothesis 2 (short version). On the basis of studies indicating that laughter can function as an attention getting device we hypothesise that after laughter is produced interlocutors will direct their gaze at the laugher.

Our Hypothesis 2 is partially invalidated in the setting of our corpus. We do not observe a significantly higher probability of gaze at the laugher *after* the onset of the laughter (Figure 8.3), but rather *before* the onset. This data can be explained considering that the participants are sitting very close, engaged in a task, and mutual attention is already granted without needing to be signalled through gaze. Our data, on the other hand, highlight the role of laughter to elicit a (laughter) response from the partner (Bavelas et al., 2002).

8.7.3 Laughter coordination

Hypothesis 3 (short version). We expect laughs where one participant joins in with another's laugh (*joining in* laughs) to be preceded by an 'inviting' gaze from their partner. Similarly we expect the person joining the laugh to gaze at the partner.

Hypothesis 3 was partly confirmed: we observed higher probability of gaze from the partner at the onset of an antiphonal or

coactive laughter, but we did not observe a higher probability of looking at the partner from the laugher preceding the production of an antiphonal laughter. We therefore do not observe the need for a 'gaze window' in order to respond to a laughter with a laughter. It would be nonetheless interesting in future work to control whether the laughter was produced by the participants in turn-initial position or not, in order to see if the observations mirror the patterns observed for speech (e.g., Bavelas et al., 2002). Nevertheless, our data show the important role played by gaze in the coordination of laughter production between participants and its role for eliciting responses from the partner, not just in terms of speech turns (Rossano, 2013).

The role of gaze for laughter coordination is particularly striking in the case of coactive laughter (i.e. both interlocutors start laughing at the same onset time), where participants seem to look at each other not only to synchronise on the simultaneous onset but also to terminate the laughter. This kind of gaze may not only serve the purpose of syncing the response, but also monitoring other non-verbal cues about the partner's disposition towards the laughable. This is an open question for future work.

8.7.4 Limitations

Our work provides evidence for the important link that gaze behaviour has for coordination in interaction, and also stresses the interaction between gaze, non-verbal behaviour and dialogue acts. Our data therefore offer new material for modelling multimodal meaning construction in interaction – important not only from a theoretical linguistic perspective, but also for the implementation of ECAS able to be more pragmatically adequate and to read non-verbal cues from the user to refine their own behaviour (e.g., if the user laughed and looked at the ECA, a likely adequate response might be to laugh back). More complex models for semantic processing are needed (to be discussed in Parts III and IV of this thesis) in order to tune ECAS behaviour to the pragmatic functions performed by the laughter, but our results suggest that gaze might

be one of the cues to be considered in order to classify the type of laughter.

Our study is limited by the small sample size analysed. We are in the process of extending our dataset, which will allow us to employ more complex statistical models able to account for several variables at the same time (e.g., laughter position in relation to the speech-turn, to the laughable placement, arousal). Cross-cultural studies (e.g., Rossano et al., 2009) showed differences in gaze behaviours between different communities (a consideration relevant for the implementation of ECAS appropriate to the user's culture). Our results therefore should not be taken as absolute, but open up the possibility for interesting comparative studies.

Part III

Laughter interpretation

Introduction to Part III

Смейся, смейся громче всех,
Милое созданье.
Для тебя веселый смех,
Для меня — страданье.
Laugh, laugh loudest,
Lovely creature.
For you – laughter,
For me – suffering.

1940 song, music by Vadim Kozin, words by Yakov Yadov

This part focusses on how laughter contributes to the meanings of utterances and what the dialogue system can do with it. We are going to explore how laughter can be integrated into the dialogue context to form a communicative intent or to modify a dialogue state. Laughter interpretation is a concern of both the natural language understanding (NLU) and the dialogue manager (DM) modules.

Firstly, in Chapter 9 we look at how laughter is integrated on a NLU level, analysing how laughter is correlated with different dialogue acts and how it can be used to help a computational model predict a dialogue act.

Then, in Chapter 10 we introduce the formalism which is used to express the rules of transformations of information-states of dialogue participants. Linear Dialogue Manager (LDM) is an implemented framework which uses proof search on top of a linear logic. It integrates the insights from KoS (Ginzburg, 2012) and information-state update (ISU) and provides the grounds for integrating laughter within a dialogue management component of a dialogue system.

Finally, in Chapter 11 we proceed with the next step of laughter interpretation, on the level of the DM, analysing how laughter can update the context of a dialogue. In this section we are only concerned with some non-humorous uses of laughter, leaving out the relationship between humour and laughter for the last part of the thesis.

9 Laughter which forms a dialogue act

The content of this chapter was previously published in Maraev, Noble et al. (2021).[1] *It has been substantially revised.*

9.1 INTRODUCTION

A dialogue manager typically receives a user *move* – a structured representation of user input produced by the natural language understanding (NLU) module, usually, either in the form of intents and entities or represented as dialogue acts (see Sections 5.2 and 4.2). Here we will look at the latter, more sophisticated and general case, investigating how laughs relate to the dialogue acts.

In Section 9.2 we start by introducing the task of dialogue act recognition and further motivating our study. Then, in Section 9.3 we look at collocations of dialogue acts and laughs, providing statistical and qualitative observations. In Section 9.4 we explore the capacity of several Transformer-based neural network models for dialogue act recognition. We conclude with discussing the implications and limitations of our findings in Section 9.5.

[1] Vladislav Maraev, Bill Noble, Chiara Mazzocconi and Christine Howes (2021). 'Dialogue act classification is a laughing matter'. In: *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Potsdam, Germany: SEMDIAL. The implementation of neural networks was done by Bill Noble. Chiara Mazzocconi put together the qualitative observations (Section 9.3.2). All authors contributed to the discussion of the results.

As discussed in Section 4.2, the concept of a dialogue act (DA) is based on that of the speech act (Austin, 1975). Breaking with classical semantic theory, Speech Act Theory considers not only the propositional content of an utterance but also the actions, such as promising or apologising, it carries out. DAS extend the concept of the speech act, with a focus on the interactional nature of most speech. Dialog Act Markup in Several Layers (DAMSL) (Core and Allen, 1997), for example, is an influential DA tagging scheme where DAS are defined in part by whether they have a forward-looking function (expecting a response) or backward-looking function (in response to a previous utterance).

Dialogue act recognition (DAR) is the task of labelling utterances with the dialogue act they perform, given a set of dialogue act tags. As with other sequence labelling tasks in natural language processing (NLP), some notion of context is helpful in DAR. One of the first machine learning models for DAR was a Hidden Markov Model (HMM) that used lexical and prosodic features as input (Stolcke et al., 2000). Blache et al. (2020) has shown that high-level syntactic features, for instance, the number of modifiers and complex clauses, can improve DAR.

Recent state-of-the-art approaches to dialogue act recognition have used a hierarchical approach, using large pre-trained language models like the Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2019) to represent utterances, and adding some representation of discourse context at the dialogue level (e.g., Ribeiro et al., 2019; Mehri et al., 2019). Recently, neural models trained on large datasets with various unsupervised tasks have found success creating sentence-level and contextual word representations.

However, Noble and Maraev (2021) observe that without fine-tuning, standard BERT representations perform very poorly on dialogue, even when paired with a discourse model, suggesting that certain utterance-internal features missing from BERT's textual

pre-training data, such as laughter, may have an adverse effect on dialogue act recognition.

Research on laughter suggests the potential of laughter to modify the interpretation of an utterance and action performed by it, as discussed in Section 3.4. For instance, Bryant (2016) shows how listeners are influenced towards a non-literal interpretation of sentences when accompanied by laughter. Similarly, Tepperman et al. (2006) shows that laughter can act as a contextual feature for determining the sincerity of an utterance, e.g., when detecting sarcasm.

Nevertheless there is a dearth of research exploring the use of laughter in relation to different dialogue acts in detail, and therefore little is known about the role that laughter may have in facilitating the detection of communicative intentions.

We investigate the disambiguation potential of laughter for interpreting the meaning of an utterance, with the central point being the dialogue act performed by it. To do so, we employ a Transformer-based model and look into laughter as a potentially useful feature. Laughs are not present in large-scale pre-trained models, such as BERT, but their representations can be learned while training for a dialogue-specific task (DAR in our case). We further explore whether such representations can be additionally learned, in an unsupervised fashion, from dialogue-like data, such as a movie subtitles corpus, and if it further improves the performance of our model.

We start with observing how dialogue acts can be classified with respect to their collocations with laughs and discuss the patterns observed in relation to the pragmatic functions that laughter can perform in dialogue.

9.3 CORPUS STUDY

In this section we analyse dialogue acts in the Switchboard Dialogue Act Corpus (SWDA) according to their collocation with laughter and provide some qualitative insights based on the statistics.

Dialogue act	*Example*	*#*	*%*
Statement-non-opinion	Me, I'm in the legal department.	72824	36
Acknowledge (Backchannel)	Uh-huh.	37096	19
Statement-opinion	I think it's great	25197	13
Agree/Accept	That's exactly it.	10820	5
Abandoned or Turn-Exit	So, -	10569	5
Appreciation	I can imagine.	4633	2
Yes-No-Question	Do you have to have any special training?	4624	2
Non-verbal	[Laughter], [Throat_clearing]	3548	2
Yes answers	Yes.	2934	1
Conventional-closing	Well, it's been nice talking to you.	2486	1
Uninterpretable	But, uh, yeah	2158	1
Wh-Question	Well, how old are you?	1911	1
No answers	No.	1340	1
Response Acknowledgement	Oh, okay.	1277	1
Hedge	I don't know if I'm making any sense or not.	1182	1
Declarative Yes-No-Question	So you can afford to get a house?	1174	1
Other	Well give me a break, you know.	1074	1
Backchannel in question form	Is that right?	1019	1
Quotation	You can't be pregnant and have cats	934	.5
Summarize/reformulate	Oh, you mean you switched schools for the kids.	919	.5
Affirmative non-yes answers	It is.	836	.4
Action-directive	Why don't you go first	719	.4
Collaborative Completion	Who aren't contributing.	699	.4
Repeat-phrase	Oh, fajitas	660	.3
Open-Question	How about you?	632	.3
Rhetorical-Questions	Who would steal a newspaper?	557	.2
Hold before answer/agreement	I'm drawing a blank.	540	.3
Reject	Well, no	338	.2
Negative non-no answers	Uh, not a whole lot.	292	.1
Signal-non-understanding	Excuse me?	288	.1
Other answers	I don't know	279	.1
Conventional-opening	How are you?	220	.1
Or-Clause	or is it more of a company?	207	.1
Dispreferred answers	Well, not so much that.	205	.1
3rd-party-talk	My goodness, Diane, get down from there.	115	.1
Offers, Options, Commits	I'll have to check that out	109	.1
Self-talk	What's the word I'm looking for	102	.1
Downplayer	That's all right.	100	.1
Maybe/Accept-part	Something like that	98	<.1
Tag-Question	Right?	93	<.1
Declarative Wh-Question	You are what kind of buff?	80	<.1
Apology	I'm sorry.	76	<.1
Thanking	Hey thanks a lot	67	<.1

Table 9.1: Frequency counts (number of utterances) and examples of dialogue acts in the SWDA corpus. The table was adapted from Jurafsky et al. (1997a).

SWDA is tagged with a set of 220 dialogue act tags which, following Jurafsky et al. (1997b), we cluster into a smaller set of 42 tags. Table 9.1 shows frequency counts for SWDA. The distribution of laughs in different dialogue acts has a rather uniform shape with a few outliers (Figure 9.1). The most distinct outlier is the *Non-verbal* dialogue act which is misleading with respect to laughter, because utterances only containing a single laughter token fall into this category. However isolated laughs can serve, for example, to acknowledge a statement, to deflect a question, or to show appreciation (see Chapter 11, Ginzburg et al., 2020 and Mazzocconi, 2019). We will add a further conjecture on this class of DAS in Section 9.4.5.

9.3.1 Method

Let us introduce our comparison schema using the other two outliers, *Downplayer* (make up 0.05% of all utterances) and *Apology* (0.04%), comparing them with the most common dialogue act in SWDA – *Statement-non-opinion* (33.27%). Downplayers are commonly produced as the adjacency counterpart to apologies (second element of the utterance pair produced by the other speaker, Schegloff and Sacks, 1973), realised, for instance by utterances like 'Don't worry' or 'It's alright'.

We take into consideration five laughter-related dimensions of an utterance, and create 5-dimensional (pentagonal) representations of DAS according to these, which we call *laughter profiles*. Each dimension's value is equal to the proportion of utterances of a given type which contain laughter:

↑ current utterance;
↖ immediately preceding utterance by the same speaker;
↗ immediately following utterance by the same speaker;
↙ immediately preceding utterance by the other speaker;
↘ immediately following utterance by the other speaker.

For instance, (9a) is an illustrative example of the phenomenon shown in Figure 9.2.

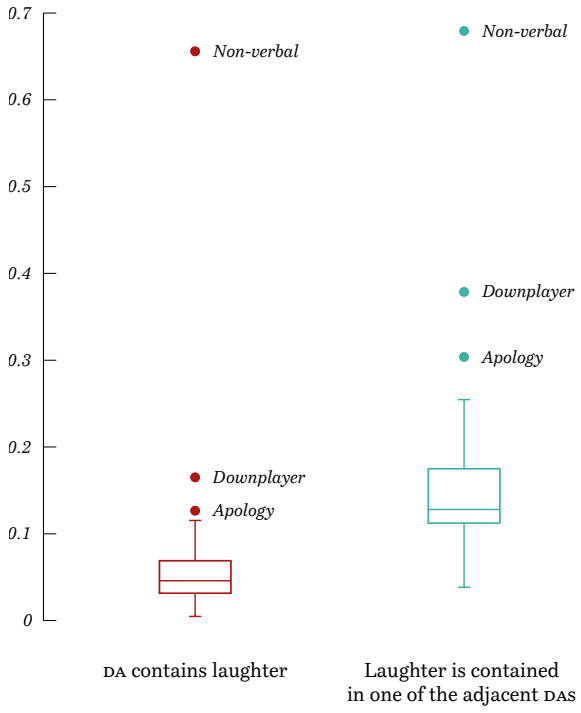


Figure 9.1: Box plots for proportions of dialogue acts which contain laughs in SWDA.

(9a) 1 *A.* I'm sorry to keep you waiting #<laughter>.#[2] *Apology*
2 *B.* #Okay# <laughter>. *Downplayer*
3 *A.* Uh, I was calling from work *Statement-non-opinion*
sw4660

We show the laughter profiles for all dialogue acts in Figure 9.3. We believe that such a depiction helps the reader form impressions about similarities between DAs based on their laughter collocations and notice the ones that stand out in some respects.

To further assess the similarity between dialogue acts based on their collocations with laughs we factorise their laughter profiles

[2]Overlapping material is marked with hash signs.

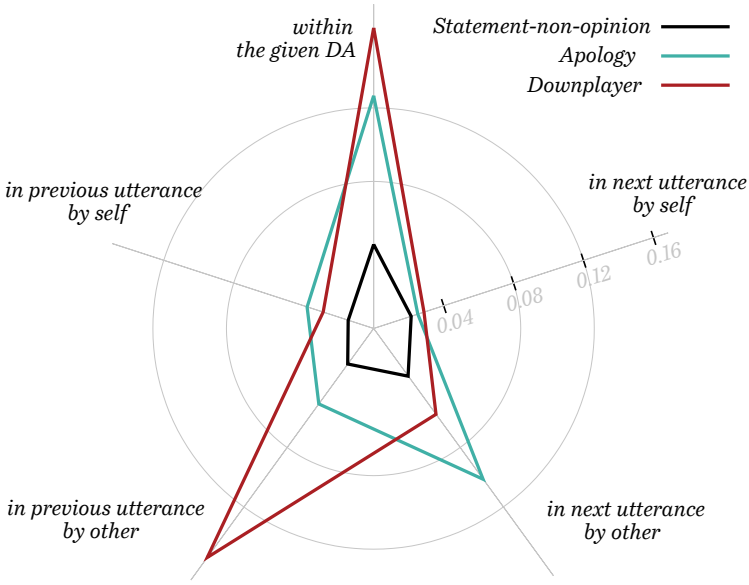


Figure 9.2: Comparison between the *laughter profiles* – pentagonal representations of laughter collocations of dialogue acts.

into 2D space using singular value decomposition (SVD). We can see that dialogue acts form some distinct clusters. The resulting plot is shown in Figure 9.4. Let us now proceed with some qualitative observations.

9.3.2 Observations

Laughter and modification or enrichment of the current DA

We observe a higher proportion of laughter accompanying the current dialogue act (↑) when the laughter is aimed at modifying the current dialogue act with some degree of urgency in order to smooth or soften it (*Action-directive*, *Reject*, *Dispreferred answer*, *Apology*), to contribute to its enrichment stressing the positive disposition towards the partner (*Appreciation*, *Downplayer*, *Thanking*), or to cue for the need to consider a less probable meaning, therefore helping in non-literal meaning interpretation (*Rhetorical-questions*).

While *Apology* and *Downplayer* have rather distinct and pecu-

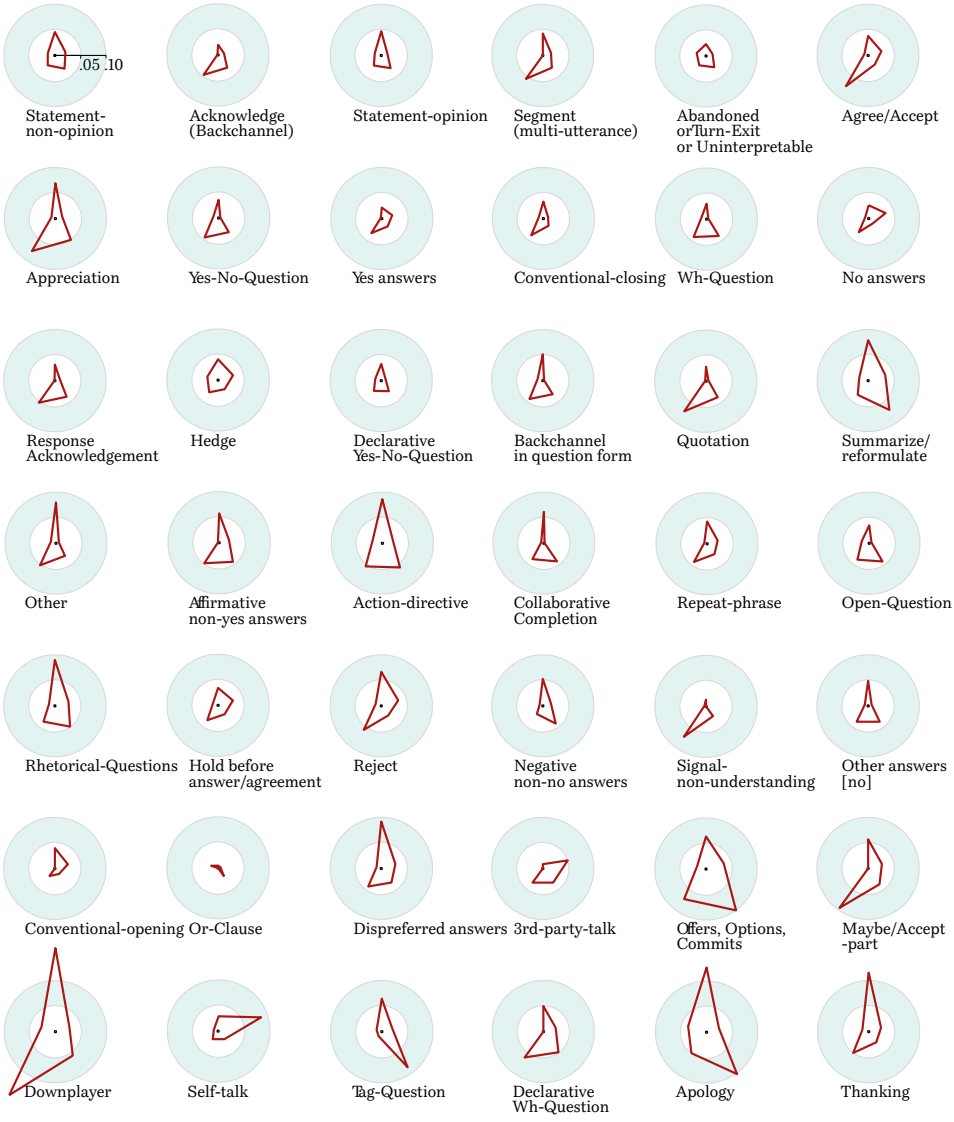


Figure 9.3: Laughter profiles: proportions of utterances which include laughter. Dimensions: ↑ current utterance; ↖ immediately preceding utterance by the same speaker; ↗ immediately following utterance by the same speaker; ↙ immediately preceding utterance by the other speaker; ↘ immediately following utterance by the other speaker. DAs are ordered by their frequency in SWDA (left-to-right, then top-to-bottom).

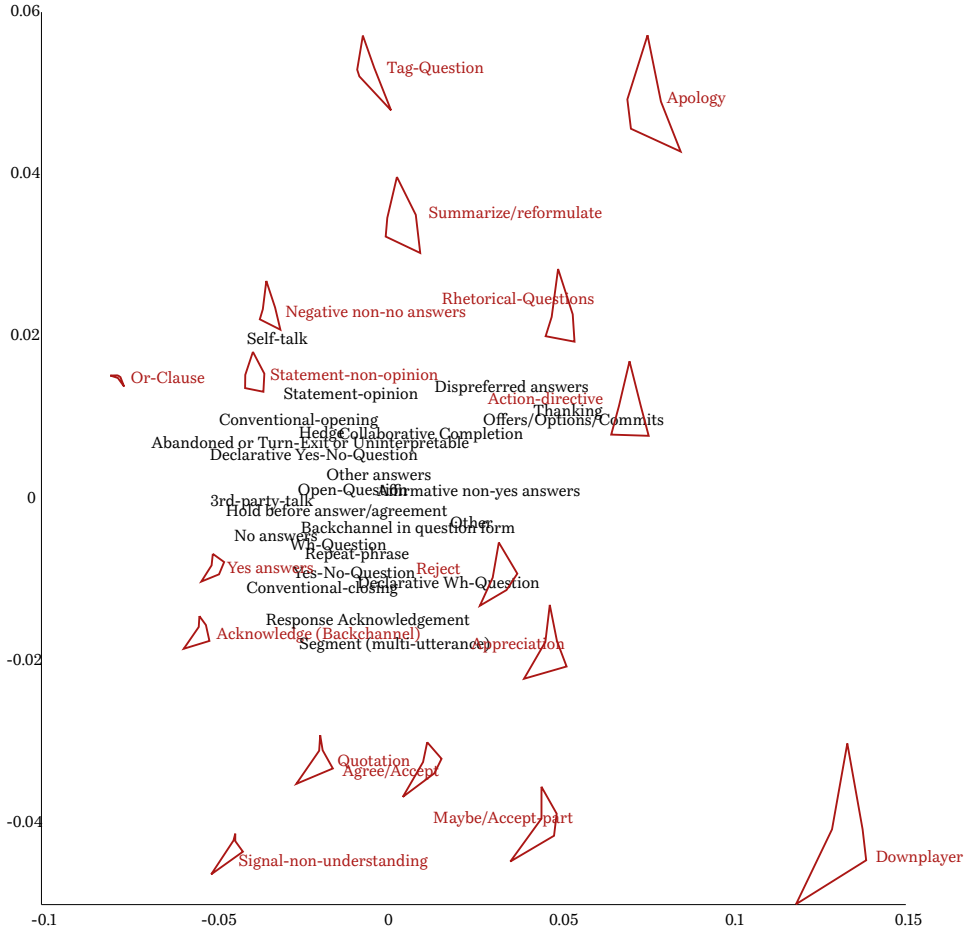


Figure 9.4: Singular value decomposition of pentagonal representations of dialogue acts. For a selection of dialogue acts we depict their laughter profiles.

liar profiles discussed in more detail below, we observe *Dispreferred answers*, *Action-directive, Offers/Options/Commits* and *Thanking* to constitute a close cluster when considering the decomposed values of the laughter profiles (see Figure 9.4, around *Action-directive* DA).

Laughter for benevolence induction and laughter as a response

The profiles observed in relation to the preceding and following turns reflect the multitude of functions that laughter can perform in interaction, stressing the fact that it can be used both to induce or invite a determinate response (dialogue act) from the partner (*Downplayer*, *Agree/Accept*, *Appreciation*, *Acknowledge*) as well as being a possible answer to specific dialogue acts (e.g., *Apology*, *Offers/Options/Commits*, *Summarise/Reformulate*, *Tag-question*).

A peculiar case is then the one of *Self-talk*, often followed by laughter by the same speaker. In this case the laughter may be produced to signal the incongruity of the action (in dialogue we normally speak to others, not to ourselves), while at the same time function to smooth the situation, for instance, when having issues of lexical retrieval, as in (9b), or some degree of embarrassment from the speaker, when questioning whether a contribution is appropriate or not, as in (9c).

(9b) 1 *A.* Have, uh, really, -
2 *A.* what's the word I'm looking for, *Self-talk*
3 *A.* I'm just totally drawing a blank <laughter>.
Statement-non-opinion
sw2372

(9c) 1 *B.* Well, I don't have a Mexi-, - *Statement-non-opinion*
2 *B.* I don't, shouldn't say that, *Self-talk*
3 *B.* I don't have an ethnic maid <laughter>.
Statement-non-opinion
sw2368

Apology and Downplayer

It is interesting to comment on the parallelisms of laughter usage in relation to *Apology* and *Downplayer*, represented in Figure 9.2 in contrast to *Statement-non-opinion*, in as much as their graphic representations are more or less mirror-images of each other and show how the dialogue acts are linked by the pragmatic functions laughter can perform in dialogue.

In both *Apology* and *Downplayer* we observe a rather higher proportion of occurrences in which the dialogue act is accompanied by laughter (↑) in comparison to other DAS (see Figure 9.3). In the case of *Apology*, laughter can be produced to induce benevolence from the partner (Mazzocconi et al., 2020), while in the case of *Downplayer* the laughter can be produced to reassure the partner about some situation that had been appraised as discomforting (classified as *social incongruity* by Mazzocconi et al., 2020, see Section 3.3) and somehow signal that the issue should be regarded as not important (Romaniuk, 2009), as in (9a) and (9d).

(9d) 1 *A.* I don't, I don't think I could do that <laughter>. *Statement-non-opinion*
2 *B.* Oh, it's not bad at all. *Downplayer*
3 *A.* It's, it's a beautiful drive. *Statement-non-opinion*
sw2711

(9a) 1 *A.* I'm sorry to keep you waiting #<laughter>.# *Apology*
2 *B.* #Okay# <laughter>. *Downplayer*
3 *A.* Uh, I was calling from work *Statement-non-opinion*
sw4660

The interesting mirror-image profiles observable in the lower part of the graph can therefore be explained by considering the relation between the two dialogues acts. We observe cases in which an *Apology* is accompanied by a laughter, and then followed by a *Downplayer*, showing that the laughter's positive effect was attained and successful. This allows us to explain both the bottom left spike (↙) observed for *Downplayer* (often preceded by an utterance by

the partner containing laughter) and the bottom right spike (↘) observed for *Apology* (often followed by an utterance by the partner containing laughter). In example (9a) both the apology and the downplayer are accompanied by laughter, while in (9e) a typical example of a laughter accompanying an *Apology* is followed by a *Downplayer*.

(9e)	1	*B.* I'm sorry <laughter>.	*Apology*
	2	*A.* That's all right.	*Downplayer*
	3	*B.* You, you were talking about, uh, uh,	*Summarise*
			sw2434

9.3.3 Politeness theory perspective

Politeness theory, as proposed by (Brown and Levinson, 1987) offers an interesting perspective on laughter in the context of dialogue acts. It introduces the following concepts that seem promising for our analysis.

Face 'The public self-image that every member wants to claim for himself' (ibid.), it consists of positive and negative aspects.

Negative face 'The basic claim to territories, personal preserves, rights to non-distraction – i.e. to freedom of action and freedom from imposition' (ibid.).

Positive face 'The positive consistent self-image or "personality" (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants' (ibid.).

Face-threatening act (FTA) What is intended to be done by either verbal or non-verbal communication, that might threaten the face of the speaker or their partner.

Face-saving tactic According to Brown and Levinson the agents (who are assumed to be rational) will seek to minimise the threat inflicted by face-threatening acts using certain strategies.

According to Brown and Levinson (1987) there are FTAS of four types, classified by the *face* which is threatened:

S- negative face of the speaker,

S+ positive face of the speaker,

L- negative face of the listener,

L+ positive face of the listener.

We claim that laughter is one face-saving tactic, following the characterisation of counter-acting with laughter the face-threatening acts caused by trouble-telling Brown and Levinson, 1987. The trouble-teller threatens their own positive face and might offer some comic relief and try to induce a hesitant laughter from the hearer (Jefferson, 1984). It is evident that this is not the only example and other FTAS can be mitigated by laughter.

There is no clearcut correspondence between DA and FTA types, and Politeness theory wasn't taken into consideration in the DAMSL annotation protocol. There are few cases for which we can conjecture a clear connection between the DA and the FTA, given that they were explicitly discussed by Brown and Levinson. Interestingly, these are the dialogue acts which are strongly, according to our corpus study, associated with laughter.

Two obvious *S*- FTAS seem to be present in DAMSL: *Thanking* and *Downplayer*. For the latter Brown and Levinson explicitly state that the speaker 'may feel constrained to minimize [the hearer's] debt or transgression, as in "It was nothing, don't mention it."' (ibid.). These dialogue acts have very similar laughter profiles.

One dialogue act that can correspond to *S*+ is *Apology,* because it is a expression of regret for some previous FTA. Other examples of *S*+ are various kinds of self-humiliating actions, which could be *Self-talk*, as in (9c) where the speaker gives a negative evaluation of their own words, and *3rd-party-talk*. These three dialogue acts have quite distinct laughter profiles. *Agree/accept* can also act as *S*+, in the case of a compliment, or S-, in the case of accepting an offer.

There are a few of clear examples of *H-* too: *Action-directive* (as in telling the hearer what to do), *Offers, Options, Commits* (putting pressure on the hearer), *Summarise/Reformulate* (suggesting to the hearer a better formulation of what was just said). The last two cases not only often contain laughter in the DA itself, but are also often followed by laughter from the partner. A similar profile is also observed for *Tag-Questions*, e.g., a question 'Right?' can put on the hearer some additional pressure to agree with the speaker.

The last category of FTAS is *H+* and it includes various forms of criticisms, contradictions and disagreements, which are often associated with laughter, as discussed Section 3.3. The associated DAS are *Reject* and *Dispreferred answers*. *Rhetorical questions* can also fall into this category, because they can potentially challenge a hearer.

Overall, Politeness theory not only provides a new perspective on laughter profiles of the dialogue acts, but also puts the concept of *social incongruity* under additional scrutiny and gives one a potential to get a more specific definition and evaluation of it.[3]

We now turn to the question of whether our qualitative observations of patters between laughs and dialogue acts can be used to improve a dialogue act recognition task.

9.4 LAUGHTER FOR DIALOGUE ACT RECOGNITION

In this section we will study whether Transformer-based neural networks (NNS) can predict how laughter interacts with the DA of an utterance. We will carry out three experiments.

1. We examine if laughter can act as a predictive feature for DA and help disambiguation of DAS (Section 9.4.3).

2. We look how dialogue-specific pre-training on data that includes laughs can affect the model performance (Section 9.4.4).

[3]We would like to thank Arash Eshghi for bringing up this topic during Vladislav's final seminar on 31 May 2022.

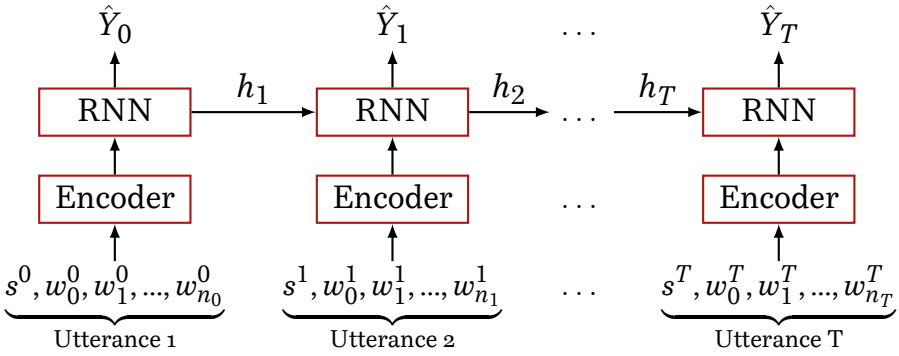


Figure 9.5: Simple neural dialogue act recognition sequence model

3. We look at the *Non-verbal* DAS and enquire into whether the model can predict more meaningful DAS (Section 9.4.5).

9.4.1 The model

We employ a simple neural architecture with two components: an encoder that vectorises utterances, and a sequence model that predicts dialogue act tags from the vectorised utterances (Figure 9.5). Since we are primarily interested in comparing different utterance encoders, we use a basic recurrent neural network (RNN) as the sequence model in every configuration.[4] The RNN takes the encoded utterance as input at each time step, and its hidden state is passed to a simple linear classification layer over dialogue act tags. Conceptually, the encoded utterance represents the context-agnostic features of the utterance, and the hidden state of the RNN represents the full discourse context.

As a baseline utterance encoder, we use a word-level convolutional neural network (CNN) with window sizes of 3, 4, and 5, each with 100 feature maps (Kim, 2014). The model uses 100-dimensional

[4]We have experimented with long short-term memory (LSTM) as the sequence model, but the accuracy was not significantly different compared to RNN. This can be explained by the absence of longer distance dependencies on this level of our model.

word embeddings, which are initialised with pre-trained gloVe vectors (Pennington et al., 2014).

For the BERT utterance encoder, we use the BERT BASE model with hidden size of 768 and 12 transformer layers and self-attention heads (Devlin et al., 2019). In our implementation, we use the uncased model provided by Wolf et al. (2019).

9.4.2 Data

We perform experiments on the Switchboard Dialogue Act Corpus (SWDA, 42 dialogue act tags), which is a subset of the larger Switchboard corpus, and the dialogue act-tagged portion of the Augmented Multi-party Interaction (AMI) Meeting Corpus. AMI Meeting Corpus (AMI-DA) uses a smaller tagset of 16 dialogue acts (AMI Project, 2005).

Switchboard	*AMI Corpus*
Dyadic	Multi-party
Casual conversation	Mock business meeting
Telephone	In-person & video
English	English
Native speakers	Native & non-native speakers
2200 conversations, 1155 in SWDA	171 meetings, 139 in AMI-DA
400k utterances	118k utterances
3M tokens	1.2M tokens

Table 9.2: Comparison between Switchboard and the AMI Meeting Corpus

Preprocessing

We make an effort to normalise transcription conventions across SWDA and AMI. We remove disfluency annotations and slashes from the end of utterances in SWDA. In both corpora, acronyms are tokenised as individual letters. All utterances are lower-cased.

Utterances are tokenised using a word piece tokeniser (Wu et al., 2016) with a vocabulary of 30,000. We add a special laughter token to the vocabulary and map all transcribed laughter to that token.

	SWDA		AMI-DA	
	F_1	acc.	F_1	acc.
BERT-NL	38.10	77.07	49.09	67.06
BERT-L	45.99	76.93	50.17	67.12
CNN-NL	37.23	75.08	38.37	63.46
CNN-L	27.59	75.40	37.94	64.30
Majority class	0.78	33.56	1.88	28.27

Table 9.3: Comparison of macro-average $F_1$and accuracy depending on using laughter in the training phase.

We also prepend each utterance with a speaker token that uniquely identifies the corresponding speaker within that dialogue.

9.4.3 *Experiment 1: Impact of laughter*

In the first experiment we investigated whether laughter, as an example of a dialogue-specific signal, is a helpful feature for DAR. Therefore, we train another version of each model: one containing laughs (L) and one with laughs left out (NL), and compare their performances in DAR task. Table 9.3 compares the results from applying the models with two different utterance encoders (BERT, CNN).

BERT outperforms the CNN on AMI-DA. On SWDA, the two encoders are more comparable, though BERT has a slight edge in accuracy, suggesting that it relies more heavily on defaulting to common dialogue act tags. On SWDA, we see small improvements in accuracy and macro-$F_1$for models that included laughter. For AMI-DA, the effect of laughter is small or even negative – the impact of laughter on performance becomes more clear in the disaggregated performance over different dialogue acts. Indeed, laughter improves the accuracy of the model even on some dialogue acts in which laughter occurs rarely in the current and adjacent utterances (see Figures 9.6 and 9.7).

Confusion matrices in Figure 9.8 provide some food for thought.

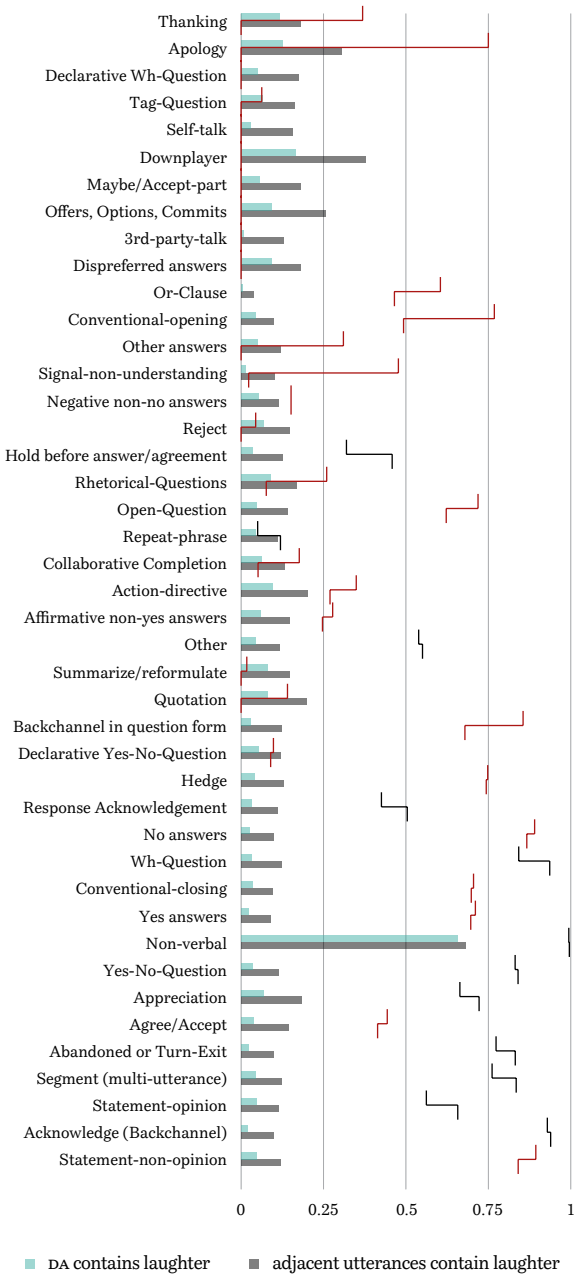


Figure 9.6: SWDA corpus. Change in accuracy for each dialogue act (BERT-NL vs BERT-L). Positive changes when adding laughter (BERT-L) are shown in red. Horizontal bars indicate how often the DA is associated with laughter.

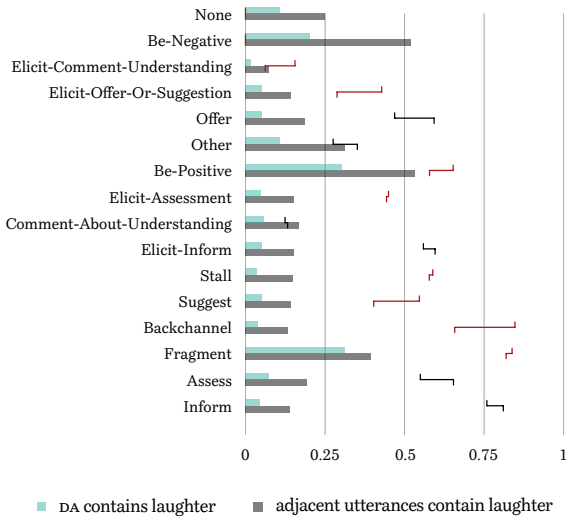


Figure 9.7: AMI-DA corpus. Change in accuracy for each dialogue act (BERT-NL vs BERT-L). Positive changes when adding laughter (BERT-L) are shown in red. Horizontal bars indicate how often the DA is associated with laughter.

Most of the misclassifications fall into the majority classes, such as *Statement-non-opinion*, on the left edge of the matrix. However, there are some important exceptions, such as *Rhetorical-questions*, which are misclassified as other forms of questions due to their surface question-like form. Importantly, laughter helps to classify rhetorical questions correctly, because in a conversation it can be used as a device to cancel seriousness or sincerity (Ginzburg et al., 2015; Tepperman et al., 2006), in this case, of a question. Therefore, questions, like the one we show in example (9f), are easier to disambiguate with the help of laughter.

(9f) (Talking about hobbies)

1 *B.* Um, as far as spare time, they talked about,
Statement-non-opinion

2 *B.* I don't, + I think, *Statement-non-opinion*

3 *B.* who has any spare time <laughter>?
Rhetorical-question

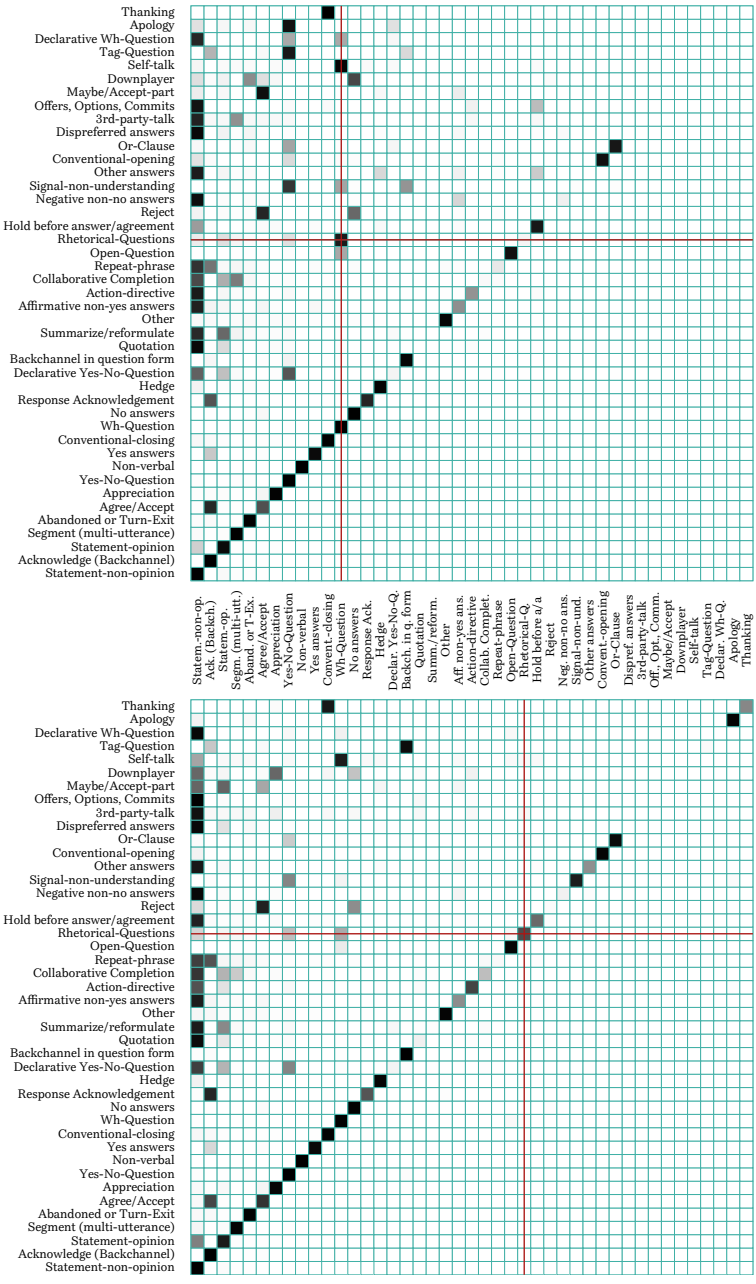


Figure 9.8: Confusion matrices for BERT-NL (top) vs BERT-L (bottom); SWDA corpus. Solid line shows the improvement in classification of rhetorical questions.

4 *B.* <laughter>. *Non-verbal*
sw2434

9.4.4 *Experiment 2: Laughter and pre-training*

As previously noted, training data for BERT does not include features specific to dialogue (e.g., laughs). We therefore experiment with a large and more dialogue-like corpus constructed from OpenSubtitles (Lison and Tiedemann, 2016) (350M tokens, where 0.3% are laughter tokens). We used a manually constructed list of words frequently used to refer to laughter in subtitles and replaced every occurrence of one of these words with the special laughter token. We then collected every English-language subtitle file in which at least 1% of the utterances contained laughter (about 11% of the total). Because utterances are not labelled with speaker in the OpenSubtitles corpus, we randomly assigned a speaker token to each utterance to maintain the format of the other dialogue corpora.

The pre-training corpus was prepared for the combined masked language modelling and next sentence (utterance) prediction task, as described by Devlin et al. (2019).

We analyse how pre-training affects BERT's performance as an utterance encoder. To do so, we consider the performance of DAR models with three different utterance encoders:

1. FT – pre-trained BERT with DAR fine-tuning;
2. RI – randomly initialised BERT (with DAR fine-tuning);
3. FZ – pre-trained BERT without fine-tuning (frozen during DAR training).

For the pre-trained (FT, FZ) conditions we perform two types of pre-training:

1. OSL – pre-training on the portion of OpenSubtitles corpus
2. OSNL – same as OSL, but with all the laughs removed.

We fine-tune and test our models on the corpora containing laughs (L).

	SWDA		AMI-DA	
	F_1	acc.	F_1	acc.
BERT-L-FT	36.75	76.60	43.37	64.87
BERT-L+OSL-FT	41.42	76.95	48.65	68.07
BERT-L+OSNL-FT	43.71	77.09	43.68	64.80
BERT-L+OSL-FZ	9.60	57.67	17.03	51.03
BERT-L+OSNL-FZ	7.69	55.29	16.99	51.46
BERT-L-RI	32.18	73.80	34.88	60.89
Majority class	0.78	33.56	1.88	28.27
SotA	–	83.1 [5]	–	–

Table 9.4: Comparison of macro-$F_1$and accuracy with further dialogue pre-training.

We observe that dialogue pre-training improves performance of the models. Fine-tuned models also perform better than the frozen ones because the latter provide less opportunities for the encoder to be trained for the specific task.

Including laughter in pre-training data improves F_1scores in most cases, except for the SWDA in the fine-tuned condition. The difference is especially pronounced for AMI-DA corpus in the fine-tuned condition (4.97 p.p. difference in F_1). The question of relevance of movies subtitle data for either SWDA or AMI-DA can be a subject for further study, including the types of laughs in the corpora. It might be the case that nature of AMI-DA is congruent with those of movie subtitles, since the participants in AMI-DA basically are role-playing being in a focus group rather than being involved in a natural dialogue.

Next, we will further investigate whether non-verbal dialogue acts can be classified as more specific dialogue acts by our model.

9.4.5 *Experiment 3: Laughter as a non-verbal dialogue act*

In this experiment, following the observations regarding the misleading character of *Non-verbal* dialogue acts, we looked at the

[5] Kozareva and Ravi (2019)

predictions that the model would give this class of dialogue acts if it wasn't aware of the *Non-verbal* class. To do so, we mask the outputs of the model where the desired class was *Non-verbal* and do not backpropagate these results. We used the BERT-L-FT for this experiment. After training we tested the resulting model on the test set containing 659 non-verbal dialogue acts, 413 of which contain laughter.

For 314 (76%) of such dialogue acts the model has predicted the *Acknowledge (Backchannel)* class and for 46 (11%) – continuations of the previous DA by the same speaker. The rest were classified as either something uninformative (the *Abandoned or Turn-Exit or Uninterpretable* class) or, from manual observation, clearly unrelated.

Acknowledge (Backchannel) can cover some uses of laughter, for instance, to show to the interlocutor acknowledgement of their contribution, implying the appreciation of an incongruity, and inviting continuation (functioning simultaneously as a continuer and assessment feedback, Schegloff, 1982), as in example (9g).

(9g) We mark continuations of the previous DA by the same speaker with a plus, and indicate misclassified dialogue acts with a star. Dialogue acts which were annotated as *Non-verbal* in SWDA are marked in red.

1	*B.*	Everyone on the boat was catching snapper, snappers except guess who.	*Statement-non-opinion*
2	*A.*	<laughter> It had to be you.	*Summarise/reformulate*
3	*B.*	<laughter> I ca-, I, -	*Uninterpretable*
4	*A.*	Couldn't catch one to save your life. Huh.	*Backchannel**
5	*B.*	That's right,	*Agree/Accept*
6	*B.*	I would go from one side of the boat to the other,	*Statement-non-opinion*
7	*B.*	and, uh,	+
8	*A.*	<laughter>.	*Backchannel*
9	*B.*	the, uh, the party boat captain could not understand, you know,	+

10 *B.* he even, even he started baiting my hook <laughter>,
Statement-non-opinion
11 *A.* <laughter>. *Backchannel*
12 *B.* and holding, holding the, uh, the fishing rod. +
sw2290

Nevertheless, these two cases clearly cannot account for all the examples discussed in the literature (e.g., standalone uses of laughter as signal of disbelief or negative response to a polar question, Ginzburg et al., 2020) and above in Section 9.3.2. Future models will therefore require a manual assignment of meaningful dialogue acts to standalone laughs.

9.5 DISCUSSION

The implications of the results obtained are twofold: showing that laughter can help a computational model to attribute meaning to an utterance and help with pragmatic disambiguation, and consequently stressing once again the need for integrating laughter (and other non-verbal social signals) in any framework aimed to model meaning in interaction (Ginzburg et al., 2020; Maraev et al., 2018).

In addition to the findings of Chapter 8, our results provide further evidence for the fact that non-verbal behaviours are tightly related to the dialogue information structure, propositional content and dialogue act performed by utterances. Laughter, along with other non-verbal social signals, can constitute a dialogue act in itself conveying meaning and affecting the unfolding dialogue (Bavelas and Chovil, 2000; Ginzburg et al., 2020).

In this work we have shown that laughter is a valuable cue for the DAR task. We believe that in our conversations laughter is informative about interlocutors' emotional and cognitive appraisals of events and communicative intents. Therefore, it should not come as a surprise that laughter acts as a cue in a computational model.

On the question of laughter impact on the DAR task, this study found that laughter is more helpful in the SWDA corpus than in

AMI-DA. Due to the nature of interactions over the phone, SWDA dialogue participants cannot rely on visual signals, such as gestures and facial expressions. Our results support the hypothesis that in SWDA, vocalizations such as laughter are more pronounced and therefore more helpful in disambiguating dialogue acts. This may also explain why our best models perform better on SWDA: more of the information that interlocutors and dialogue act annotators rely on is present in SWDA transcripts, whereas AMI-DA annotators receive clear instructions to pay attention to the videos (AMI Project, 2005). This finding is consistent with that of Bavelas et al. (2008) who demonstrate that in face-to-face dialogue, visual components, such as gestures, can convey information that is independent from what is conveyed by speech.

Laughter can be used to mark the presence of an incongruity between what is said and what is intended, coined as *pragmatic incongruity* by Mazzocconi et al. (2020). In those cases laughter is especially valuable for disambiguating between literal and non-literal meaning, as we have shown for rhetorical questions, a task which is still a struggle for most NLP models and dialogue systems.

There is abundant room for further progress in determining how other speech-related information, such as prosody and disfluencies, can be incorporated into a DAR model. Stolcke et al. (2000) showed that dialogue acts can have specific prosodic manifestations that can be used to improve DAR. Furthermore, in relation to laughter, its form (duration, arousal, overlap with speech) can be informative about its function and position with respect to the laughable (Tian et al., 2016; Mazzocconi et al., 2020). Incorporating such information is crucial if models pre-trained on large-scale text corpora are to be adapted for use in dialogue applications.

10 Dialogue management with Linear logic

The content of this chapter integrates two papers: Maraev, Bernardy and Ginzburg (2020)[1] *and Maraev, Bernardy and Howes (2021)*[2].

10.1 YET ANOTHER DIALOGUE MANAGER?

A key aspect of dialogue systems design is the coherence of the system's responses. In this respect, a key component of a dialogue system is the dialogue manager, which selects appropriate system actions depending on the current state and the external context.

Although there has been a lot of development in dialogue systems in recent years, only a few approaches reflect advancements in *dialogue theory*. Our aim is to closely integrate dialogue systems with work in theoretical semantics and pragmatics of dialogue. This field has provided accounts for linguistic phenomena intrinsic to dialogue such as non-sentential utterances (Schlangen, 2003; Fernández et al., 2007; Ginzburg, 2012), clarification requests Purver, 2006; Ginzburg, 2012 and self-repair (Ginzburg et al., 2014;

[1]Vladislav Maraev, Jean-Philippe Bernardy and Jonathan Ginzburg (2020). 'Dialogue management with linear logic: the role of metavariables in questions and clarifications'. In: *Traitement Automatique des Langues (TAL)* 61.3, pp. 43–67. Maraev initiated and planned the study. Maraev and Bernardy (Bernardy did the main work in Sections 10.2–10.4) wrote the parts of the original paper that appear here with input from other authors.

[2]Vladislav Maraev, Jean-Philippe Bernardy and Christine Howes (2021). 'Non-humorous Use of Laughter in Spoken Dialogue Systems'. In: *Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2021)*, pp. 33–44. Maraev initiated and planned the study. The original manuscript was written by Maraev with input from Bernardy and Howes.

Hough and Purver, 2012), where the resolution is intuitively tied to the coherence of what is being said.

To this end, a formal and in particular a logical representation is instrumental. This chapter is concerned with the representation of participant states and transitions in a unified logical framework.

Even though the progress in bridging dialogue management and theoretical research of dialogue is promising, we believe that it is crucial to use formal tools which are appropriate for the task, so that the formalisation and implementation of dialogue semantics closely matches the mental picture that experts have. In the view of Dixon et al. (2009) this is best done by representing the information-state of the agents as updatable sets of propositions. Subsets of propositions in the information state can be treated independently, and, therefore, a suitable and flexible way to represent updates is as propositions in linear logic. We adopt this view here, and further argue for it in the body of the chapter.

We further extend the framework of Dixon et al. (2009) to deal with lack of clarity (and certain cases of non-probabilistic ambiguity). Indeed, asking a question is often not done in one utterance which leaves nothing to interpretation. Typically, in a conversation a question and its answer may be many utterances apart, and the intermediate utterances form insertion sequences (Schegloff, 1968), for instance a follow-up clarification request and a corresponding answer. The insertion sequences are, in turn, conditioned on the preliminaries for the original question (Levinson, 1983, Chapter 6). In this study, we deal with question-answering and clarification requests in a unified way, within a framework of dialogue management and using linear logic formalisation.

In line with linguistic approaches which use metavariables for underspecified elements of the context (e.g., Kempson et al., 2001), we propose here the use of *metavariables*, thereby leveraging much research on unification and proof search in various logical frameworks. That is, metavariables will stand in for any piece of information which is left to further interpretation. In particular, in this chapter we explore the potential of using metavariables in the representation of question-answer exchanges.

Larsson (2002) proposed the use of Prolog (and hence, proof search), as a dialogue management framework. However, the lack of linear hypotheses means that destructive information-state updates are sometimes awkward to represent. In addition, he does not consider the use of metavariables to represent uncertainty – even though Prolog in principle has the capacity to do it.

Using a solid logical basis (Bratko, 1986; Girard, 1995) which corresponds well with the intuition of information-state based dialogue management, we are able to provide a fully working prototype[3] of the components of our framework:

1. a new proof-search engine based on linear logic, modified to support inputs from external systems (representing inputs and outputs of the agent);

2. a set of rules which function as a core framework for dialogue management (in the style of KoS theoretical account). The rules which we present below are provided to this engine, in the same form (modulo typesetting);

3. several examples which use the above to construct potential applications of the system. The engine is able to run domain-specific rules and generic rules together, forming a working system.

10.2 PROOF SEARCH AS A PROGRAMMING LANGUAGE

The prevailing tradition in formal semantics, including in most pieces of work cited above, is to represent (declarative) statements as propositions, formalised in an underlying logic (often first-order logic). In particular, in linguistic theories based on intuitionistic logic (such as Theory of Types with Records (TTR), see Chapter 15), true statements corresponds to propositions which admit a proof.

There is a long history of using proof search as a declarative programming paradigm, where the programmer specifies *axioms*

[3]Source code and documentation are available at https://github.com/GU-CLASP/ProLin.

and *rules of inference* which model their application domain. Typically such a system of axioms and rules represents a database of facts. For example, the axiom $(\mathit{Leave}\ 55\ \mathit{Valand}\ 11.50)$ can model the fact that bus 55 leaves from Valand at 11:50. The rule $(\mathit{Leave}\ x\ \mathit{Valand}\ y \rightarrow \mathit{Arrive}\ x\ \mathit{CentralStationen}\ (y + 45\ \mathit{minutes}))$ can represent travelling times on a certain line.

Then, the user may define a query (or goal) as a logical formula. The system can then search for a proof of the goal as a way to query the database of facts. Often, goals contain *metavariables,*[4] which play the role of unknowns for unification: their value can be fixed to any term for a goal to be reached. For example, the goal $(\mathit{Leave}\ x\ \mathit{Valand}\ y)$ corresponds to a request to list all the buses leaving from Valand (as x) together with their departure time (as y).

Because statements are propositions, it is natural to use proof-search as a means to represent possible moves in dialogue seen as a game (Larsson, 2002).

10.3 LINEAR LOGIC AS A DM FRAMEWORK

Typically, and in particular in the archetypal logic programming language Prolog (Bratko, 1986), axioms and rules are expressed within the general framework of first-order logic. However, several authors (Dixon et al., 2009; Martens, 2015) have proposed to use linear logic (Girard, 1995) instead. For our purposes, the crucial feature of linear logic is that hypotheses may be used *only once.* For example, one could have a rule $\mathit{IsAt}\ x\ \mathit{Valand}\ y \multimap \mathit{IsAt}\ x\ \mathit{CentralStationen}\ (y + 45\ \mathit{minutes})$. Consequently, after firing the above rule, the premise $(\mathit{Is}\ x\ \mathit{Valand}\ y)$ becomes unavailable for any other rule. Thereby the linear arrow $\multimap$ can be used to conveniently model that a bus cannot be at two places simultaneously.[5]

[4] Here, we use the convention that metavariables start with a lowercase letter, and constants (including predicates) with an upper case.

[5] If several arrows are present in a rule (such as $A \multimap B \multimap C$) then both A and B are consumed and C is produced.

In general, the linear arrow corresponds to *destructive state updates*. Thus, the hypotheses available for proof search correspond to the *state* of the system. In our application they will correspond to the *information state* of the dialogue participant.[6]

This way, firing a linear rule corresponds to triggering an *action* of an agent, and a complete proof corresponds to a *scenario*, i.e. a sequence of actions, possibly involving action from several agents. Hence, the actions realised as actual interactions constitute the observable dialogue. That is, an action can result in sending a message to the outside world (in the form of speech, movement, etc.). Conversely, events happening in the outside world can result in updates of the information state (through a model of the perceptual subsystem).

At any point in the scenario, the multiset of available *linear hypotheses* represents the current information-state of the agent which is modelled. To clarify, the information-state (typically in the literature and in this thesis as well), corresponds to the state of a *single* agent. Thus, a scenario is conceived as a sequence of actions and updates of the information state of a single agent a, even though such actions can be attributed to any other dialogue participant b. (That is, they are a's representation of the actions of b.)

To reiterate, in our implementation, the information-state can be queried using *rules* (such as those we list below). Because they are linear, these hypotheses can also be removed from the state.

It is important to note that we will not forego the unrestricted (i.e. non-linear) implication ($\rightarrow$). Rather, both implications will co-exist in our implementation, thus we can represent simultaneously transient facts, or states (introduced by the linear arrow) and immutable facts (introduced by the unrestricted arrow). Besides, thus,

[6]We note, that in linear logic, facts (or hypotheses) do not come in a hierarchy. Either we have a fact, or we don't. However, in second-order variants of intuitionistic logic, like the one we use, one can conveniently wrap propositions in constructors, to indicate that they come with a qualification. For example, we can write *Unsure P* to indicate that the proposition P may hold (for example if clarification is required).

we have a *fixed* set of rules (which remain available even after being used), such as $IsAt\ x\ Valand\ y \multimap IsAt\ x\ CentralStationen\ (y + 45\ minutes)$ above. Each such rule manipulates a part of the information state (captured by its premises) and leaves everything else in the state unchanged.

10.4 QUESTION-ANSWERING WITH METAVARIABLES

In this subsection we show how a metavariable can represent what is being asked, as the unknown in a proposition. A first use for metavariables is to represent the requested answer of a question.

We represent a question by a predicate P over a type A. That is, using a typed intuitionistic logic:

$$A : Type$$
$$P : A \rightarrow Prop$$

The intent of the question is to find out about a value x of type A which makes $(P\ x)$ true, or at least entertained by the other participant. We provide several examples in Table 10.1. It is worth stressing that the type A can be large (for example asking for any location) or as small as a boolean (if one requires a simple yes/no answer). We note in passing that, typically, polar questions can be answered not just by a boolean but by qualifing the predicate in question, for example 'maybe', 'on Tuesdays', etc. (Table 10.1, last two rows). In this instance $A = Prop \rightarrow Prop$.

One complication is polar questions phrased in the negative (Cooper and Ginzburg, 2012); for example: 'Doesn't John like bananas?'. In this instance, a simple 'no' answer can be ambiguous, and a possible model would be a multi-valued kind of answer ('Yes he does.' represented as *DefiniteYes*; 'No he doesn't.', represented as *DefiniteNo*, 'No' as *AmbiguousNo*, and 'He does in the weekend.' as *Qualifier OnWeekend*):

$$\begin{array}{l} Q\ Multi\ (\lambda x.\mathbf{case}\ x\ \mathbf{of} \\ \qquad\qquad AmbiguousNo \rightarrow Trivial \\ \qquad\qquad DefiniteNo \quad\ \rightarrow \neg\ (Like\ John\ Bananas) \end{array}$$

question	A	P	reply	x
Where does John live?	*Location*	$\lambda x.\mathit{Live\ John\ x}$	in London	*ShortAnswer* *Location* *London*
Does John live in Paris?	*Bool*	$\lambda x.\mathbf{if}\ x$ **then** $(\mathit{Live\ John\ Paris})$ **else** $\mathit{Not}\ (\mathit{Live\ John\ Paris})$	yes	*ShortAnswer* *Bool* *True*
What time is it?	*Time*	$\lambda x.\mathit{IsTime\ x}$	It is 5am.	*Assert* $(\mathit{IsTime}\ 5.00)$
Does John live in Paris?	$\mathit{Prop} \to \mathit{Prop}$	$\lambda m.m\ (\mathit{Live\ John\ Paris})$	yes	*ShortAnswer* $(\mathit{Prop} \to \mathit{Prop})$ $(\lambda x.x)$
Does John live in Paris?	$\mathit{Prop} \to \mathit{Prop}$	$\lambda m.m\ (\mathit{Live\ John\ Paris})$	from January	*ShortAnswer* $(\mathit{Prop} \to \mathit{Prop})$ $(\lambda x.\mathit{FromJanuary}\ (x))$

Table 10.1: Examples of questions and the possible corresponding answers. The type A is the type of possible short answers. The proposition $P\ x$ is the interpretation of a short answer x. The x column shows the formal representation of a possible answer, either in short form or assertion form.

$$\begin{aligned} &\mathit{DefiniteYes} && \to \mathit{Like\ John\ Bananas} \\ &\mathit{Qualifier\ m} && \to m\ (\mathit{Like\ John\ Bananas})) \end{aligned}$$

To represent ambiguity in the case of *AmbiguousNo*, we make the answer provide no information, in the form of a trivial proposition (which is always true regardless of context). This is a natural account, because the meaning of short answers (such as 'no') always depends on the context. ('Paris' does not mean the same thing in the context of 'Where do you live?' as in the context 'Where were you born?'). Additionally, in the framework of a full dialogue management system, the *AmbiguousNo* case should be treated as not resolving the question (the question effectively remains unanswered). However, in such a framework, it is always possible to receive a biasing answer ('I don't know') or no answer whatsoever. Even more complications are possible, by the introduction of cases such as rhetorical questions. We deem such complications beyond the scope of the current implementation.

Within the state of the agent, if the value of the requested answer is represented as a metavariable x, then the question can be represented as: $Q\ A\ x\ (P\ x)$. That is, the pending question (Q denotes a question constructor) is a triple of a type, a metavariable x, and a proposition where x occurs. We stress that $P\ x$ is *not* part of the information state of the agent yet, rather the fact that the above question is the question under discussion (QUD) is a fact. For example, after asking 'Where does John live?', we have:

$$haveQud : QUD\ (Q\ Location\ x\ (Live\ John\ x))$$

Resolving a question can be done by communicating an answer. An answer to a question $(A : Type; P : A \rightarrow Prop)$ can be of either of the two following forms:

a ShortAnswer is a pair of an element $X : A$ and its type A, represented as $ShortAnswer\ A\ X$, or

an Assertion is a proposition $R : Prop$, represented as $Assert\ R$.

Therefore, one way to process a short answer is by the *processShort* rule:

$$\begin{aligned} processShort : (a : Type) \rightarrow (x : a) \rightarrow (p : Prop) \rightarrow \\ ShortAnswer\ a\ x \multimap QUD\ (Q\ a\ x\ p) \multimap p \end{aligned}$$

Above we use Π type binders to declare (meta)variables (written here $(a : Type) \rightarrow$, $(x : a) \rightarrow$, etc.). This terminology will make sense to readers familiar with dependent types. For others, such binders can be thought as universal quantification ($\forall a, \forall x$, etc.), the difference being that the type of the bound variable is specified.[7]

We demand in particular that types in the answer and in the question match (a occurs in both places). Additionally, because x occurs in p, the information state will mention the concrete x which was provided in the answer. For example, if the QUD was

[7]The reader worried about any theoretical difficulty regarding mixing linear and dependent types is directed to Atkey (2018) and Abel and Bernardy (2020).

$(Q\ Location\ x\ (Live\ John\ x))$ and the system processes the answer *ShortAnswer Location Paris*, then x unifies with *Paris*, and the new state will include *Live John Paris*.

To process assertions, we can use the following rule:

$$\begin{aligned} processAssert : (a : Type) \rightarrow (x : a) \rightarrow (p : Prop) \rightarrow \\ Assert\ p \multimap QUD\ (Q\ a\ x\ p) \multimap p \end{aligned}$$

That is, if i) p was asserted, and ii) the proposition q is part of a question under discussion, and iii) p can be unified with q (we ensure this unification by simply using the same metavariable p in both roles in the above rule), then the assertion resolves the question. Additionally, the metavariable x is made ground to a value provided by p, by virtue of unification of p and q. For example, 'John lives in Paris' answers both questions 'Where does John live?' and 'Does John live in Paris?' (there is unification), but, not, for example 'What time is it?' (there is no unification). Note that, in both cases (*processAssert* and *processShort*), the information state is updated with the proposition posed in the question.

10.5 LINEAR DIALOGUE MANAGER

In this section we integrate our question answering framework within more a complete dialogue manager (DM). We call it Linear Dialogue Manager (LDM). We stress that LDM models the information-state of only one participant. Regardless, this participant can record its own beliefs about the state of other participants. Figure 10.1 shows how LDM can be integrated into a spoken dialogue system. In general, the core of LDM is comprised of a set of linear-logic rules which depend on the domain of application. However, many rules will be domain-independent (such as generic processing of answers). We present these rules in the following subsections.

10.5.1 Interface with language understanding and generation

To be useful, a DM must interact with the outside world, and this interaction cannot be represented using logical rules, which can

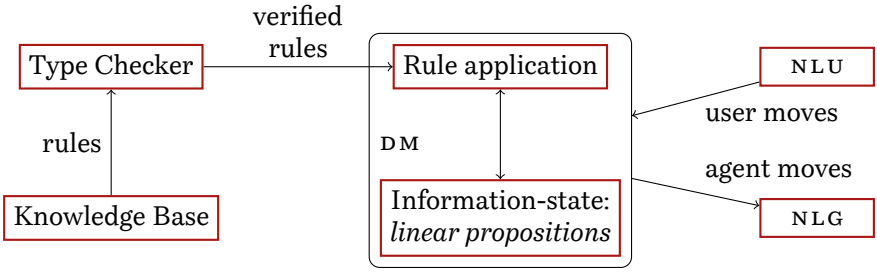


Figure 10.1: Architecture of a spoken dialogue system with a dialogue manager based on a linear logic framework.

only manipulate data which is already integrated in the information state. Here, we assume that the information that comes from sources which are external to the dialogue manager is expressed in terms of semantic interpretations of moves, and contains information about the speaker and the addressee in a structured way. We provide 5 basic types of moves, specified with a speaker and an addressee, as an illustration:

Greet	*spkr addr*
CounterGreet	*spkr addr*
Ask	*question spkr addr*
ShortAnswer	*vtype v spkr addr*
Assert	*p spkr addr*

These moves can either be received as input or produced as outputs. If they are inputs, they come from the natural language understanding (NLU) component, and they enter the context with $Heard : Move \rightarrow Prop$ predicate. For example, if one hears a greeting, the proposition $Heard\ (Greet\ S\ A)$ is added to the information state/context, without any rule being fired – this is what we mean by an external source.

If they are outputs, to be further used by the natural language generation (NLG) component, some rule will place them in *Agenda*. For example, to issue a countergreeting, a rule will place the proposition $Agenda\ (CounterGreet\ A\ S)$ in the information state.

Thereby each move is accompanied by the information about who has uttered it, and towards whom it was addressed. All the moves are recorded in the *Moves* part of the participant's dialogue gameboard, as a stack, represented as a *Cons*-list in our implementation.

Additionally, we record any move *m* which one has yet to actively react to, in a hypothesis of the form *Pending m*. We cannot use the *Moves* part of the state for this purpose, because it is meant to be static (not to be consumed). *Pending* thus allows one to make the difference between a move which is fully processed and a pending one.

10.5.2 Initial state

In general, we start with empty *QUD* and *Agenda*. A non-empty *QUD* can be prepared if in a certain conversational genre (Ginzburg, 2012) some open questions are assumed from the start. The *Agenda* might not be empty if one wants the system to initiate the conversation. There are also no moves: nothing has been said by either party.

$$\begin{array}{ll} \_ :: QUD & Nil \\ \_ :: Agenda & Nil \\ \_ :: Moves & Nil \end{array}$$

(We often do not care about the proof object witnessing a propositions, in which case we denote it with an underscore.)

10.5.3 Hearing

The capacity of 'hearing' or, in other words, starting the processing of semantic representations of utterances from the NLU component is implemented with the following rule:

$$\begin{array}{l} hearAndRemember : \\ \quad (m : DP \to DP \to Move) \to \\ \quad (x\ y : DP) \qquad\qquad\quad \to \\ \quad (ms : List\ Move) \qquad\quad \to \\ \qquad Heard\ (m\ x\ y) \multimap Moves\ ms \multimap HasTurn\ x \multimap \end{array}$$

$$
\begin{array}{l}
[\_ :: \mathit{Moves}\ (\mathit{Cons}\ (m\ x\ y)\ \mathit{ms});\\
\ \_ :: \mathit{Pending}\ (m\ x\ y);\\
\ \_ :: \mathit{HasTurn}\ y]
\end{array}
$$

where $(m\ x\ y)$ is a semantic representation of the utterance. Here we produce a record, whose fields will all be added to the information state. The rule demands that participant x has the turn and, as a result, turn was taken by his partner y.[8] The *DP* type stands for *dialogue participant*. As a result we do several things: i) place the move in a move list for further references (*PushMove*), ii) record the turn-switching (which in a complete system may not apply to all cases – then additional hypotheses would be added), and iii) prepare to process the move (*Pending*).

10.5.4 Uttering

The capacity of 'uttering' represents an ability to generate information for the NLG component. The NLG component is represented by the *Agenda* that contains a move that is just about to be uttered.

$$
\begin{array}{l}
\mathit{utterAndRemember}:\\
\quad (m : \mathit{DP} \to \mathit{DP} \to \mathit{Move}) \to\\
\quad (\mathit{ms} : \mathit{List\ Move}) \qquad\qquad \to\\
\quad (x\ y : \mathit{DP}) \qquad\qquad\qquad \to\\
\qquad \mathit{Agenda}\ (m\ x\ y) \multimap \mathit{Moves}\ \mathit{ms} \multimap \mathit{HasTurn}\ x \multimap\\
\qquad\quad [\_ :: \mathit{Utter}\ (m\ x\ y);\\
\qquad\quad \ \_ :: \mathit{Moves}\ (\mathit{Cons}\ (m\ x\ y)\ \mathit{ms});\\
\qquad\quad \ \_ :: \mathit{HasTurn}\ y]
\end{array}
$$

Here also we take care of turn-taking in the same rule. As a result, the system consumes the *Agenda* and passes the move to the NLG component. The move is also memorised in the *Moves* stack.

[8] For now we have a very simple model of turn-taking, which can be improved in many ways: certain moves may not induce turn-change, there can be more than two participants, etc.

10.5.5 *Basic adjacency: greeting*

We can show how basic move adjacency can be defined in the example of countergreeting preconditioned by a greeting from the other party:

$$
\begin{array}{l}
\mathit{counterGreeting} : \\
\quad (x\ y : \mathit{DP}) \rightarrow \\
\quad\quad \mathit{HasTurn}\ x \multimapinv \mathit{Pending}\ (\mathit{Greet}\ y\ x) \multimap \\
\quad\quad\quad \mathit{Agenda}\ (\mathit{CounterGreet}\ x\ y)
\end{array}
$$

10.5.6 QUD *incrementation*

Another important rule accounts for pushing the content of the last move. In KoS both statements and queries induce the QUD to increment; statements transform a proposition to a question under discussion. In the case where the latest move is *Ask*, it's content can be put on top of the questions under discussion (*QUD*) stack:

$$
\begin{array}{l}
\mathit{pushQUD} : \\
\quad (q : \mathit{Question}) \quad\quad\quad \rightarrow \\
\quad (\mathit{qs} : \mathit{List}\ \mathit{Question}) \rightarrow \\
\quad (x\ y : \mathit{DP}) \quad\quad\quad\quad \rightarrow \\
\quad\quad \mathit{Pending}\ (\mathit{Ask}\ q\ x\ y) \multimap \mathit{QUD}\ \mathit{qs} \multimap \\
\quad\quad\quad \mathit{QUD}\ (\mathit{Cons}\ q\ \mathit{qs})
\end{array}
$$

10.5.7 *Integrating the answers*

If the user asserts something that relates to the top *QUD*, then the *QUD* can be resolved and therefore removed from the stack. The corresponding proposition p is saved as a *UserFact*.[9] This rule extends the abstract rule that was introduced in Section 10.4.

$$
\begin{array}{l}
\mathit{processAssert} : \\
\quad (a : \mathit{Type}) \rightarrow (x : a) \rightarrow (p : \mathit{Prop}) \rightarrow
\end{array}
$$

[9] For the current purposes we only remove the top QUD, but in a more general case we can implement the policy that can potentially resolve any QUD from the stack.

$$
\begin{array}{l}
(qs : List\ Question) \rightarrow \\
(dp\ dp1 : DP) \qquad \rightarrow \\
\quad Pending\ (Assert\ p\ dp1\ dp) \qquad \multimap \\
\quad QUD \quad (Cons\ (Q\ dp\ a\ x\ p)\ qs) \multimap \\
\qquad [\_ :: UserFact\ p; \\
\qquad\ \ \_ :: QUD\ qs]
\end{array}
$$

Short answers are processed in a very similar way to assertions:

$$
\begin{array}{l}
processShort : \\
\quad (a : Type) \rightarrow (x : a) \rightarrow (p : Prop) \rightarrow \\
\quad (qs : List\ Question) \rightarrow \\
\quad (dp\ dp1 : DP) \qquad \rightarrow \\
\quad\ \ Pending\ (ShortAnswer\ a\ x\ dp1\ dp) \multimap \\
\quad\ \ QUD \quad (Cons\ (Q\ dp\ a\ x\ p)\ qs) \qquad \multimap \\
\qquad [\_ :: UserFact\ p; \\
\qquad\ \ \_ :: QUD\ qs]
\end{array}
$$

10.5.8 Questions and clarifications

We use a uniqueness check (operator $\rightarrow_!$ checks whether metavariable is unique and made ground, $\rightarrow_?$ checks for solutions which are non-unique or cannot be made fully ground; see Maraev, Bernardy and Ginzburg, 2020) to determine whether the system can resolve the question (*produceAnswer*) or needs to initiate a clarifying side sequence (*produceCR*).

$$
\begin{array}{l}
produceAnswer : \\
\quad (a : Type) \rightarrow (x : a) \rightarrow_! (p : Prop) \rightarrow \\
\quad (qs : List\ Question) \rightarrow \\
\quad\ \ QUD\ (Cons\ (Q\ USER\ a\ x\ p)\ qs) \multimap \\
\quad\ \ p \multimapinv \\
\qquad [\_ :: Agenda\ (ShortAnswer\ a\ x\ SYSTEM\ \ USER); \\
\qquad\ \ \_ :: QUD\ qs; \\
\qquad\ \ \_ :: Answered\ (Q\ USER\ a\ x\ p)]
\end{array}
$$

$$
\begin{array}{l}
produceCR : \\
\quad [a : Type; x : a; p : Prop; qs : List\ Question; \\
\quad\ \ \_ :: QUD\ (Cons\ (Q\ USER\ a\ x\ p)\ qs); \_ :: p] \rightarrow_?\ CR
\end{array}
$$

The clarifying side sequence itself (*CR*) is meant to be specified by a dialogue developer, possibly informed by machine-learning systems, because it is domain-specific and the choice of the spectrum of possible options is wide.

10.5.9 Grounding

Grounding is a major issue to be tackled in a DM due to the importance of dealing with the uncertainty about the speaker's intention, as discussed in Section 4.3, and, additionally, of NLU and automatic speech recognition (ASR) modules. In this subsection we provide a sketch of basic grounding strategies and moves related to them, which will be further used to model laughter in Chapter 11.

Dialogue systems deal with confidence scores from the ASR and NLU components, which reflect the uncertainty in user queries. For simplicity we will represent the confidence score t on the basis of three confidence threshold levels ($T_1 < T_2$), where *RED* corresponds to $t < T_1$, *YELLOW* to $T_1 < t < T_2$, and *GREEN* to $T_2 < t$. Colour-coded confidence scores would accompany user moves, e.g., the *Ask* move such as 'What time is it?' can be represented as follows:

Ask (*Q U Time t0* (*IsTime t0*)) *U S YELLOW*

Here we illustrate the possibility of extending the system with Interactive Communication Management (ICM) moves and grounding strategies, replicating Larsson's (2002) account for grounding and feedback. ICM moves are used for coordination of the common ground in dialogue, which expresses, for instance, explicit signals for integrating the incoming information and updating common ground (dialogue gameboard in our implementation). The basic type for the ICM move is the following:

ICM level polarity content

where *level* corresponds to the level of grounding, according to the *action ladder* model discussed in Section 4.3, (contact, perception, understanding, acceptance), *polarity* is either positive or

negative, and the optional value *content* corresponds to a component of the common ground in question. For instance, the move (*ICM Per Neg None*) would correspond to the utterance 'I didn't understand what you said' or 'Pardon', and the move (*ICM Und Pos q*) can be realised as the utterance 'You are asking me what time it is' if the QUD *q* corresponds to the question from the *Ask* move exemplified above.

Next, we modify our basic *pushQUD* rule defined in Section 10.5.6 to support different system behaviours depending on the confidence score. In the *GREEN* case, a question from the user *Ask* move is integrated into *QUD*, and the ICM move displaying positive acceptance feedback, i.e. 'okay', (*ICM Acc Pos None*) is put on the *Agenda*. In the *YELLOW* case, the system should additionally report positive understanding, e.g., 'You want to know about time', so add a (*ICM Und Pos q*) move to the *Agenda*.

```
pushQUDGreen :
  (q : Question)        →
  (qs : List Question) →
  (x y : DP)            →
    Pending (Ask q x y GREEN) ⊸
    Agenda as ⊸
    QUD qs    ⊸
      [_ :: QUD (Cons q qs);
       _ :: Agenda (Cons (ICM Acc Pos None) as); ]

pushQUDYellow :
  (q : Question)        →
  (qs : List Question) →
  (x y : DP)            →
    Pending (Ask q x y YELLOW) ⊸
    Agenda as ⊸
    QUD qs    ⊸
      [_ :: QUD (Cons q qs);
       _ :: Agenda (Cons (ICM Acc Pos None)
          (Cons (ICM Und Pos q) as)); ]
```

For *RED* confidence score, the system issues an interrogative ICM query, such as 'I understood you're asking me about the time, is that correct?'. In this case a special type of *QUD* is introduced, namely a question about whether question q is correctly understood.

```
icmINTConfirm :
  (q : Question) →
  (x y : DP)     →
    Pending (Ask q x y RED) ⊸
    Agenda as ⊸
    QUD qs    ⊸
      [_ :: QUD (Cons (Q Bool x
          (      if x then UND q
                 else UNDN q)) qs);
       _ :: Agenda (Cons (ICM Und Int q) as)]
```

Processing answers related to such a type of *QUD* will be done as usual. For instance, a short 'yes' or 'no' will be treated here as a boolean, and depending on the answer the context will contain either *PendingUserFact* (*UND* q) or *PendingUserFact* (*UNDN* q).

Regardless of the particular answer, once the ICM question is answered, it is removed from the *QUD* stack, so that the *QUD* stack is restored to the originally asked question. In our system, this is taken care of by the generic handling of *ShortAnswer*s. Thus, in the case of a positive answer to such a query, there is nothing particular to do.

In the negative case, the ICM move about understanding that the question was not q is issued.

```
icmINTneg :
  (q : Question)   →
  (x y : DP)       →
  (c : Confidence) →
    PendingUserFact (UNDN q) ⊸
    Agenda as ⊸
    Agenda (Cons (ICM Und Pos (QuestionIsNot q)) as)
```

How ICM moves are converted to natural language utterances, depending on q, is a NLG issue. For instance,

$$
\begin{array}{l}
\mathit{ICM\ Und\ Pos} \\
\quad (\mathit{QuestionIsNot} \\
\qquad (\mathit{Q\ U\ Time\ t0}\ (\mathit{IsTime\ t0})))
\end{array}
$$

can become the (rather tedious) utterance 'So, you are not asking me what time it is', whereas more sophisticated queries with more arguments can be resolved with a shorter utterance depending on the arguments that are *made ground* (see 10.4). For instance, in the context of an interaction at a food kiosk:

$$
\begin{array}{l}
\mathit{ICM\ Und\ Pos} \\
\quad (\mathit{QuestionIsNot} \\
\qquad (\mathit{Q\ U}\ (\mathit{Prop} \rightarrow \mathit{Prop})\ \mathit{m0}\ (\mathit{m0\ WantOlives}))
\end{array}
$$

could become a simple 'Sorry, let's forget olives.'

10.6 SUMMARY

This chapter has described the method used for designing a DM which we then further extend to support different kinds of laughter (Chapter 11) and the rhetorical resources that can be used for humour (Chapter 16). We have explained the main conceptual decisions behind the rules related to grounding and question answering, and given a few examples of such rules. We regard further explanations of our LDM framework as tangential to the main scope of the thesis and refer our readers to our previous publication (Maraev, Bernardy and Ginzburg, 2020), which contains more examples and further discussions.

This chapter lays the groundwork for treating laughter within a DM uniformly with other moves and actions. In the next Chapter we treat laughter as such a move and see how it can be integrated into the dialogue state using the LDM framework.

11 Non-humorous laughter in dialogue management

The content of this chapter was previously published in Maraev, Bernardy and Howes (2021)[1]*. It has been substantially revised.*

11.1 INTRODUCTION

In this chapter we consider laughter from a utilitarian perspective, and attempt to determine which kinds of laughs can be relevant for dialogue systems.

As discussed in Section 5.4, there have been attempts to produce laughs as a way to mimic human behaviour and align with it, as well as laughing avatars mainly focussed on laughter as a reaction to jokes. Here we take a rather different approach. We start from examples of usage of laughter in real task-oriented dialogue and then propose ways in which these behaviours can be reproduced in a dialogue system, and, more specifically, in its dialogue management component.

The example (11a) below is an excerpt from a role-play dialogue collected by Howes et al. (2019) for their Directory Enquiries Corpus (DEC) (Bondarenko et al., 2020). Dialogue participants were playing the roles of a caller and an operator, respectively asking for the phone numbers of certain named businesses. Half of the dialogues

[1]Vladislav Maraev, Jean-Philippe Bernardy and Christine Howes (2021). 'Non-humorous Use of Laughter in Spoken Dialogue Systems'. In: *Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2021)*, pp. 33–44. Maraev initiated and planned the study. The original manuscript was written by Maraev with input from Bernardy and Howes.

happened in a noisy environment, with many mishearings and laughs induced.

(11a)
56 *Caller.* er the next one is er tanfield chambers
57 *Operator.* santias?
58 *Caller.* tanfield like t- T A N
59 *Operator.* sorry i don't hear you again please?
60 *Caller.* er T A N
61 *Operator.* C?
62 *Caller.* tanfield
63 *Operator.* A
64 *Operator.* N
65 *Caller.* yeah
66 *Caller.* and then field
67 *Operator.* and then seal?
68 *Caller.* chambers
69 *Operator.* <laugh> sorry i hear you quite poorly
70 *Operator.* let's try again
71 *Operator.* C?
72 *Caller.* yeah sorry the traffic is crazy around here
73 *Operator.* I know <laugh> don't worry
74 *Operator.* so C
75 *Operator.* A
76 *Caller.* er
77 *Caller.* tanfield T like thomas

DEC 22_KL_loc2

Let's look at the first laughter (line 69). We can see that the operator's question 'and then seal?' (line 67) was not addressed and this piece of information was not grounded. 'C?' (line 71) refers to the restart from the beginning (it was 'Tanfield', but she has heard an initial 'C'). The negative feedback provided by the operator (line 69) entails extra effort from the caller – she needs to restart her request from the beginning – this obligation is somewhat intrusive and may require extra smoothing (see Section 9.3.3 and Mazzocconi,

2019; Raclaw and Ford, 2017). For our purposes we will treat this laughter as accompanying negative feedback.

For a dialogue system designer this poses an empirical question, namely, would it be useful to soften negative feedback with laughter? For instance, feedback associated with a local failure (e.g., a speech recognition failure), such as 'Sorry, I didn't understand' or 'Sorry I didn't hear you'. It may also be useful where negative feedback is the result of an external query, for example when something is not found in the database, and can accompany a system request to start over, as in example (11a).

As we discussed in Chapter 9, the reaction to an apology can also be accompanied by laughter, as with the second laugh in (11a) on line 73. We do not think that users often apologise to a dialogue system, as it is usually the dialogue system which is at fault, but this might be different for special cases of systems that aim at more naturalistic behaviour.

This chapter addresses the following research question: how can these laughs be accounted for in a task-oriented dialogue system?

This chapter is organised as follows. Section 11.2 presents a small typology of laughter types which we think should be accounted for in task-oriented dialogue systems. In Section 11.3 we show a formal account for the aforementioned types of laughter. We conclude with a brief discussion of our findings and further laughter-related issues in Section 11.4.

11.2 TYPES OF LAUGHTER

In this section we outline some types of laughter that can be of special interest to task-oriented dialogue systems and can be accounted for within our proposed framework.

As we have mentioned in Section 10.5.9, in accord with Allwood (1995), Clark (1996) and Larsson (2002) we consider four action levels that are involved in dialogue. Here we discuss what can happen at each level of action – contact, perception, understanding and reaction – with respect to laughter.

11.2.1 Contact and perception levels

Troubles related to establishing and maintaining a stable communication channel can lead to laughter. One such example would be delays in communication, for instance over an unreliable network, which might lead to a person already speaking at the moment when the communication is only supposed to be established. Obvious examples of such cases are caused by signal jitter over video conference platforms like Zoom.

Lack of perception basically indicates things that haven't been heard correctly (cases similar to (11a)). Also, it seems that interruptions, like sudden noise or echo, and events related to them might be quite surprising and laughter would be a natural reaction (see Section 11.4 of laughter related to surprise).

11.2.2 Understanding level

The lack of pragmatic understanding relates to the kinds of incongruities that are caused by the violation of the principle of conversational relevance. This is very useful for dialogue systems, because they are prone to errors in this realm. It is often the case that incorrect natural language understanding (NLU) or automatic speech recognition (ASR) can lead to prioritising irrelevant results (for example, in cases of out-of-scope user queries), which can cause user confusion and, therefore, laughter. This type of laughter can be treated as negative feedback, and (11b) and (11c) exemplify such uses of laughter.

(11b) From a dialogue between a virtual assistant (Diana) and a person with autism spectrum disorder (Mark). The audience – other people remotely present at their conversation over Zoom – react to irrelevance of Diana's reply with laughter.

1 *Mark.* Diana, what is money?
2 *Diana.* I am Diana, a virtual interlocutor.
3 *Audience.* <laughter>

https://youtu.be/6RN4B5Z88TQ at 2:10

(11c) 1 *Brian.* Would you like tea or coffee?

2 *Katie.* yes
3 *Brian.* <laughter>

constructed example

A dialogue system can also be unsure about what has been understood. In such cases, the system should demonstrate a lower degree of commitment to what has been said as a part of a display of understanding. For example, in case of feedback regarding the user input, when the system repeats the input after the user, it can be useful to include laughter in verbatim repeats, which would mean: yes, I understood this, but I might be wrong. This can also be useful for a system's actions taken based on low confidence results. The example in (11d) demonstrates this in a natural interaction – Dbillon's laughter accompanies a reprise clarification request for the fragment of Danny's utterance. Even though we are aware of the obscene essence of this fragment, it still has a form of a clarifying feedback and can be treated this way.

(11d) 1231 *Danny.* Guess what they they've nick-named me? write on all your lockers like they re-arrange your name.
1232 *Dbillon.* Oh, do they?
1233 *Danny.* They put on mine pussy lips
1234 *Dbillon.* Pussy lips? <laughter> why's that?
1235 *Danny.* Well, instead of [last or full name] they put pussy lips
1236 *Dbillon.* <laughter> Well, it's a sad thing if you can't laugh at yourself isn't it, eh?
1237 *Danny.* I have I haven't scrubbed it off or anything cos it is quite funny

BNC KPA

11.2.3 Reaction level

On this level what has been understood can be either accepted or rejected for the current purpose. Acceptance laughter can typically be related to a reaction to humour or an apology (see next section).

Laughter can also serve as a positive answer to a question, like in (11e), line 339. Here the 2 year old child confirms that she was actually making a joke. This example additionally stresses that laughter is highly ambiguous and one needs to look at different aspects of the context to correctly interpret it.

(11e) from Mazzocconi and Ginzburg (2022)

329 *Mum.* Did you help Daddy make the coffee?
330 *Mum.* Where did you make the coffee?
331 *Child.* Tea
332 *Mum.* Tea? there was no tea!
333 *Mum.* did you make the coffee in the bathroom?
334 *Mum.* no! Where did you make the coffee?
335 *Mum.* where did you make the coffee this morning?
336 *Child.* <smiling> upstairs
337 *Mum.* upstairs!?
338 *Mum.* That's a joke, right?
339 *Child.* <laughter/>
340 *Mum.* <laughter> yeah </laughter> you're making a joke!
341 *Mum.* you know that coffee +…
342 *Mum.* there's no kitchen upstairs!
343 *Child.* listen Mommy
344 *Mum.* what?
345 *Child.* nursie Daddy
346 *Mum.* nursie Daddy, that's another joke!
347 *Mum.* you're being funny now, oh, no, we do not draw on clothes

Providence Corpus (Demuth et al., 2006) – Naima 02004

Ginzburg et al. (2020) consider some uses of standalone laughter as cases of a negative response to a polar question (11f) or a signal of disbelief in a previously uttered assertion.

(11f) from Ginzburg et al. (2020), 'Context: Bayern München goalkeeper Manuel Neuer faces the press after his team's (*Dreierkette* [three-in-the-back]) defence has proved highly prob-

lematic in the game just played (which they won 3-2 against Paderborn)' (ibid.)

1	*Journalist.*	(smile) Dreierkette auch 'ne Option? (Is the three-at-the-back also an option?)
2	*Manuel Neuer.*	fuh fuh fuh (brief laugh)

In Section 11.3 we show how this kind of laughter as negative response like (11f) can be handled by the dialogue manager.

11.2.4 Laughter and intrusion

As we discussed in relation to example (11a), in natural dialogue, intrusion is frequently associated with laughter. In (11g), the caller reacts with a compassionate laughter to the apology given by the operator. The second laugh in (11a) might be interpreted by the operator in a similar way.

(11g)

162	*Operator.*	still not finding it
163	*Operator.*	having problems with this one
164	*Caller.*	okay
165	*Caller.*	er maybe i can find
166	*Caller.*	er the place myself but thank you very much for the information
167	*Operator.*	no problem <u>sorry for not finding the the last one</u>
168	*Caller.*	<laugh>
169	*Caller.*	no worries
170	*Caller.*	thank you

DEC 16_HG_loc2

We also observe that laughter can clearly accompany asking for a favour by the same speaker. In example (11h) the operator asks the caller if they can start from the beginning, which can be treated as an intrusion of some sort[2], therefore asking for a favour, and the apology is accompanied by laughter.

[2]In Politeness theory (see Section 9.3.3) this can be treated as face-threatening act (FTA) threatening the negative face of the listener.

(11h)

59	*Caller*	B as in bicycle
60	*Operator*	yeah
61	*Caller*	then you have R
62	*Caller*	I
63	*Operator*	R
64	*Caller*	G
65	*Operator*	I
66	*Operator*	okay sorry no- now i lost the track okay can we it start from the beginning <laugh> sorry
67	*Caller*	okay
68	*Caller*	yes we can
69	*Operator*	maybe you can just say the uh say words
70	*Caller*	yeah no no problem

DEC 24_LK_loc2

11.3 FORMAL TREATMENT OF NON-HUMOROUS LAUGHTER

We base our formal approach on the Linear Dialogue Manager (LDM) framework which we described in detail in Chapter 10. Overall, our approach benefits from the treatment of grounding and special Interactive Communication Management (ICM) moves in a domain-general way.

11.3.1 Laughter as an answer

Laughter as a negative reaction to interrogative feedback in the case of a low confidence ASR or NLU result can be illustrated by the following dialogue.

(11i)

1	*User*	I would like to order a vegan bean burger.	*Ask*
2	*System*	I understood you'd like to order a beef burger. Is that correct?	*ICM Und Int*
3	*User*	<laughter>	*ShortAnswer Bool False*

constructed example

Here we can treat laughter as a short negative answer, similar to 'No'. In the case of the interrogative ICM move, such an answer can be processed using the *icmINTneg* rule defined in Section 10.5.9.

This can be treated as a recovery strategy for different dialogue system's reactions that were not intended by its designers. This approach can be extended to other cases of user feedback, for instance, to cover the cases with higher confidence score where the system produces *ICM Und Pos q* move.

In the more sophisticated example in (11f), Manuel Neuer's laughter can be handled by our generic rules for integrating questions under discussion (QUDS) – *pushQUD*. For this purpose we need to consider polar questions as expecting an answer of *Prop* → *Prop* type (see Table 10.1). Recalling the example:

(11f)	1	*Journalist.*	(smile) Dreierkette auch 'ne Option? (Is the three-at-the-back also an option?)
	2	*Manuel Neuer.*	fuh fuh fuh (brief laugh)

Ginzburg et al. (2020)

and a type for a question, from Section 10.4:

$$A : Type$$
$$P : A \rightarrow Prop$$

in this case,

$$A = Prop \rightarrow Prop$$
$$P = \lambda m.m\ IsOptionDreierkette$$

The brief laughter by Manuel Neuer can be represented as:

$$[\![\, fuhfuhfuh]\!] = ShortAnswer$$
$$(Prop \rightarrow Prop)\ (\lambda x.Laughable\ x)$$

where the modification of the proposition, resulting in

$$Laughable\ IsOptionDreierkette$$

has a very basic meaning: this proposition *is the laughable* (see Section 3.3), without being more specific about the laughter function. One can also consider being more specific, simply treating the laughter as a negation

$$\mathit{ShortAnswer}\ (\mathit{Prop} \rightarrow \mathit{Prop})\ (\lambda x.\mathit{Not}\ x),$$

which is appropriate in the case of Neuer's laughter, but in the general case laughter has a more nuanced meaning, including exactly the opposite – a positive answer as in (11e). For such cases it is more useful to keep laughter as a modifier of a proposition and disambiguate it using other aspects of the context.

Note that besides treating laughter as an answer, it can act as a device to reject a question – or, in the *action ladder* terms (see Section 4.3), decline the possibility of acceptance of a question. In this case the current QUD ought to be removed, which can be implemented as an alternative reaction to laughter. Overall, the decision to treat laughter as either an answer or as a rejection signal depends on the form of laughter (Section 3.5), as well as accompanying multimodal context which includes pose, gestures and gaze (Chapter 8).

11.3.2 Laughter which accompanies feedback

Laughter can act as a part of the realisation of ICM moves performed by natural language generation (NLG) component. It seems to us that in particular ICM moves the use of laughter can be considered 'safe'. For instance, the system's ICM move of the form

$$\begin{aligned}&\mathit{ICM\ Und\ Pos}\ (\mathit{QuestionIsNot}\\ &\quad(\mathit{Q\ U}\ (\mathit{Prop} \rightarrow \mathit{Prop})\ \mathit{m0}\ (\mathit{m0\ WantOlives}))\end{aligned}$$

(the system shows understanding that it wasn't the olives that the user was asking about) can be realised as a natural language utterance like 'Okay, let's forget olives, hehe', whereas laughter is used as a smoothing device to mitigate the awkwardness of system failure. Larsson (2002) often included an apology 'Sorry' in some of the

ICM moves, e.g., 'Sorry, I didn't understand that'. With some possible caveats, we can sometimes include a slight laughter in such moves, especially if system is getting a bit repetitive and produces (*ICM Und Neg*) too often. Considering the evidence for laughter often accompanying apology (as a separate dialogue act) presented in Section 9.3, this can be regarded as mimicking natural behaviour in dialogue.

11.4 DISCUSSION

In this chapter we have shown how some types of laughter can be accounted for in a task-oriented spoken dialogue system. In Chapter 10 we proposed our own proof-theoretic architecture of a dialogue manager based on the KoS framework and extended it with some grounding strategies. Based on this, we have shown how certain types of laughter can be processed within the dialogue manager and natural language generator, namely: laughter as negative feedback, laughter as a negative answer to a polar question and laughter as a signal accompanying system feedback.

In the following subsections we discuss several issues related to laughter for future work, but only merely touching the main subject of this chapter.

11.4.1 Surprise

Intuitively, laughter is related to events that are not expected in interaction. One of the ways to establish some degree of natural behaviour for a dialogue system would be to react sincerely to these kind of surprising events. A possible measure for a system's surprisal is how confused it is by the user input. A natural measure for this from information theory is *perplexity*, a probability-based metric. For N words in an evaluation set $W = w_1 w_2 \ldots w_N$, the average perplexity per word is computed as follows:

$$PP(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i \mid w_1 \ldots w_{i-1})}} \tag{11.1}$$

Given a language model, we can employ a threshold defined by perplexity which the system can use to act as being surprised, e.g., by saying 'Ha-ha, I did not expect this!'

Similarly, perplexity can be inferred from tracking a dialogue state in a Dialogue State Tracking task (Mrkšić et al., 2017), which is a common task in statistical approaches to dialogue systems. Or, following Noble and Maraev (2021), the recurrent neural network (RNN) trained on a large dialogue corpus as a representation of dialogue context (see Section 9.4.1) can be used to calculate perplexity.

Laughter as a reaction of surprise can relate to the levels of feedback, for example, a user surprised by a pragmatically incoherent system's reply can laugh (Section 11.3.1). But here surprise is taken in isolation, as a measure in its own right.

11.4.2 Awkwardness and time-saving

In (11j), 'under' is produced after a long pause (line 26) and therefore indicates awkwardness in producing the phonetic spelling, which made the operator wait – therefore making the situation uncomfortable to the caller, so laughter was used for smoothing it. We will come back to this example to illustrate humorous incongruities in Chapter 16.

(11j)

17	*Caller.*	okay so it starts with a
18	*Caller.*	L
19	*Operator.*	L?
20	*Caller.*	as in london
21	*Operator.*	yes
22	*Caller.*	A as in america
23	*Operator.*	america
24	*Caller.*	er U
25	*Caller.*	as in er ((pause: 1.2s))
26	*Caller.*	er under
27	*Caller.*	<laugh>
28	*Operator.*	under yes

DEC 28_NM_loc2

In the follow-up excerpt (11k) from the same dialogue, the user's awkwardness continues and she accompanies it with laughter. Firstly, she laughs (line 139) demonstrating that she has given up finding any phonetic spelling for 'K', giving the opportunity to the operator to continue. Her second laugh smooths her slight embarrassment after the situation was resolved by the operator.

(11k)

134	*Caller.*	O for oslo
135	*Operator.*	O for oslo
136	*Caller.*	again O for oslo
137	*Operator.*	O for oslo
138	*Caller.*	and K for er ((pause: 1.6s))
139	*Caller.*	<laugh>
140	*Operator.*	as in king?
141	*Caller.*	k- king <laugh> yeah
142	*Operator.*	yes
143	*Caller.*	thank you
144	*Operator.*	that's it?
145	*Caller.*	that's it

DEC 28_NM_loc2

We can hypothesise that in a dialogue system these examples can be handled as follows. For a system, there are operations which the developer knows are going to take time due to technical constraints, but are expected to be immediate by the user. In this case the system can produce similar behaviour to that in (11j) (lines 25–27): 'er… (pause) [comes up with an answer] <laughter>'.

Conversely, the system can detect patterns of filled pauses accompanied with laughs from the user and treat them as turn-release cues. It can be a signal of either that there is something that confused the user, or that she genuinely could not come up with an answer due to certain difficulties. A downplayer (e.g., 'Don't worry', as discussed in Section 9.3) or laughter in response could also be appropriate system feedback in such situations. We consider these ideas as topic for further empirical investigations.

Laughter related to smoothing retrieval difficulties can be indicative. Consider the case of language tutoring. In the Anki flashcard app,[3] the system provides users with a word in one language on the front side of the card and the user should provide a translation. The user then gets the correct response from the back of the card and evaluates her own response (was this card Hard, Good or Easy to recall). If we consider making a similar conversational app, indications of retrieval issues – filled pauses ('er em…') and follow-up smoothing by laughter – can lead to the decision to flag this card as 'Hard' and provide corresponding feedback (11l).

(11l) 1 *System.* What is the Swedish for donkey?
2 *User.* er em … åsna?.. <laughter>
3 *System.* Yes, that was tough, but it is correct! (system marks the card as 'Hard')

constructed example

Another example is taken from the Fresno et al. (2022) study of code-switching in language-tutoring dialogues. In (11m) the student is experiencing difficulties with retrieval of a word 'journey' in his second language (English). He provides a word in his first language (Italian, 'traghetto'), requests a word in English and produces a brief laugh associated with his awkwardness due to not knowing the word in English and asking the teacher.

(11m) 105 *Student.* yeah <pause> and during the <pause> *traghetto*
106 *Student.* what's the name? <laughter>
107 *Teacher.* journey.
108 *Student.* journey?
109 *Teacher.* journey.
110 *Student.* journey.
111 *Teacher.* mm.
112 *Student.* during the journey.

ESF Second Language Data Bank, liean24i.1.cha

[3]https://apps.ankiweb.net

11.4.3 *Approaches to evaluation*

Each of the aforementioned improvements has to be a subject for evaluation within the dialogue system. We expect to see that these improvements will be reflected in the following evaluation criteria.

Some of the improvements would fall into a checklist-style set of objective criteria, like being able to understand laughter as negative feedback, or as a signal of surprise. The same goes for a system's laughter as an appropriate reaction to conversational humour.

Another portion of the features can be evaluated only subjectively, for example, it is a question of user preference whether it is okay for a system to accompany asking for favour (e.g., 'Let's start over!') with laughter. For this purpose, subjective evaluation methods like Subjective Assessment of Speech System Interfaces (SASSI) (Hone and Graham, 2000) can be used. We optimistically expect that characteristics such as naturalness and likeability would increase and annoyance would decrease.

11.4.4 *Summary*

In this chapter we looked at how laughter can be integrated in the dialogue context from the point of view of a dialogue manager (DM). So far we have only been concerned with non-humorous contexts, with a few exceptions. In the last part of the thesis we specifically address the properties of dialogue contexts: what is *about* them that makes them laughable?

Part IV

Laughter and humorous incongruities

Introduction to Part IV

И смех, и грех.
And laughter and sin.

Russian saying

The key goal of this part is to provide an account of humour from an interactive perspective. Even though this can be also considered as a first step towards understanding the idiosyncratic properties that relate to a laughable, here we mostly focus on jokes: constructed ones, taken from joke books and natural examples from dialogue corpora.

A key aspect of understanding laughter, given that the laughable is identified, is to understand which aspects of the context are evoked by the laughable. These characteristics of the context are usually known under the rather nebulous term *incongruity*, the definition of which greatly varies in the literature, as we mentioned in Section 3.3.

This chapter paves the way to a precise definition of incongruity in dialogue and develops an implementable account of it, which includes determining the rhetorical aspects of context and how they are used to calculate incongruity.

The contribution of this part is two-fold:

1. We identify the aspects of incongruity calculation as a prerequisite for a contribution to be a (humorous) laughable.

2. We analyse how these aspects can help understand conversational humour, applicable to dialogue systems.

This part is organised as follows. Chapter 12 discusses existing accounts of humour incongruity and motivates the importance of

dialogicity in humour. It introduces the notions of *enthymemes* and *topoi* which are central for our analysis. In Chapter 13 we present our approach and explain four elements of it. Chapter 14 presents a first case study of a joke within the framework of our interactive approach to humour. Chapter 15 presents a detailed implementable account of the role of integration of new information for the perception of humour. Chapter 16 discusses the implications for dialogue systems.

12 Humour and interaction

Sections 12.2 and 12.3 of this chapter were previously published in Maraev, Breitholtz, Howes, Larsson et al. (2021)[1] *and went through minor revisions.*

12.1 WHY AN INTERACTIVE PERSPECTIVE ON HUMOUR?

In this chapter we attempt to motivate the role of interaction in humour comprehension. The main rationale behind our perspective is that, firstly, laughter is an interactive phenomenon (see Section 4.1) and to understand it in relation to humour one needs to consider humour and laughter within the same cognitive system. Secondly, many written jokes involve interactive elements which require the reader to understand multiple perspectives on the situation that is described in a joke (Lepore and Stone, 2014, Ch. 11).

Conversational humour can be defined as very spontaneous and highly contextual. And we can be sure that one joke, which evokes laughter in one context, can, in a different context, lead to a scowl, or even to more serious action: from a dismissal from a job to a prison sentence.

Typically, the literature on humour is more focussed on *humour competence* (Attardo, 2010 among the works reviewed in Section 12.2), which is crucial for making a clear cut distinction between non-humour and *verbal* humour. Here we don't share the same goal, and do not address the question of what humour *is*. We deal with

[1] Vladislav Maraev, Ellen Breitholtz, Christine Howes, Staffan Larsson and Robin Cooper (2021). 'Something Old, Something New, Something Borrowed, Something Taboo: Interaction and Creativity in Humour'. In: *Frontiers in Psychology* 12, p. 1443. All authors contributed to the discussion and research of examples. Maraev, Breitholz, and Howes wrote the paper with contributions from Larsson and Cooper.

the relation between laughter and humour, and argue that for such a purpose one needs to primarily be concerned with performance aspect of humour (in Section 4.1 we already discussed why linguistic performance should not be neglected for dialogue in general).

Having given a brief motivation for the role of interaction for humour, we will now move on to review existing linguistic theories of humour.

12.2 THEORIES OF HUMOUR

One of the most prominent theories which emphasises the importance of linguistic understanding of humorous texts is Victor Raskin's Semantic-Script Theory of Verbal Humour (SSTH) (Raskin, 1985). SSTH introduces influential notions of script opposition and script overlap, which are often used as the basis for 'incongruity theories' (e.g., Oring, 2016) of humour. The main hypothesis of SSTH is formulated as follows (Raskin, 1985, p. 99):

> A text can be characterized as a single-joke-carrying text if both of the text conditions … are satisfied:
>
> 1. The text is compatible, fully or in part, with two different scripts;
> 2. The two scripts with which the text is compatible are opposite in a special sense...

The 'special sense' of opposite, is realised within 'relatively few binary categories which are essential to human life' (Raskin, 1985, p. 113).

The General Theory of Verbal Humour (GTVH) (Attardo and Raskin, 1991), extends the remit of SSTH, to differentiate between verbal and referential humour, and to account for the relative degree of similarity between different jokes. In GTVH, the notion of script opposition is the most abstract of the six 'knowledge resources' which the creator of a joke may draw upon. Our model, which takes the dialogicity of jokes as its core insight, is also compatible with the GTVH, providing a finer-grained way of describing the resources

used in humour. We take these to be based on general resources for interaction.

These, the most well-known theories of humour (SSTH, GTVH) are only concerned with humour competence (Attardo, 2010). They abstract away from the actual process of joke comprehension and do not include processing as a crucial condition for humour (Ritchie, 2018). Acknowledging Ritchie's claim about a lack of actual explanations regarding how jokes are processed as text, we view the dialogicity of joke processing as a crucial condition for getting a humorous effect that may result in amusement, a smile or laughter.

In recent decades incongruity-resolution theories became influential (Hempelmann and Attardo, 2011; Hurley et al., 2011). The key assumption is that most of jokes require a *resolution* step, accounting for the decrease in the oddity of the situation as a joke unfolds. However, many scholars point out that the key concepts in this group of theories also lack precise definitions (Ritchie, 2004; Morreall, 2011; Warren and McGraw, 2016).

Ritchie (2018) emphasises the importance of explicating these so-called 'theory-internal' concepts in 'theory-external' terms which will arise from more general explanations relying on underlying cognitive processes, such as text comprehension. We agree with this principle in general, and in our case we explicate our approach to humour in terms of a wider theory of incremental reasoning in dialogue.

Considerable research has been done in conversational humour, mainly studying it on qualitative and sociological grounds, for instance, Hay (2000) studies gender differences in humour production, Davies (1984) considers the group activity of 'joint joking' and highlights different styles of such an activity, and Günther (2003) provides an analysis of canned jokes and corresponding laughs in the British National Corpus (BNC) based on Conversational Analysis (CA). Our goal here is to provide an account of humour in jokes with a potential to scale down to mechanisms at play in naturally occurring humorous episodes.

One attempt to define incongruity to explain the properties of a laughable was made by Ginzburg et al. (2015) and Ginzburg et al.

(2020), who used the notions of topoi and enthymemes to account for potential clashes. In this thesis, following Breitholtz and Maraev (2019), we claim that incongruity in jokes can be cast in terms of *topoi* (as a resource to account for different ways of opposing) and *enthymemes* (as an argument occurring in a dialogue or text, and involving one or more topoi) that arise from specific interactional experiences. We see the ability to manipulate incongruity in this way as being central to conversational and situational humour. In the next section we will elaborate on these principles.

12.3 HUMOROUS CONVERSATION

Many theories of humour focus exclusively on written versions of jokes with an idealised non-present audience. However, in reality, humour is always based in an interactive context, and, we argue, the cognitive and social mechanisms managing dialogue processes like turn taking, repair, grounding and contextual enrichment, are also the mechanisms that allow us to produce and interpret both linguistic and non-linguistic humorous events.

The dialogicity of jokes and other humorous events is reflected in the emphasis on the sequential structure of jokes in many studies of humour (see for example Suls, 1972; Ritchie, 2018). At each increment there is a potential for participants in a humorous exchange to interpret things differently. This is often exploited in jokes. For example, the joke in (12a) plays upon the fact that the perspectives of the two characters are different and this fact is revealed to the joke hearer incrementally.

(12a) A senior citizen is driving on the highway. His wife calls him on his cellphone and in a worried voice says, 'Herman, be careful! I just heard on the radio that there was a madman driving the wrong way on Route 280'. Herman says, 'Not just one, there are hundreds!'

in Ritchie (2018) cited from Hurley et al. (2011)

The example above illustrates dialogicity *within* a joke. The joke is set up as a dialogue between two characters with different takes

on the situation. However, interaction is also fundamental in joke telling (or joke reading or joke interpretation) *events*. For example, the author of a joke book might not direct a particular joke at a specific individual. However, she must have some idealised audience in mind, one that is likely to get the joke. This means that even in contexts such as social media, humour is inherently dialogical. It is not only important to look at the cases when humorous tweet receives an explicit response, like in (12b), but also consider the opportunity to respond, which may or may not be taken up. In (12b) the Twitter user *J* is making a joke referring to the social distancing rules introduced in the pandemic and the trope that men sometimes exaggerate their height on dating sites.

(12b) 1 *J.* guys will stand 5'8" from you and call it 6 feet
2 *K.* Most guys can't tell what six inches look like let alone six feet…

https://twitter.com/andrealongchu/status/1246445034792239106

One important consequence of the dialogicity of humour is the possibility that interlocutors might interpret the same piece of discourse in distinct ways, just as characters within a joke can, and the source of humour is often a play on this potential for multiple interpretations. This potential is a consequence of, among other things, the inferential nature of language in use.

Jokes, like any piece of discourse that in some way involves implicit meaning, necessitate drawing on some kind of resources about the world (Yus, 2003) in order to reason from what is explicitly said. These resources could be facts, judgements about people and society, etc. which underpin inferences and associations made by interlocutors. We suggest analysing humorous interactions in terms of *enthymemes*, arguments where the conclusion does not follow by necessity, usually because one or more premises are not explicit in the discourse. The rules of thumb warranting enthymemes are referred to as *topoi*. Ducrot (1988, 1980) and Anscombre (1995) argue that topoi are essential not only for argumentation but for all kinds of interaction, as they supply implicit information and

principles of reasoning which must be recognised by an interlocutor for enthymematic discourse to make sense. For example, if Alice is going out on a rainy day, and Bob advises her to take an umbrella, it is implicit that the umbrella provides protection from the rain. If Bob in the same situation tells Alice to put on a sun hat, the comment would either not make sense to Alice, or be taken as irony due to general practices associated with umbrellas and sunhats and different types of weather.

In our analysis it is the juxtaposition of contrasting topoi which creates a humorous effect. For example, (12b) relies on two contrasting topoi: a corona-specific *safe-distance* topos that people should stay 6 feet apart and a topos, associated for example with dating apps and web sites, that men who are 5'8" tall often claim to be 6', with *6 feet* as a point of overlap between the two. We will return to this example in Chapter 13.

Topoi may be very generally applicable, such as the topos of gravity, which holds in most contexts on earth. However, often topoi are specific to, or at least more strongly associated with, particular socio-cultural contexts. This may be the citizens of a nation, the members of a sub-cultural group or people in a particular age span, such as children in school. Also, just as new topoi emerge when new situations arise, established topoi gradually disappear as norms and circumstances change. For example, consider the joke in (12c):

(12c) 'What game does a lady's bustle resemble?'
'Back-gammon!'

This joke is a word play on the name of the game backgammon and gammon as a joint of meat, the rear leg of a pig, implying that this is what a bustle[2] looked like. This fashion of making your backside look huge was much ridiculed at the time, and there was even a particular genre of 'bustle jokes'. Today, there is still an overarching topos that changing the way you look to appear more attractive is slightly ridiculous. However, this applies to things like

[2]a type of woman's undergarment used in the mid to late 19th century that added volume to the back part of the skirt just below the waist

botox, but not to dying one's hair (unless you are a man). So, even if we know what a bustle is, the humour is less obvious to us than it would have been to a 19th century person who had access to a topos that if x uses a bustle, x is vain and slightly ridiculous, while no similar topos existed for example with respect to corsets.

Another example is (12d):

(12d) 'Dear Postnord Customer! The Corona pandemic poses big challenges for our company. How can we claim that we sought you, but that no-one was at home?'

https://twitter.com/Muralgranskaren/status/1250294659957452809

The joke is a fabricated message from the Swedish postal service Postnord, which has a bad reputation for service in general. The topoi that are relevant for interpreting this joke are that since there is a pandemic, people are at home, and that Postnord tend to make excuses for not delivering, blaming the recipient or sender for not having met the conditions for delivery.

The basic topos that this joke evolves around is the principle that if someone is at home and there is a parcel for them, the parcel can be delivered. We represent that in (12.1) below. In our semi-formal notation, the premises are shown above the line and the conclusion below, as is standard. The wiggly line denotes a not strictly logical chain of reasoning, as opposed to for example an if-then sequence separated by a straight line, which indicates a logical inference. These are not intended to be complete formal representations, but rather as a convenient and clear way of representing our intuitions about topoi and enthymemes. More complete formal representations are shown in Breitholtz (2020) and in Chapter 15. The topos that if someone is at home they can receive a delivery is acceptable to most people. It is also relatively uncontroversial that if the opposite were true, that the person who is expecting a delivery is not at home, the parcel cannot be delivered, as seen in (12.2).

$$\frac{x\_\text{is\_a\_parcel\_for}\_y \quad y\_\text{is\_at\_home}}{\text{deliver}\_x\_\text{to}\_y} \tag{12.1}$$

$$\frac{x\_\text{is\_a\_parcel\_for}\_y \quad y\_\text{is\_not\_at\_home}}{\text{not\_deliver}\_x\_\text{to}\_y} \tag{12.2}$$

Topos (12.2) licences an argument that a particular parcel hasn't been delivered by Postnord since the recipient was not at home, and is thus applicable to situations where the premises above the wiggly line in (12.2) are instantiated. However, there is a third topos in play here – one saying that Postnord claims (possibly falsely) that they are unable to deliver parcels since recipients are not at home.

$$\frac{x\_\text{is\_a\_parcel\_for}\_y \quad \text{A\_claims}\_y\_\text{is\_not\_at\_home}}{\text{unable\_to deliver}\_x\_y} \tag{12.3}$$

An argument based on (12.2) is acceptable (though possibly mistrusted due to (12.3)) in situations where the claim that the recipient is not at home is true or at least not clearly false. However, in times of lockdown, where the vast majority of people are at home most of the time, this is very unlikely to be the case.

12.4 WHAT'S IN A LAUGHABLE?

In the work of Mazzocconi et al. (2020) incongruity goes beyond humorous laughables and is considered to be a more general notion. It scales up to other types of laughables, including non-humorous ones. As we mentioned in Section 3.3, Mazzocconi (2019) defines incongruity as '[involving] a clash between a general inference rule (a topos) and a localized inference (an enthymeme)'. This definition allows very wide coverage of incongruities associated with laughables. For instance, in example (12e) the ironic laughter reverses the meaning of 'history ended with Ronald Reagan' which is opposite to what was in fact said by the lecturer. In Chapter 11 we have shown how such meaning reversal can be operationalised within

the dialogue manager (DM). Ginzburg et al. (2020) indicate that such reversed meaning clashes with the sincerity topos 'If *A* says *p*, then *A* means *p*'. This topos reflects a social convention of some sort which is violated by this humorous (ironical) contribution. This allows one go beyond the scope of written jokes and analyse humour as a more complex phenomenon.

(12e) *Lecturer.* … And then of course you've got Ronald Reagan … and <laughter> history ended with Ronald Reagan.

BNC JSM

Nevertheless, we cannot be sure that laughter-inducing inferences are always localised. For instance, the humorous exchange in (12b) involves some clashing allusions which are beyond the actual text. Furthermore, inference rules do not necessarily need to be general, but rather can be created ad-hoc. In the following chapters we will explore this further, looking at humour as a complex phenomenon which is based on dialogical inference mechanisms.

13 The interactive approach to humour

The content of this chapter was previously published in Maraev, Breitholtz, Howes, Larsson et al. (2021).[1] *Section 13.5 was extended and the rest of the chapter went through minor edits.*

13.1 THE ELEMENTS OF THE THEORY

The aim is to build our approach around the elements that arise from considering humour as a dialogue phenomenon, and relate them to existing theoretical notions. Four elements of our approach are discussed in the following sections:

- Mapping the joke to existing rhetorical resources (Section 13.2)
- New resources, their acquisition and decay (Section 13.3)
- The process of integration of new information (Section 13.4)
- Emotional appraisal of the content (Section 13.5)

Our claim is not that any one of these elements is either necessary or sufficient for humour. Indeed the same utterance can be experienced as humorous by one agent and not humorous by

[1] Vladislav Maraev, Ellen Breitholtz, Christine Howes, Staffan Larsson and Robin Cooper (2021). 'Something Old, Something New, Something Borrowed, Something Taboo: Interaction and Creativity in Humour'. In: *Frontiers in Psychology* 12, p. 1443. All authors contributed to the discussion and research of examples. Maraev, Breitholz, and Howes wrote the paper with contributions from Larsson and Cooper.

another agent in the same situation. Furthermore, a single agent can experience something as serious at one time and then at another time find it funny (as in 'We laugh about it now, but it was deadly serious at the time'). The presence of one or all of these elements is, then, no guarantee that something will be experienced as humorous.

We do, however, hypothesise that anything which is experienced as funny will have at least one, and often several, of these elements.

13.2 MAPPING THE JOKE TO EXISTING RHETORICAL RESOURCES

13.2.1 What gets integrated?

By *rhetorical resources* here we mean *topoi*, rules of thumb that warrant enthymemes in the dialogue context (see Section 12.3). We call *integration* a formal way to combine a piece of discourse with existing knowledge and a dialogue state, in a similar way to how moves were integrated in Chapters 10 and 11.

In this section we will look in more detail at example (13a) (a part of (12b), introduced earlier), and see how a coronavirus-related topos has been creatively combined with another topos for humorous effect. Informally, we can speak about two topoi here: the *safe distance* topos and the *dating website* topos, with '6 feet' as a point of overlap between the two topoi.

(13a) Guys will stand 5'8" from you and call it 6 feet.

Information which is present in the joke needs to be integrated with pre-existing knowledge. The joke brings a few puzzles when processed, which require additional creative effort from the listener. Why do guys call the distance 6 feet when it is 5'8"? How easy is it to notice 4" difference in distance? Why does this relate to guys specifically, and not to people in general? Overall, some imagination is required from the listener.

But what can this imagination be based on? We argue that connotations of the words used play an important role, and this can be

expressed in terms of the topoi that are available for conversational participants (including the audience). In any given situation or context there will be several topoi which are potentially applicable, but some will be more salient than others. In the case of topoi related to the coronavirus, these are particularly salient as they are directly related to people's everyday lived experience. Much humour relies on the existence of the multiplicity of applicable topoi in any given context. More generally, jokes are often based on the asymmetry of the salience of topoi (we refer the interested reader to Chapter 15 for discussion).

More formally we can speak of two crucial topoi; during the coronavirus pandemic people should stand 6 feet apart (to prevent the spread of the disease), which we represent as (13.1), and the topos that guys exaggerate their height on dating apps (13.2).

$$\frac{x\_\text{is}\_\text{a}\_\text{person} \quad y\_\text{is}\_\text{a}\_\text{person}}{x\_\text{and}\_y\_\text{should}\_\text{stand}\_6'\_\text{apart}} \tag{13.1}$$

$$\frac{x\_\text{is}\_\text{a}\_\text{guy} \quad x\_\text{is}\_5'8''}{x\_\text{claims}\_\text{he}\_\text{is}\_6'} \tag{13.2}$$

In order to see what mechanisms are required for the creative process of comprehension let us modify the joke slightly, to see which components are required to make it comprehensible and humorous.

13.2.2 Relocating the joke to the UK

First, let's move our joke to the UK, where people refer to height in imperial units, but the coronavirus social distancing rule is formulated as 'Stay 2 metres apart from anyone not in your household'.[2] Therefore (13.1) requires one or several additional premises in order to be processed. We can see (at least) two possible reasoning patterns: one option (13.3) is to add the premise that person x and person y are located in the USA. Another option (13.4) is to reason by seeking an analogy of the corona specific 2-metre rule, that is

[2]https://www.gov.uk/guidance/national-lockdown-stay-at-home

the 6 feet rule. Overall, taking the additional premises into account, the humour ought to be less obvious (and perhaps less funny) for a Brit as compared to an American, although this can be further investigated in an empirical study.

$$\frac{x\_\text{is}\_\text{a}\_\text{person} \quad y\_\text{is}\_\text{a}\_\text{person} \quad x\_\text{and}\_y\_\text{are}\_\text{in}\_\text{USA}}{x\_\text{and}\_y\_\text{should}\_\text{stand}\_6'\_\text{apart}} \tag{13.3}$$

$$\frac{x\_\text{and}\_y\_\text{should}\_\text{stand}\_2\text{m}\_\text{apart} \quad 2\text{m}\_\text{is}\_\text{analogous}\_\text{to}\_6'}{x\_\text{and}\_y\_\text{should}\_\text{stand}\_6'\_\text{apart}} \tag{13.4}$$

13.2.3 Relocating the joke to Europe

In Europe, feet are not used at all in measurements. Let us try to adjust the joke to the European metrics and corona-specific rules by changing the coronavirus restriction. Given the restriction, in the joke we will need to 'call it' 2 metres. But what about the height? We have two alternatives: direct conversion of 4" difference (13b) or picking some arbitrary height, which is 'not good enough' for dating (13c).

(13b) Guys will stand 189 cm from you and call it 2 metres.

(13c) Guys will stand 1.6 metres from you and call it 2 metres.

Here the corona-specific topos, similar to (13.1) but involving 2 metre distance is invoked, but not the *dating website* one, because 2 metres is commonly considered 'too much' for a height. Therefore the joke basically doesn't work, as the inference (13.5) does not hold (shown with crosses) in the given context.

$$\begin{array}{c} x\_\text{is}\_\text{a}\_\text{guy} \quad x\_\text{is}\_1.6\text{m} \\ \times \\ x\_\text{claims}\_\text{he}\_\text{is}\_2\text{m} \end{array} \tag{13.5}$$

The joke also does not pass the direct measurement conversion test:

(13d) Guys will stand 172 cm from you and call it 183 cm.

Here the coronavirus social distancing topos is no longer salient and the 'dating website' topos is not salient either. One of the possible ways to encourage associations with online dating is to substitute the very precise 183 cm by (say) 185 cm. But this would not make it humorous, just bizarre and possibly far-fetched: one can think of it as a riddle, and the solution to it is to convert cm to inches, think of it in an American context and only then get to the humour.

13.2.4 Guys

One more thing to test is to break the compatibility with the *dating website* script, or, more specifically, topos (13.2) which constitutes it, and is itself based on the more general topos that being tall (but not too tall, as discussed in Section 13.2.3, above) is considered an attractive quality in men (at least in Western societies), such that men who do not meet the tallness criterion of attractiveness may be inclined to claim that they do in situations involving searching for a partner.

(13e) People will stand 5'8" from you and call it 6 feet

Although the (USA-specific) corona social distancing topos still applies here, (13e) does not invoke the same associations between height exaggeration on dating apps because 'people' usually encompasses both men and women. There is no common topos about women exaggerating their height to attract a date, and different norms apply. As with 2 metres for men, discussed above, 6' is generally considered excessively tall for a woman. This means that even if there were an equivalent topos about women exaggerating their height on dating apps, the heights in question would be e.g., 5'4" and 5'9", which would not be compatible with the corona social distancing topos.

13.3 NEW RESOURCES, THEIR ACQUISITION AND DECAY

13.3.1 Humour and novel events

Informally we think about the creation of humorous discourse as involving something old and something new. In the case of the new jokes around the coronavirus pandemic, this means that well established and generally accepted topoi are combined in some way with topoi which are novel. We have already seen examples of this in (12b) and (12d).

However, this is not always the case, as jokes can be created from two or more different topoi which are already available to a competent language user. In this case, what is new, we argue, is the relationship which is established between the topoi. As the combinations become overused through repeated exposure, these lose their novelty and the jokes lose their humorous effect.

The novelty of a topos is not fixed, either. Repeated exposure to a topos means that the novelty value decreases and the possibility to make jokes using the topos in creative ways also diminishes. It seems likely that there is a quantifiable relationship between the novelty of topoi or combinations of topoi and how humorous they are perceived to be (as seems to be the case with so-called 'Dad jokes' which may induce laughter in children who have not encountered them before, but groans from more experienced members of the language community), but this is an empirical question for future work.

We take our examples from the coronavirus pandemic precisely because it has led to large quantities of new information and topoi becoming widespread in society. This rapid introduction of new topoi (in this case related to the coronavirus pandemic) has lead to many instances of humorous creativity in the form of jokes, memes, videos and funny exchanges, rapidly disseminated through social media. This makes the coronavirus pandemic a perfect case study for exploring the acquisition of new resources and their further integration with previously existing ones.

13.3.2 *The case of toilet paper on the verge of the pandemic*

The decrease of novelty of topoi is particularly clear where many new topoi quickly became shared – available as resources for a community of language users – in a short space of time, due to exceptional circumstances. In the case of the coronavirus pandemic of 2020, early on in the pandemic (before many countries went into lockdown) people started panic buying certain goods such as toilet paper. This led to the topos 'if you are going to be in lockdown, you need plenty of toilet roll', with a chain of reasoning from existing topoi that can be paraphrased as: if something is essential then you don't want to run out of it (13.6), and if you don't want to run out of something then you should buy lots of it (13.7).

$$\frac{x\_\text{is\_essential}}{\text{do\_not\_run\_out\_of\_}x} \tag{13.6}$$

$$\frac{\text{do\_not\_run\_out\_of\_}x}{\text{buy\_lots\_of\_}x} \tag{13.7}$$

As detailed in Breitholtz (2020), two linked topoi can also be expressed in a single topos, such as (13.8). Given the new premise that during a lockdown toilet roll is an essential item (13.9), and that during a lockdown there are limited opportunities for buying goods leads to a more specific version of topos (13.10). This led to jokes such as that in (13f) when the new topoi first became shared, but these typically did not persist as the context changed and it became clear that buying toilet paper was still possible during lockdown.

$$\frac{x\_\text{is\_essential} \quad \text{do\_not\_run\_out\_of\_}x}{\text{buy\_lots\_of\_}x} \tag{13.8}$$

$$\frac{x\_\text{is\_toilet\_paper} \quad \text{lockdown}}{x\_\text{is\_essential}} \tag{13.9}$$

$$\frac{x\_\text{is\_toilet\_paper} \quad \text{lockdown} \quad \text{do\_not\_run\_out\_of\_}x}{\text{buy\_lots\_of\_}x} \tag{13.10}$$

(13f) 'Why did the chicken cross the road?'
'She saw a shop with some toilet rolls left.'

In addition to the corona specific new topoi and the pre-existing old topoi, getting the joke in (13f) also requires knowledge of the joke frame in English of the classic chicken joke (13g), which the lockdown chicken joke subverts and exploits.

(13g) 'Why did the chicken cross the road?'
'To get to the other side'

Interestingly, while the classic chicken joke is usually considered to just be absurdist,[3] subverting the notion of a chicken crossing the road for exactly the same reason a person would (which even small children can grasp), it originally may have had a double meaning relying on knowledge that where you go when you're dead can be referred to as 'the other side', which was well known when the chicken joke first appeared (presumably some time before it is first attested in print in a 1847 New York periodical), though may be a less accessible topos now (or completely unavailable, as with the 'bustle' example (12c)). This additional knowledge that a (suicidal) chicken crossing a road is likely to be hit by a car and killed adds another level to our understanding of the joke.[4] This ability to get the joke at different levels is characteristic of jokes – which rely on interlocutors having different (and possibly multiple) interpretive resources available.

13.3.3 Outdated jokes about the 1918 flu pandemic

The dynamic nature of which topoi are salient in a particular situation also means that certain humorous comments which would not have been interpretable to us (or at least would have required a significant effort to understand) before the coronavirus pandemic

[3]Wikipedia, for example, describes it as anti-humour, https://en.wikipedia.org/wiki/Why_did_the_chicken_cross_the_road%3F.

[4]See e.g., https://www.esquire.com/uk/life/news/a12346/the-upsetting-true-meaning-of-that-why-did-the-chicken-cross-the-road-joke/

now become comprehensible due to our new salient topoi, which are analogous to many from the 1918 flu pandemic, such as (13h) and (13i), about 'flu' masks, which are also prevalent in the coronavirus pandemic (though usually referred to as 'face' masks).[5]

(13h) 'Flu masks improve the appearance of many men, but when worn by women, they take much of the joy and beauty out of life.'

(13i) 'Every woman secretly believes she would be fascinating in a harem veil. Wearing a flu mask is a good, safe way to try the effect.'

Other jokes which may not be so obvious to a modern audience, such as (13j) rely on the context of the 1918 flu pandemic occurring at the same time as the first World War, with the Allies fighting the Germans led by Kaiser Wilhelm II. This joke can, however, be updated to the 2020 context by simple substitution of both the disease and a controversial figure, as in (13k). Whether you find this funny or not will also depend on your political persuasion, which also relies on your acceptance of a number of associated topoi.

(13j) 'The Kaiser and the Flu are running neck and neck in the world's popularity contest.'

(13k) 'Vladimir Putin[6] and the coronavirus are running neck and neck in the world's popularity contest.'

13.4 THE PROCESS OF INTEGRATION OF NEW INFORMATION

13.4.1 Accommodation and forced reinterpretation

A common technique for creating humorous effect is importing a topos from a different domain or a type of situation to the context

[5]Taken from https://www.smithsonianmag.com/history/memes-1918-pandemic-180975452/

[6]The joke target was changed from Donald Trump who had been in the draft version of this thesis until March 2022.

of the joke. This involves *accommodation* (Lewis, 1979; Beaver and Zeevat, 2007), integration of new information which is in some way conveyed or hinted at in an utterance but not explicitly stated. Accommodation is frequent in dialogue and often happens seamlessly as the things we accommodate are non-controversial (Larsson, 2002; Breitholtz, 2020). We believe that many or even all utterances which involve reasoning require the accommodation of topoi. Normally, although not accessible before the utterance in question, this accommodated information is more or less salient in the current domain or context. However, the cases in which a humorous effect is created seem to require accommodation of topoi that are not the most salient and that need to be integrated from a different domain or context.

In previous examples (12b) and (12d), we have seen how jokes rely on combining something old (in these examples, from pre-corona times) and something new (corona-related). We argue that these jokes can be described as accommodating a new topos (from the corona pandemic context) into an old context (males boasting in online dating sites and late mail delivery, respectively for (12b) and (12d)). However, the accommodation effect comes out even more clearly when a new topos is integrated into the context of a more clearly defined existing joke structure, such as knock-knock jokes. In such cases, the joke structure is assumed to be familiar to the hearer(s), and the jokes rely on jointly establishing the context of the well-known joke structure.

To make clear how these jokes rely on access to the topos to be integrated, we will look at a dialogic exchange where a dialogue participant *lacks* sufficient knowledge of the context that the topos to be accommodated is to be integrated from. (In this example, the integrated topos is neither new nor corona-related, although it can be assumed that the joke was perceived as more funny when the integrated topos was more recent and more salient than it is now.) The excerpt is an example of explicit joke telling from the British National Corpus (BNC). In this extract, 6-year old David reproduces the knock-knock joke (in line 3799) without understanding its meaning. We can say that he does not understand what is incon-

gruous about the Avon lady knocking, which is what (allegedly!) makes the joke funny.

(13l) Phillip (46), Jane (40), Christopher (9), David (6) – at home having breakfast. Overlapping material is shown in square brackets.

3797	*David.*	Knock, knock.
3798	*Jane.*	Who's there?
3799	*David.*	The Avon lady, your bell's broken!
3800	*Phillip.*	The Avo- Avon lady?
3801	*David.*	Mm mm.
3802	*Phillip.*	What does she do?
3803	…	
3814	*David.*	Dad, I don't know what an Avon lady does.
3815	*Phillip.*	What does she do?
3816	*David.*	I don't know.
3817	*Phillip.*	Mm mm!
3818		Oh!
3819		Well she doesn't come here.
3820	*David.*	She fixes bells.
3821	*Phillip.*	<laughing: No>
3822	*David.*	Well what [does she do?]
3823	*Jane.*	[Guess] can't you?
3824	*Christopher.*	<talking from other room> She rings the bell, she rings.
3825		And she
3826	*Jane.*	She co-
3827	*Christopher.*	<unclear>
3828	*Phillip.*	Okay.
3829		Thanks Chris.
3830	*Jane.*	She's somebody who comes to the door and tries to sell you some make-up and perfume and toys and things.

BNC KCH

To understand this joke at least two things are required: a) knowledge of the general structure of knock-knock jokes and b) cultural knowledge of the Avon lady being a door to door salesperson (for Avon make-up products) who, according to the longstanding advertising campaign, rings the bell (leading to the advertising slogan 'Ding Dong, Avon Calling' becoming a well-known phrase.[7] This joke breaks the pattern of knock-knock jokes, as 'knock-knock' doesn't generally bear any sense apart from being a set-up for an upcoming pun from the joke teller.

When Phillip asks David to explain the joke (which is not for Phillip's benefit, but because he does not expect his son to have access to the appropriate topos), David (line 3820) proposes a topos which is compatible with the joke (someone who fixes bells would expect a broken doorbell, and therefore knock at the door). This topos is rejected by his father, Phillip (line 3821), although the rejection is accompanied by laughter, which indicates a mismatch between David's topos and the actual one. David's explanation is treated by Phillip as a humorous episode, albeit an unintentional one. Later, Christopher (line 3824) explains what the Avon lady does, which may help David to get the joke, and Jane (line 3829) also adds more information which might help David to understand.[8]

Next, we will show an example of integrating of a new (corona-related) topos into a old (pre-corona) joke context. Here, the context is again clearly identifiable (erotic role-play) although perhaps more loosely structured than the 'knock-knock' joke. It is a prime example of integration and also highlights the temporal dynamic of dialogic jokes by invoking a so-called 'garden path' mechanism (Attardo and Raskin, 1991). Ritchie (2018) calls this type of joke construction the 'forced reinterpretation'. The joke teller has two possible interpretations of the joke set up in mind, or, more specifically, two topoi which can underpin the communicated enthymeme. Using the sequential ordering of the information in the joke, the joke

[7] see e.g., https://www.youtube.com/watch?v=66IWgU9AAis from 1956.

[8] Note that understanding a joke and finding it funny are not the same thing. We do not go into this distinction here.

teller boosts the salience of one of the topoi, nudging the listener towards one of the possible interpretations. This encourages the listener to accommodate this particular topos. The punch line then subverts this accommodation, revealing another interpretation of the joke.

(13m) 'Darling…fancy putting on a nurse's uniform?'
'Ooh, cheeky boy…you feeling horny?'
'Nah…we've run out of loo roll'

A teller of the joke in (13m) presents an enthymeme in the first two utterances of the joke; this enthymeme can be rephrased as 'If A is persuading B into wearing a nurse's uniform, then A is feeling horny' and it is an instance of a topos similar to (13.11), namely that a (sexy) nurse's uniform may be worn as part of an erotic role play situation.

$$\frac{x\_\text{wears\_nurse's\_uniform}}{x\_\text{is\_involved\_in\_erotic\_role\_play}} \tag{13.11}$$

The joke teller plays on this accommodation, presenting the final utterance, which explicitly negates this assumption ('Nah…') and providing the new reason for wearing a nurse's uniform. The reasoning behind understanding the punchline is unfolded as follows: due to the coronavirus lockdown restrictions people in general are not allowed to go out. However, these restrictions do not apply to key workers (including nurses). In the UK, for example, in the lockdown of Spring 2020, special shopping hours were introduced for National Health Service (NHS) staff, who were also exempt from quarantine restrictions. In the situation projected in the joke the reasoning is based on the lockdown specific topos that if one pretends to be a nurse, one is allowed to buy toilet paper.

13.4.2 The role of incrementality

In order to create a humorous effect it is not only inferences which play a crucial role, but also the *order* in which they are made. This is pointed out by Ritchie (2018) as a major critique against the

Semantic-Script Theory of Verbal Humour (SSTH) (Raskin, 1985) which claims that we can consider the text to be 'joke-carrying' without sequential and procedural factors. We believe that one reason that order matters has to do with the integration process, in the sense that a familiar context first needs to be established so that the integration of a new topos creates a humorous effect, by forcing the hearer to infer and accommodate the new topos. This is an attempt to explain more specifically *why* order matters, in terms of participants' real-time inferential work on the level of topoi in dialogue.

Let's consider the following reformulation of the 5'8" joke (12b) which we claim is significantly less funny:

(13n) Guys keep their distance just like they lie about their height on Tinder. They will stand 5'8" from you and call it 6 feet.

Here the first sentence is the crucial inference that is assumed to be made by the listener of the joke. In our opinion, making the inferred overt ruins the humour, or at least makes the joke much less amusing. This emphasises the importance of the process of integrating new information by the listener, and the corresponding assumptions that are made by the joke teller.

Another example is given in (130), a modified version of (12a). Here, merely adjusting the order in which information is introduced, without making anything more explicit, seems to make a joke less funny (but perhaps more confusing).

(130) A senior citizen is driving on the highway and confronts hundreds of cars driving the wrong way. His wife calls him on his cellphone and in a worried voice says, 'Herman, be careful! I just heard on the radio that there was a madman driving the wrong way on Route 280'.

To sum up, one might argue that all jokes that combine topoi from different contexts are examples of taking something from one context into the other. However, the integration aspect is more

clearly brought out when one context (often but not necessarily a joke-related context) is first established, and then an unexpected topos from a different context is introduced.

13.4.3 What if it is not possible to integrate a topos?

Sometimes the information given in dialogue is so puzzling that it is impossible to integrate it, because there is no knowledge resource which is available. Let's consider the example in (13p), as it could be told in a dialogue situation. We use a constructed example[9] because it allows us to abstract away from complex cultural and social assumptions as well as situational context, and treat the discourse on a level of very basic assumptions.

This joke is a good illustration of how interlocutors build a common ground incrementally, agreeing on and refuting topoi drawn on to underpin the dialogue.

(13p)
1 *A.* How do you put an elephant into a fridge?
2 *B.* Hmm, I don't know?
3 *A.* Open the door, put the elephant inside, close the door.
4 *B.* Haha okay
5 *A.* How do you put a giraffe into the fridge?
6 *B.* Open the door, put the giraffe inside, close the door?
7 *A.* Wrong! Open the door, get the elephant out, put the giraffe inside, close the door.

constructed example

The question evokes a topos about how to put things in fridges, which is in some way restricted to the kitchen domain. In this context, the idea of how to put something into a fridge is obvious, and also restricted to things that are (usually) food, and of the right size. This leads the interlocutor *B* to say that he does not know how to put an elephant into a fridge.

The joke-telling genre indicates in this instance that *A*'s question ('How do you put *x* in a fridge') is not really a request for information

[9] Based on the joke that appears at: http://jeremy.zawodny.com/blog/archives/009023.html

but has an answer which is known to *A* and which is to be revealed to *B*. On the other hand, the question is odd, which leads B (or the audience) to expect a non-trivial answer. It is important to draw attention to this because it is this oddity that provokes a chuckle from the listener when the triviality is revealed.

One way of characterising 'oddity' is in terms of congruity (or incongruity) with regard to salient topoi. The activity of putting something in a fridge is associated with a particular sequence of events. However, this sequence of events or actions will work more or less well to create the state of x being in the fridge. We can think of a scale of oddity for these kinds of questions (Table 13.1):

degree	*example*
Trivial	'How do you put a cheese in a fridge?'
Tricky	'How do you put a big cake in a fridge?'
Odd	'How do you put an elephant in a fridge?'

Table 13.1: Degrees of oddity

We can think of *trivial* and *odd* as eliciting incongruity. The trivial question addresses something that is considered to be known, and the odd one addresses something ridiculously impossible. A nice example of a trivial question is 'Why did the chicken cross the road?' Questions are usually not supposed to address knowledge that can be easily inferred from the question (crossing the road entails getting to the other side of it).[10] This can be also be explained by violation of Grice's Maxim of Quantity (Grice, 1975): the answer 'to get to the other side' does not provide any additional information, and is thus superfluous.

A *tricky* question requires some non-trivial resolution, for example:

(13q) 'How do you put a wedding cake in a fridge?'
'You will need to remove one of the shelves.'

[10] However, as we point out in section 13.3.2 this joke has additional level of interpretation.

In (13p), giving the answer (line 3), *B* relaxes the implausibility of the elephant being put inside a fridge with no additional non-trivial actions. B accepts the required sequence of actions and acknowledges that (line 4). But is this enough to answer the question about a giraffe?

B (line 6) gives an answer based on his newly acquired story-world, where elephants fit into fridges. But, apparently what B has acquired is not enough: putting a giraffe into the fridge requires several other assumptions to be accommodated.

1. Even given that the fridge is 'magical', and big enough to fit an elephant, it is still not big enough to fit two big animals.

2. The joke-teller is talking about the very same fridge (this is especially important for languages in which there is no definite article)

3. Even if *B* understands that *A* is talking about the same fridge, it is not obvious that it already has an elephant inside, since it has never been explicitly said that an elephant has been put into a fridge.

In Chapter 15 we will formalise the integration process in this joke using Theory of Types with Records (TTR).

13.5 EMOTIONAL APPRAISAL OF THE CONTENT

13.5.1 The appraisal

As we noted in the Section 13.4, the elements of a joke, such as premises and conclusions, are presented to the listener for accommodation. We recall that they can be derived from rhetorical resources, either overtly present or alluded to in the joke. Here we are agnostic about whether we are concerned with individual elements or their combinations.

Any element can be appraised as a reference to a sensitive subject or an insult. We borrow the term *appraisal* from cognitive

theories of emotions in consensus with the Ginzburg et al. (2020) approach to the appraisal of laughables (see Section 3.6). For instance, in (13n) the message (the topos) which was communicated covertly is that guys often exaggerate their height. Here the topos contains a criticism, therefore it can be considered sensitive, because direct criticisms are not acceptable in some cultures, and appraised as improper.

In this section we are going to describe necessary considerations for emotional appraisal of jokes which include their taboo and improper aspects.

13.5.2 Relation to taboo

Taboo subjects are those which it is not (usually) acceptable to talk about in a given society. This may be because it is repulsive (as with bodily functions, such as faeces and vomit) or because it is considered morally unacceptable (such as adultery, incest or cannibalism). Many societies have taboos about sex and death, with other taboos (for example about particular types of food) demonstrating that taboos are based on specific cultural norms. Several of the jokes in this chapter involve an element of taboo. For example, in (13m), the initially evoked topos is about erotic role play, which is taboo in most contexts. An example is the joke (13r), below.

(13r) Since everybody has now started washing their hands, the peanuts at the bar have lost their taste.

Here the communicated topoi are that the taste of people's fingers greatly contributes to the taste of communal bowls of peanuts, and if people don't wash their hands there will be traces of many things on their hands. In particular, there is a topos that people do not wash their hands after going to the toilet, so the peanuts will contain traces of urine or faecal matter – a classic taboo subject. This topos is also the basis of an urban myth claiming that there was a scientific study done on bowls of bar peanuts which found traces of a number of different urine samples.[11]

[11] https://www.cottagesmallholder.com/peanuts-or-peenuts-2054/

What counts as a taboo also depends on the context of the interaction (in a patient doctor interaction, for examples, bodily functions may be legitimately discussed) and is also gradient with certain topics being seen as more or less improper depending on the situation. We therefore extend the discussion in this section to cover topics which are not considered to be outright taboos, but are considered improper in some contexts.

13.5.3 Impropriety

The aspect of joke impropriety is often associated with the work of Freud (1905), who distinguishes *tendenziös* ('tendencious') elements in jokes, which refer to either hostility or obscenity, both of which directly relate to violations of social norms, including the norms of conversation.

In the witty remark by Mark Twain below (13s) the improper purpose of this quote – namely, to covertly communicate the insult that Wagner has no ideas is achieved by contrasting the unusual topos set up in the first sentence (the conclusion about the legality or otherwise of musical composition is not usually a consideration) and its counterpart that takes a person composing music and the legality of this as its premises and generates the conclusion that the composer has no ideas whatsoever. The humour is additionally enhanced through the contradiction of the common topos that Wagner is a great composer (and great composers usually have lots of ideas).[12]

(13s) There is no law against composing music when one has no ideas whatsoever. The music of Wagner, therefore, is perfectly legal.

According to Freud, humour arises from direct and indirect mentions of socially unacceptable topics. However, this cannot explain the amusement caused by jokes which already establish the impropriety or taboo in the set-up. Ritchie (2018, p. 145) provides a nice example of the point of the joke being concerned not with

[12] See Yablo (2014, p. 179) for discussion of how this enthymeme can be solved.

impropriety on the general level, which was already established in the set up, but on a more precise version of the set up, revealed in the punch line. Ritchie doesn't seem to think that amusement triggered by the joke can be explained by the Freudian view: 'If a topic can be mentioned in the set-up of a joke without creating humour, it is hard to see why an indirect mention should be the cause of amusement.' (ibid.)

(13t) *recited by Ritchie (2018) from Tibballs (2000)*
A woman was in bed with her husband's best friend when the phone rang. After hanging up, she turned to her lover and said: 'That was Jim, but don't worry, he won't be home for a while. He's playing cards with you.'

We agree that if one analyses the improper content – the adulterous liaison – on a general level, it should not be more amusing in the punchline than in the set-up. However, our approach provides greater granularity based on which topoi are available at different points in comprehension of the joke: (i) the setup invokes the improper topos of the adulterous wife, and (ii) the punchline invokes another improper and contrasting topos of adulterous husband through employing additional inference mechanisms enlisted in the previous sections.

Even harmless jokes like the elephant joke (13p) are somewhat reminiscent of impropriety, in the way that social norms are bent and the listener is challenged (cf. the face of the listener in Politeness theory). In this regard it is deemed important to draw a connection to work of Mazzocconi et al. (2020) on *social incongruity*, the type of incongruity that is related to non-humorous laughter (see Section 3.3). Social incongruity is closely related to incongruities discussed in this section, therefore we can claim that studies into humorous incongruities can affect not only laughables that are *pleasant incongruity* laughables which are typically associated to humour. For instance, topoi, which are evoked in humorous situations may be the same as the ones that cause embarrassment.

Ritchie (2018) distinguishes joke detection, recognition and appreciation. Detection refers to judging about the status of the text, recognition – to 'getting a joke', and appreciation results in amusement of various degrees which may lead to laughter. In the case of our framework, detection and recognition would relate to understanding the set of topoi which are required to get the joke, which can be also acquired during the joke comprehension process. In order to see a potential connection between humour and laughter as an indicator of amusement, one needs to look into the emotional appraisal of the joke content and at whether laughter is an appropriate reaction to a joke. We have already discussed this aspect in relation to deliberately withholding laughter in Section 3.6.

13.6 SUMMARY

This chapter has described the interactive approach to humour that we propose for tackling the jokes using an established approach to dialogical inferences. The general aim of such a theory is to potentially scale down to conversational humour and laughter, and furthermore signal what the laughter is about (see laughter causes in Section 3.3).

Our approach is comprised of four elements: mapping to existing rhetorical resources (topoi), acquisition and decay of topoi, integration and accommodation of topoi, and emotional appraisal. In the following Chapter 14 we will discuss all four elements of our approach in relation to one particular example of a joke.

14 Case study I. Muffins and bagels: application of the theory

14.1 INTRODUCING THE JOKE

Let us informally explain our account for the short joke (14a), using the notions that were explained in the previous chapter. This joke involves interaction, therefore it is illustrative of which elements of our approach should be taken into account in the analysis. We believe that this explores new angles, from which the jokes can be approached within a linguistic theory.

(14a) During film production in Kyiv famous Soviet actor Rolan Bykov dropped by a tiny bakery. A placid shopkeeper stands behind the counter.

1 *Customer.* Are these bagels fresh?
2 *Shopkeeper.* No.
3 *Customer.* What about the muffins?
4 *Shopkeeper.* Better get the bagels.

https://zamos.ru/humor/persona/20667/ (in Russian)

The joke relies on generally available knowledge, that fresh is better than stale when talking about bread, as well as knowledge which can be acquired from the context, which is that Soviet bakeries (as well as many other establishments of planned economy) often experience shortages.[1] Then, the listener of the joke imagines the customer's and shopkeeper's states of mind and incrementally updates them upon receiving new information from the joke teller.

[1] We will later argue that this information is not crucial for comprehending the humour.

Thus, the listener comes to the conclusion (along with several other inferences) that the shopkeeper has extremely stale muffins. This image is somewhat extreme and improper as is the shopkeeper's attitude in selling products that are no longer fresh.

In the following sections we will discuss how the four elements of the interactive approach to humour are involved in this joke.

14.2 RESOURCES

With regard to the domain set up by the joke, we can talk about a very general topos of freshness, which is applicable for most grocery products, except maybe some cheeses and the infamous Swedish example of Surströmming.[2]

The context of the joke is that A goes into a bakery, presumably to buy bread or cakes. A first asks about the freshness of the bagels. The shop assistant, B, responds that they are not fresh. A, thinking about getting muffins instead, asks whether those are fresh, and B responds that A better get the bagels. This short dialogue is underpinned by (at least) the following topoi:

1. The topos saying that if some food is fresh, one might consider buying it:

$$\frac{x\_\text{fresh}}{x\_\text{buy}} \tag{14.1}$$

2. The topos saying that if you have to choose between two food items, and one is fresher that the other, you should choose to buy the fresher one:

$$\frac{x\_\text{fresher\_than}\_y}{x\_\text{buy}} \tag{14.2}$$

[2] However, we noticed that even a single mention of this fermented herring can provoke a burst of laughs.

3. The topos saying that there is a preference for food which is fresher, basically, 'the fresher – the better':

$$\frac{x\_\text{fresher}\_\text{than}\_y}{x\_\text{better}\_\text{than}\_y} \tag{14.3}$$

4. Reversed (14.3) is also salient in the bakery domain:

$$\frac{x\_\text{better}\_\text{than}\_y}{x\_\text{fresher}\_\text{than}\_y} \tag{14.4}$$

General knowledge in the form of topoi is defeasible as can be seen with a counterexample. For instance, knowledge about freshness, expressed by the topoi above would not be alluded to in a modified version of the joke, in a setting of a Swedish grocery shop starting from the utterance 'Is this Surströmming fresh?'. This utterance of course can be considered funny but for reasons underpinned by topoi which are different from the ones underpinning the bagel joke above.

14.3 ACQUISITION AND DECAY

Topoi can have lifespans that vary. Even though the joke in (14a) can be perceived as funny without relying on resources with a shorter lifespan, such resources can act as an enhancer of a humorous effect. Even though Ritchie (2018) uses the term 'enhancers' and 'joke maximisers' for factors such as joke-teller diction, repetition and audience admiration, we can as well talk about community-specific topoi here. For the joke (14a) one such enhancer is the fact that the action takes place in the Soviet shop where:

1. The lack of food supplies could lead to a rather gloomy image of a shop with only stale muffins and bagels constituting the stock (or its leftovers after standing in a hour-long queue).

2. The attitude of the shopkeeper which can be described by a famous joke about state capitalism 'So long as the bosses

pretend to pay us, we will pretend to work'. This licenses his indifference about the state of the groceries or, if time and mood permits, his eagerness to sell stale products, therefore the joke resonates with such a personality.

14.4 INTEGRATION OF NEW INFORMATION

In this section we will draw attention to the dialogical nature of the information flow within the joke (14a). After the first utterance the customer has communicated that she considers buying some bagels and that the freshness of the bagels will have an impact on her willingness to buy them. When the shopkeeper replies 'no', we know that the bagels are not fresh, and naturally, the customer starts inquiring about the freshness of other baked goods. Then, to the customer's surprise (as well as to the joke-listeners), the shopkeeper advises to her buy bagels that, as he just has claimed, are not fresh.

The key inference here is that *the muffins are extremely stale*, but there are also a couple of optional ones, which can be counted as enhancers:

- There is nothing fresh left in this shop.
- The shopkeeper offers something which is not good for sale.

The latter inference we get 'for free', as one of the steps towards making the key inference. It is ambiguous whether the shopkeeper actually makes an offer, as can be perceived from the surface form of his utterance, or is just being sarcastic about the miserable condition of the muffins.

In Table 14.1 we sketch how key inferences could be made. Each of the enlisted premises might have been produced explicitly in the dialogue or drawn from the rhetorical resources.

We can assume that a topos along the lines of 'don't buy non-fresh food' is accommodated in the dialogue. If the shopkeeper had not agreed with this, they would have said something like 'they are not fresh, but they are actually best when they are a few days old',

		Customer. Are these bagels fresh?
	x_is_a_customer	
	b_is_a_bagel	explicit
		Shopkeeper. No.
	y_is_a_shopkeeper	explicit
α	b_is_stale	explicit
		Customer. What about the muffins?
	m_is_a_muffin	explicit
		Shopkeeper. Better get the bagels.
β	b_is_better_than_m	
γ	b_is_fresher_than_m	from β and (14.4)
	b_is_offered_by_y	from γ and a topos similar to (14.2)
	m_is_very_stale	from α and γ

Table 14.1: Inferences in the 'muffins' joke

or similar. The second exchange evokes the topos that if one food item is fresher than another, you should buy the fresher one. Both of these topoi seem acceptable, and most people would accept them. However, in this case, these topoi are accommodated which, when instantiated *in this particular context*, lead to inconsistent conclusions ('buy fresher bagels' and 'don't buy non-fresh bagels'). We believe that this is a more fine-grained way to account for Raskin's notion of script opposition discussed in Section 12.2.

14.5 APPRAISAL

Sometimes we can say that the combination of premises and conclusions can be emotionally appraised and this is what produces the effect, resulting in laughter or amusement, or otherwise rejecting the joke as 'gross' or 'inappropriate'.

Figure 14.1: Bakery department of a Soviet shop. The inscription reads 'Don't touch!'. (Peter Turnley, 1990)

We argue that all these inferences can be appraised as something improper.

The muffins are extremely stale. The unpleasant image of stale products can be deemed as improper. There is even a variant of the joke where the shopkeeper not only communicates that the bagels are not fresh, but specifies their condition even further ('old, stale and mouldy', in one of alternative versions of the joke).[3]

The shopkeeper offers something which is not good for sale. This is a socially unacceptable action in most cultures, even though a salesperson might trick a customer, it is still not appreciated and is the subject of many jokes. For instance, Jewish jokes[4] evoke

[3]https://www.anekdot.ru/id/-120214070/ (in Russian), bagels are substituted by *prianik*, traditional Russian sweet bread.

[4]Rolan Bykov was of a Jewish origin, and there are a few instances of this joke being recast into a Jewish joke.

a *canny* image of their characters (Davies, 2011), which might be perceived as a good quality (i.e. sharp-witted), but not necessarily positive (i.e. crafty).

There is nothing fresh left in this bakery. This inference is optional and wouldn't be evoked in all cases. Such a gloomy image clashes with typical expectations from a bakery (at least in the West). Even though in Soviet times this hasn't been an entirely unlikely image, all the preceding dialogue in the joke gives the impression that initially the customer at least had some choice.

14.6 SUMMARY

In this chapter we exemplified our interactive account through analysis of a joke which is also a short dialogue. The joke therefore requires the listener to take different perspectives of its participants and understand the implicit assumptions that underpin it. In regard to integration process we have only sketched the step-by-step inference process without any precise formalisation of the dialogue context. We will give a formal account of such a process in the next chapter using another example.

15 Case study II: The elephant riddle or further analysis of the integration process

The content of this chapter was previously published in Breitholtz and Maraev (2019).[1] *It has gone through minor revisions.*

15.1 THE ELEPHANT RIDDLE

In this chapter we give a formal account of the integration of new information during the course of a dialogue in the elephant-in-the-fridge riddle, which was first mentioned in Section 13.4 and is restated below.

(13p)
1 *A.* How do you put an elephant into a fridge?
2 *B.* Hmm, I don't know?
3 *A.* Open the door, put the elephant inside, close the door.
4 *B.* Haha okay
5 *A.* How do you put a giraffe into the fridge?
6 *B.* Open the door, put the giraffe inside, close the door?
7 *A.* Wrong! Open the door, get the elephant out, put the giraffe inside, close the door.

constructed example

[1] Ellen Breitholtz and Vladislav Maraev (2019). 'How to Put an Elephant in the Title: Modeling Humorous Incongruity with Topoi'. In: *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. London, United Kingdom: SEMDIAL. Both authors contributed to the discussion and research of examples. Breitholtz and Maraev wrote the parts of the original paper that appear here.

We will start by introducing the theoretical framework that we use here (Section 15.2), then we will explain how this framework can be used to model the process of accommodation and integration of new information (Section 15.3). We will then use these tools to analyse this particular joke (Section 15.4). We conclude with a discussion in Section 15.5.

15.2 THEORY OF TYPES WITH RECORDS

The formal framework we use is Theory of Types with Records (TTR), a rich type theory successfully employed to account for a range of linguistic phenomena, including ones particular to dialogue (Cooper and Ginzburg, 2015).

In TTR, agents perceive an individual object that exists in the world in terms of being *of a particular type*. Such basic judgements performed by agents can be denoted as $a : Ind$, meaning that a is an individual, in other words a is a *witness* of (the type) *Ind*(ividual). This is an example of a *basic* type in TTR, namely types that are not constructed from other types. An example of a more complex type in TTR is a *ptype* which is constructed from predicates, e.g., $disturb(a, b)$, 'a disturbs b'. A witness of such a type can be a situation, a state or an event. To represent a more general event, such as 'one individual disturbs another individual' *record types* are used. Record types consist of a set of fields, which are pairs of unique labels and types. A record type which corresponds to the aforementioned situation is the following:

$$\begin{bmatrix} \mathrm{x} & : & Ind \\ \mathrm{y} & : & Ind \\ \mathrm{c}_{\mathrm{disturb}} & : & disturb(\mathrm{x}, \mathrm{y}) \end{bmatrix} \tag{15.1}$$

The witnesses of record types are *records*, consisting of a set of fields which are pairs of unique labels and values. In order to be of a certain record type, a record must contain at least the same set of labels as the record type, and the values must be of a type mentioned in the corresponding field of the record type. The record may contain additional fields with labels not mentioned in

the record type. For example, the record (15.2) is of a type in (15.1) iff $a : Ind$, $b : Ind$, $s : disturb(a, b)$ and q is of an arbitrary type (for instance, degrees Celsius).

$$\begin{bmatrix} \mathrm{x} & = & a \\ \mathrm{y} & = & b \\ \mathrm{c}_{\mathrm{disturb}} & = & s \\ \mathrm{c}_{\mathrm{temperature}} & = & q \end{bmatrix} \tag{15.2}$$

TTR also defines a number of type construction operations. Here we mention only the ones that are used in this chapter.

1. *List types*: if T is a type, then $[T]$ is also a type – the type of lists each of whose members is of type T. Additionally, we use a type of non-empty lists, written as ${}_{ne}[T]$, which is a subtype of $[T]$ where $1 \leqslant i \leqslant n$. The list $[a_1, \ldots, a_n] : [T]$ iff for all i, $a_i : T$. We assume the following operations on lists: constructing a new list from an element and a list (cons), taking the first element of list (head), taking the rest of the list (tail).

2. *Function types*: if T_1 and T_2 are types, then so is $(\lambda r : T_1.T_2)$, the type of functions from records of type T_1 to record type T_2. Additionally, T_2 may *depend* on the parameter (the witness of type T_1 passed to the function).

3. *Singleton types*: if T is a type and $x : T$, then T_x is a type. $a : T_x$ iff $a = x$. In record types we use manifest field notation to represent a singleton type. Notations $\begin{bmatrix} \mathrm{a} & : & T_x \end{bmatrix}$ and $\begin{bmatrix} \mathrm{a} = \mathrm{x} & : & T \end{bmatrix}$ represent the same object.

15.3 FORMAL ACCOUNT OF THE INTEGRATION PROCESS

Similarly to what we have employed in previous chapters, following Ginzburg (2012) and Larsson (2002), we model the progress of dialogues in terms of the *information states* of the dialogue participants. In our analysis we focus on the part of a dialogue participant's information state that is considered to be shared. That is, what has in

some way been referred to in the dialogue, or what it is necessary to integrate in the information state for a dialogue contribution to be interpreted in a relevant way. As mentioned in Chapter 4, we refer to this shared part of an interlocutor's information state as the Dialogue GameBoard (DGB) of that participant. We are particularly interested in how individual agents draw on individual (and sometimes distinct) resources. We will therefore use separate DGBS for each agent, rather than letting the DGB represent a God's eye notion of context. For example, although a topos may be of central relevance in the dialogue, it does not appear on the DGB until it has been made explicit, or until something has been said which has caused it to be accommodated. We model the DGB as a record type where labels are associated with types, as in (15.3).

$$\begin{bmatrix} \text{rhet\_resources} : \begin{bmatrix} \text{topoi} : [\textit{Topos}] \end{bmatrix} \\ \text{dgb} : \begin{bmatrix} \text{eud} : [\textit{Enthymeme}] \\ \text{topoi} : [\textit{Topos}] \end{bmatrix} \end{bmatrix} \tag{15.3}$$

The record type in (15.3) represents the type of the information state of a dialogue participant with regard to enthymematic reasoning. In the DGB we find the enthymemes under discussion ('eud') and the topoi that have been evoked in the conversation. For a topos to be added to the DGB of a dialogue participant, it must have been accommodated by the participant. The field 'rhet_resources' (rhetorical resources) represents the topoi that are available to a speaker for inventing and interpreting arguments, it constitutes a part of the information state which is private.

15.4 RHETORICAL REASONING IN THE ELEPHANT RIDDLE

We model enthymematic inferences and the topoi that underpin them as functions from situations of particular types to other types of situation. For example, one topos relating to the situation described in the elephant dialogue in (13p), could be represented as a function from a situation of a type where someone opens the door of the fridge, puts an object inside, and shuts the door, to a type of

situation where the same object is in the fridge. We see this topos, τ_1, in (15.4):

$$\tau_1 = \lambda r : \begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y},\mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y},\mathrm{z},\mathrm{x}) \\ \mathrm{c}_{\mathrm{small}} & : & \mathit{small}(z) \\ \multicolumn{3}{c}{\begin{bmatrix} \mathrm{s} & : & \mathit{in}(r.\mathrm{z}, r.\mathrm{x}) \end{bmatrix}} \end{bmatrix}. \tag{15.4}$$

A topos is to be seen as a non-monotonic principle of reasoning (Breitholtz, 2020), and as such the conclusion does not follow necessarily and in all cases. Just like the principle that if x is a bird, x flies, does not apply to situations where the bird in question is a penguin, there might be a number of situations where a topos about how food gets into a fridge does not apply. Relevant to the situation at hand is an exception regarding the size of the object. Thus, we include the constraint 'small' to restrict the use of the topos to things that are small enough to fit into a fridge. τ_1 is part of B's rhetorical resources, that is, a collection of topoi that are available for B to use as warrants in reasoning. The situation suggested by A's question conveys an enthymeme ε_1 like that in (15.5).

$$\varepsilon_1 = \lambda r : \begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{elephant}} & : & \mathit{elephant}(\mathrm{z}) \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y},\mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y},\mathrm{z},\mathrm{x}) \\ \multicolumn{3}{c}{\begin{bmatrix} \mathrm{s} & : & \mathit{in}(r.\mathrm{z}, r.\mathrm{x}) \end{bmatrix}} \end{bmatrix}. \tag{15.5}$$

To integrate a topos which underpins an enthymeme under discussion, the topos accessed in the rhetorical resources of the dialogue participant must be relevant with regard to the enthymeme conveyed in the discourse. We define this as the enthymeme being a *specification* of the topos. An enthymeme ε is a specification of a topos τ iff the antecedent type of ε is a subtype of the antecedent type of τ, and, for any situation r, the result of applying ε to r, is a subtype of the result of applying τ to r as shown in (15.6).

$$\begin{array}{ll} \tau & = T_1 \rightarrow T_2 \\ \varepsilon & = T_3 \rightarrow T_4 \\ T_3 \sqsubseteq T_1 & \\ \text{for any } r, \varepsilon(r) \sqsubseteq \tau(r) & \end{array} \tag{15.6}$$

However, since the antecedent type of τ_1 involves a constraint 'small', which is not present in the antecedent type of ε_1, ε_1 is not a specification of τ_1. Interlocutor B does not have access to other relevant topoi regarding how you put an elephant into a fridge, and replies that he does not know the answer to the question.

A's next utterance evokes another topos τ_2 where the size constraint is removed, and the enthymeme under discussion is thus a specification of τ_2, which is integrated in B's DGB according to the update rule in (15.8) below. This rule takes an information state bound to r and an additional specification constraint bound to e.

$$\tau_2 = \lambda r : \begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y}, \mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y}, \mathrm{z}, \mathrm{x}) \\ \left[\mathrm{s} \ : \ \mathit{in}(r.\mathrm{z}, r.\mathrm{x}) \right] & & \end{bmatrix} . \tag{15.7}$$

$$\mathcal{F}_{integrate\_shared\_topos} =$$
$$\lambda r : \begin{bmatrix} \text{rhet\_resources} & : & \begin{bmatrix} \text{topoi} & : & [\mathit{Topos}] \end{bmatrix} \\ \text{dgb} & : & \begin{bmatrix} \text{eud} & : & [\mathit{Enthymeme}] \\ \text{topoi} & : & [\mathit{Topos}] \end{bmatrix} \end{bmatrix} \cdot$$
$$\lambda e : \begin{bmatrix} \text{t} & : & \mathit{Topos} \\ \text{c}_1 & : & r.\mathit{rhet\_resources.topoi}(\text{t}) \\ \text{c}_2 & : & \mathit{specification}(\mathsf{fst}(r.\text{dgb.eud}), \text{t}) \end{bmatrix} \cdot$$
$$\begin{bmatrix} \text{dgb} & : & \begin{bmatrix} \text{topoi} = \mathsf{cons}(e.\text{t}, r.\text{dgb.topoi}) & : & [\mathit{Topos}] \end{bmatrix} \end{bmatrix} \tag{15.8}$$

A then moves on to the second punchline of the joke, asking how to fit a giraffe into the fridge. Which enthymeme is under discussion at this point is not obvious – *B* could interpret the situation in (at least) two ways. Either, the question is how to fit a giraffe into any fridge, or into the fridge that is already occupied by the elephant. On either of these interpretations, the enthymeme under discussion ε_2 in (15.9) is similar to ε_1, with the exception that the individual z is associated with the constraint 'giraffe' rather than 'elephant'. Alternatively, an individual is added which is associated with the constraint 'giraffe' without any other individual or constraint being removed.

$$\varepsilon_2 = \lambda r : \begin{bmatrix} \text{x} & : & \mathit{Ind} \\ \text{y} & : & \mathit{Ind} \\ \text{z} & : & \mathit{Ind} \\ \text{c}_{\text{giraffe}} & : & \mathit{giraffe}(\text{z}) \\ \text{c}_{\text{fridge}} & : & \mathit{fridge}(\text{x}) \\ \text{c}_{\text{agent}} & : & \mathit{agent}(\text{y}) \\ \text{c}_{\text{open}} & : & \mathit{open}(\text{y}, \text{x}) \\ \text{c}_{\text{put}} & : & \mathit{put\_in}(\text{y}, \text{z}, \text{x}) \\ \begin{bmatrix} \text{s} & : & \mathit{in}(r.\text{z}, r.\text{x}) \end{bmatrix} & & \end{bmatrix} . \tag{15.9}$$

However, since the size constraint is now gone, it should not matter. *B*'s DGB now looks like this:

$$\begin{bmatrix} \text{dgb} : \begin{bmatrix} \text{eud} = [\varepsilon_2, \varepsilon_1] & : [\mathit{Enthymeme}] \\ \text{topoi} = [\tau_2] & : [\mathit{Topos}] \end{bmatrix} \end{bmatrix} \tag{15.10}$$

B evaluates whether the enthymeme ε_2 is underpinned by the topos already integrated in the DGB, and since the addition of a giraffe, including or excluding the elephant, does not matter since the size restriction from τ_1 is dropped in τ_2, this means that τ_2 can be used to warrant ε_2 (15.11). B thus replies, in accordance with this reasoning, that you behave in the same way to put a giraffe into a fridge as you do with an elephant, that is, you open the door, put the giraffe in, and close the door.

$$\begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{giraffe}} & : & \mathit{giraffe}(\mathrm{z}) \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y},\mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y},\mathrm{z},\mathrm{x}) \end{bmatrix} \sqsubseteq \begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y},\mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y},\mathrm{z},\mathrm{x}) \end{bmatrix} \tag{15.11}$$

A takes advantage of the fact that B draws on the topos on B's DGB, τ_2. However, A's final punchline evokes a third topos, τ_3 (15.12), which introduces a new constraint regarding the ability of an elephant and a giraffe to be in the fridge at the same time, possibly some kind of size restriction. Which is of course incongruous in relation to B's previous information state. Thus, taking advantage of the set up of B's DGB at each exchange in the dialogue, A is able to create mismatches in B's DGB, making use of at least one of his available topoi, τ_3, (see the end of Section 13.4.3 for other possible topoi which challenges B's τ_2). In the case that A would be asked to justify this punchline, the answer could be along the following: 'The fridge is huge but not enormous enough to fit two big animals'.

$$\tau_3 = \lambda r : \begin{bmatrix} \mathrm{x} & : & \mathit{Ind} \\ \mathrm{y} & : & \mathit{Ind} \\ \mathrm{z} & : & \mathit{Ind} \\ \mathrm{c}_{\mathrm{size}} & : & \mathit{huge\_not\_enormous}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{fridge}} & : & \mathit{fridge}(\mathrm{x}) \\ \mathrm{c}_{\mathrm{agent}} & : & \mathit{agent}(\mathrm{y}) \\ \mathrm{c}_{\mathrm{open}} & : & \mathit{open}(\mathrm{y}, \mathrm{x}) \\ \mathrm{c}_{\mathrm{put}} & : & \mathit{put\_in}(\mathrm{y}, \mathrm{z}, \mathrm{x}) \\ \left[\mathrm{s} \ : \ \mathit{in}(r.\mathrm{z}, r.\mathrm{x}) \right] & & \end{bmatrix} . \tag{15.12}$$

15.5 DISCUSSION

The aim of this chapter was to illustrate how the integration process required for joke processing in dialogue situations can be formally explained by means of enthymemes and topoi. Crucially, our example highlights the asymmetry between the joke teller and joke listener in terms of availability of particular rhetorical resources to them.

The scope of this study was limited by using constructed examples and abstracting away from real dialogue data. A further study with more focus on data from spoken language corpora is therefore suggested.

For this study we used TTR, which is an established framework for expressing rhetorical resources within the dynamics of dialogue (Breitholtz, 2020). In the future we may consider casting the aforementioned definitions in the Linear Dialogue Manager (LDM) implementation which we use in this thesis and which has rudimentary support for enthymematic reasoning, which we discuss next in Chapter 16.

The current study indicates the importance of having resources such as topoi, that enable an agent to reason using non-logical arguments, for building future dialogue systems with a capability to recognise and understand humour. We come to this topic in the final chapter.

16 Implications for dialogue systems

Sections 16.2 and 16.4.1 of this chapter were previously published in Maraev, Breitholtz and Howes (2020).[1] *Section 16.5 was previously published in Maraev, Breitholtz, Howes and Bernardy (2021).*[2] *They have been substantially revised.*

16.1 THE AIM OF THIS CHAPTER

In this chapter we give some further directions for developing conversational humour for spoken dialogue systems in accord with our theory which we introduced in Chapter 13. First we will briefly observe the current state-of-the-art in computational humour and then turn to our suggestions regarding acquiring and maintaining rhetorical resources for a dialogue system.

16.2 COMPUTATIONAL HUMOUR

A considerable amount of literature has been published on computational humour, highlighting the importance of understand-

[1] Vladislav Maraev, Ellen Breitholtz and Christine Howes (2020). 'How do you make an AI get the joke? Here's what I found on the web'. In: *First AISB Symposium on Conversational AI (SoCAI)*. Maraev planned the study and wrote the parts of original paper that appear here with input from the other authors.

[2] Vladislav Maraev, Ellen Breitholtz, Christine Howes and Jean-Philippe Bernardy (2021). 'Why Should I Turn Left? Towards Active Explainability for Spoken Dialogue Systems.' In: *Proceedings of the Reasoning and Interaction Conference (ReInAct 2021)*. Gothenburg, Sweden: Association for Computational Linguistics, pp. 58–64. Maraev initiated and planned the study. Original manuscript was written by Maraev with input from the other authors.

ing humour for dialogue systems (e.g., Raskin and Attardo, 1994; Hempelmann, 2008; Binsted et al., 1995).

A number of authors have investigated *humour generation*, mainly using template-based approaches inspired by humour theories. Examples of generated humorous texts are puns (Ritchie, 2005), lightbulb jokes (Raskin and Attardo, 1994), humorous names (Ozbal and Strapparava, 2012) and acronyms (Stock and Strapparava, 2005).

Much of the current literature on *humour recognition* pays particular attention to either detecting salient linguistic features, such as stylistic features (Mihalcea and Strapparava, 2005), handcrafted humour-specific features (Zhang and Liu, 2014) and N-gram patterns (Taylor and Mazlack, 2004), or latent semantic structures (Taylor, 2009; Yang et al., 2015). Yang et al. (2015), in addition, focus on humour anchors, i.e. words or phrases that enable humour in a sentence. More recent approaches (e.g., Chen and Soo, 2018) use neural networks to recognise humour. There have also been substantial work on the detection and analysis of puns (e.g., Miller et al., 2017). Chowdhery et al. (2022) introduced a Transformer-based neural model and showcased its capacity to generate explanations when prompted with a joke.

So far, however, there has been little discussion about detecting humour in an interactive setting. For example, recent studies were mostly concerned with scripted dialogues, such as TV series like 'Friends' and 'The Big Bang Theory'. Purandare and Litman (2006) used both prosodic and linguistic features, Bertero and Fung (2016) used a text-based deep learning approach and Patro et al. (2021) combine features from video and text in multi-modal self-attention based models. All these studies presuppose that the chunks of dialogues which are followed by laughs are humorous, and the rest are non-humorous. The main weakness of this approach is that in natural dialogues laughter is not necessarily associated with humorous content (see Chapter 3), in contrast to scripted comedy shows. In addition to this, not all events that are perceived as humorous provoke laughter. Even though laughter in conversations can be predicted with a fairly high accuracy (see Chapter 7), it is

still not indicative of whether the preceding content was humorous as opposed to, for example, the laughter having been used to soften a bold opinion expressed by of one of the interlocutors.

In regard to human-computer interaction (HCI) studies, Shah and Warwick (2017) presents some evidence that in Turing test experiments, failed attempts at humour can indicate behaviour which is judged as machine behaviour, even if humour is typically produced by real humans. This indicates some shared characteristics between an irrelevant conversational contribution typical to a machine and a humorous contribution from a human. Dybala et al. (2010) emphasises the importance of the 'two-stage' approach to humour in dialogue systems, where the system tracks the emotional state of the user, produces humour as a reaction to certain states, and analyses the user's further emotional reaction. Additionally, several studies (e.g., Morkes et al., 1999; Dybala et al., 2008; Dybala et al., 2009) show that dialogue systems equipped with humorous capabilities influence human-computer dialogue in a positive way. In these studies, humour is considered in relation to jokes or humorous comments generated by the dialogue system, but here we wish to draw attention to more subtle humorous incongruities.

16.3 HUMOROUS LAUGHTER IN DIALOGUE SYSTEMS

Consider the following task-oriented dialogue taken from the Directory Enquiries Corpus (DEC) (Bondarenko et al., 2020) (see Chapter 11 in for further details about the corpus):

(16a)
17 *Caller.* okay so it starts with a
18 *Caller.* L
19 *Operator.* L?
20 *Caller.* as in london
21 *Operator.* yes
22 *Caller.* A as in america
23 *Operator.* america
24 *Caller.* er U
25 *Caller.* as in er ((pause: 1.2s))

26 *Caller.* er under
27 *Caller.* <laugh>
28 *Operator.* under yes

DEC 28_NM_loc2

In this excerpt the caller experiences issues with coming up with phonetic spellings for certain letters. The first laugh (line 27) deserves attention, as it seems that it reflects on both a pleasant incongruity and a social one (smoothing), according to the taxonomy of Mazzocconi et al. (2020). The pleasant incongruity is due the fact that the phonetic spelling of 'U' as in 'under' is incongruous with the preceding ones: a preposition vs. proper nouns.

We will now consider how a spoken dialogue system developer and designer could use the elements of our theory (Chapter 13), while thinking about how this short exchange can be implemented.

Existing resources The dialogue system needs to maintain a set of resources capturing ways of phonetic spelling.

Acquisition and decay of the resources The system should allow the user to come up with a way of spelling words out and teach this to a system. For instance, it might be names of the rivers (or greek letters, or Star Wars characters). When this spelling out convention is established, it can also be violated in a humorous way. It is a question for further research whether the violation of more established conventions would appear to be more humorous.

Integration of new information This relates to describing the conceptualisation of certain behaviours, or how the new way of spelling out the name turns into a (local) convention which can be humorously subverted later on. This is the mechanism which is potentially generalisable to other conversational conventions.

Emotional appraisal Some city names are 'inappropriate', or have some 'interesting' connotations. P as in Pyongyang – would

that be funny? Or B as in Bognor?[3] This is another type of incongruity which is different from the one that violates established conventions.

In the next section we will address the aspects related to rhetorical resources for dialogue systems. We resist the temptation to touch upon the other aspects of our theory in order to keep the thesis within bounds.

16.4 RHETORICAL RESOURCES

16.4.1 Acquiring the resources

In order for a dialogue system to be able to understand humour (whether in joke form or just in general humorous interactions), we therefore believe that they need the same resources as required to understand human reasoning in dialogue (since jokes are often underpinned by general socio-cultural topoi, see Section 12.3); that is a library of topoi (Breitholtz and Cooper, 2011; Breitholtz, 2014; Breitholtz, 2020). In order to acquire such a resource, we propose to mine enthymemes from a variety of sources.

In everyday interactions, people often reason in the form of enthymemes:

(16b) 10554 *Dave.* …you're gonna be home from football until four, you gonna have your dinner, want a bath.
10555 *Lee.* Yeah, but I might not go to school tomorrow.
10556 *Dave.* Why?
10557 *Lee.* Cos of my cough.
10558 *Dave.* How can you play football and not go to school then?
10559 *Lee.* Cos I was going out in the fresh air, I'm alright when I'm out in the fresh air.
10560 *Dave.* So why aren't you going to school then?
10561 *Lee.* I'm in the class room all day dad.

BNC KBE

[3] 'Bugger Bognor', as King George V was alleged to have said on his deathbed.

In (16b) one enthymeme is constituted by Lee's assertion that he might not go to school and the reply to David's 'why?'-question, introduced by 'cos' (because). Why-questions are known to invoke enthymemes in dialogue, offered as reasons (Schlöder et al., 2016), and we can exploit this fact to look for enthymemes in spoken dialogue corpora such as the British National Corpus (BNC). Schlöder et al. (2016) report that the spoken dialogue portion of the BNC contains 2256 why-questions, and using the same search tool, SCoRE (Purver, 2001) shows 4972 utterances in the same corpus start with 'because' or 'cos'.

In text-based resources, such as Wikipedia, or online reviews (which have previously been used for reconstructing enthymemes by Rajendran et al., 2016) and Reddit, we can rely on different structures and keywords to search for potential enthymematic arguments (for example, 'therefore', 'since', 'once', 'so').

The following steps can constitute the process of mining enthymemes:

1. dependency parsing and pattern-based extraction of enthymeme candidates based on their surface structure

2. annotation, whether or not the extracted structure is an enthymeme and annotation of the premise(s) and the consequence(s) of it.

3. enthymeme classification (for example, keywords like 'since' can just relate to a time frame)

4. enthymeme parsing, that will lead to a semantic representation of an enthymeme (assuming that we have a semantic parser which is equipped for this task)

Importantly, many common enthymemes in everyday use (such as 'if one is ill one should not go to school' employed by Lee in (16b)) are generally considered to be common knowledge and do not need to be spelled out when they are invoked in dialogue (even though they are often culturally specific). However, even these topoi

must be learned by children, and a search in the CHILDES corpus[4] (MacWhinney, 2000), shows a distinct peak of 'why?'-questions in children at around age 3 (Breitholtz and Howes, 2020), and there is evidence that children's why-questions promote explanations (Bova and Arcidiacono, 2013) and involve sentential answers (Moradlou et al., 2021). Thus, we hypothesise that we may be able to mine enthymemes based on common knowledge from a corpus of child-directed speech, such as CHILDES, using the same methods as described above.

We hypothesise that a great amount of data that is needed for interpreting jokes can be extracted from ontologies, especially ones that aim at collecting data for improving general knowledge of computers, such as ConceptNet (Speer et al., 2017). The importance of ontologies was emphasised by the proponents of Ontological Semantic Theory of Humour (OSTH) (Raskin, 2017). However, we do not consider them a last resort, but rather a core component of understanding humour and other dialogical reasoning, such as explaining the mismatches between the arguments presented in (16b).

Following this, extracted enthymemes can be clustered in order to induce more general rules, or relate to one of the already described topoi, such as the Aristotelian topos of 'the more and the less' (Breitholtz, 2020). The gist of this topos is that a small thing is contained in a large thing – for example, if you can build a castle you can build a cottage, or if you can run a marathon then you can run a half marathon.

16.4.2 Using the resources

Having rhetorical resources available can help dialogue systems understand humorous contributions by the user and act (laugh, for instance) accordingly. Moreover, this should allow dialogue systems to generate humour from the resources which are known to them. One can adapt a number of rules representing topoi that

[4] Using ChildFreq (Bååth, 2010)

could enrich the current information-state of a dialogue system. For instance:

- prefer things that are fresher
- prefer things which are better
- prefer things which are cheaper
- more expensive things are better

All of these topoi can meet a counterexample which has the potential of subverting them in a humorous way.

- Some things are not good when they are fresh, therefore utterances like 'this wine is straight from the vineyard' and 'take these pickles, they are very fresh' can be deemed as mildly humorous.
- Rolls Royce is a good car, but it would be a bad advice to recommend it to a person with a middle class income who is struggling to choose a family station wagon car. 'This Rolls Royce is nice and spacious with room for the whole family – you should buy it' would be quite an ironic recommendation.
- Preferences for either more or less expensive products can be subverted, if the price difference is miniscule. The justification of recommended product based on a price difference of 1 cent seems ridiculous.[5]

A subverted statement can be treated as an enthymeme which clashes with a salient topos. The precise definition of a clash still remains to be tested, but the account of Ginzburg et al. (2015) can be taken as a point of departure. The question of how a dialogue system

[5]The author once overheard a conversation between two wealthy-looking mid-school boys in the pharmacy in a rich area in *Podmoskovye*. 'You should take this *hematogen* (nutrition bar), it is one ruble more expensive!' (1 ruble was roughly 1 euro cent).

can generate such clashing enthymemes remains unanswered at present.

Considerably more work needs to be done to describe how to choose the most salient topos from the available resources. A reasonable approach to tackle these issues could be to employ Bayesian networks, following Maguire (2019) who combines them with topoi to represent world knowledge in order to model conditionals. Another approach would be to seek approaches within the field of general artificial intelligence where there are accounts for reasoning under uncertainty and insufficient resources (e.g., Wang et al., 2018).

User preferences can be cast in topoi, such as preferences for quality over price (for online shopping) or scenery over efficiency (for car navigation, see Section 16.5). A dialogue system that meticulously follows one of the topoi could produce 'unintentional' humour, justifying its preference for a product by 1 cent price difference. At the same time it can be used to 'mock' the user preferences, i.e. highlighting one's preference for cheaper products by offering something that is negligibly cheaper – 'You generally prefer products which are cheaper – here's your yoghurt x which is 1 cent cheaper than yoghurt y'.

To develop a full picture of humour in dialogue systems, additional studies will be needed that would implement other aspects of our theory: decay of a rhetorical resource, accommodation and emotional appraisal.

16.5 EXTENDING LDM WITH RHETORICAL RESOURCES

In Chapter 10 we presented a Linear Dialogue Manager (LDM) which we used as a platform for implementing the theoretical notions that we discuss in this thesis. With respect to rhetorical resources we show how LDM can be extended to reason using topoi. It appears to be easier to discuss the extension of the system with a simple example, adapted from Breitholtz (2020). This example does not involve any humour but merely shows how enthymematic reasoning can be tackled in a dialogue manager (DM).

(16c) 1 *User.* How can I get home?
2 *System.* Via the bypass.
3 *User.* Why the bypass?
4 *System.* Because this route is the shortest.

constructed example

For dealing with lines 1–2 of example (16c), let's assume that system has access to the following facts from the knowledge base which represent three possible routes to home, via three different roads.

$$\_ :: \mathit{Route\ Bypass\ Home};$$
$$\_ :: \mathit{Route\ ParkLane\ Home};$$
$$\_ :: \mathit{Route\ BridgeRoute\ Home};$$

Assuming that the question under discussion is $(\mathit{Q\ USER\ Road\ x\ (Route\ x\ Home)})$, and the choice of hypothesis is pseudo-random, we can see that x unifies with any of three facts, therefore using the *produceAnswer* rule, the system can produce a short answer $(\mathit{ShortAnswer\ Road\ Bypass\ SYSTEM\ USER})$, which can be realised as 'Via the bypass'.

Now let us turn to the argumentative part of the dialogue. We would need to use a domain specific representation of a question by adding an additional predicate: $(\mathit{Q\ USER\ Road\ x\ (Pick\ (Route\ x\ Home))})$. The knowledge base can also be extended with some additional facts about the qualities of the routes.

$$\_ :: \mathit{Shortest}\ \ (\mathit{Route\ Bypass\ Home});$$
$$\_ :: \mathit{Cheapest}\ (\mathit{Route\ ParkLane\ Home});$$
$$\_ :: \mathit{Prettiest}\ \ (\mathit{Route\ BridgeRoute\ Home});$$

To represent enthymematic reasoning (lines 3–4 of (16c)), we will introduce the reasoning pattern represented by the following rule:

$$\mathit{toposShortest} : (x : \mathit{Road}) \rightarrow (y : \mathit{To}) \rightarrow$$
$$(\mathit{qs} : \mathit{List\ Question}) \rightarrow$$

$$
\begin{array}{l}
QUD\ (Cons\ (Q\ USER\ Road\ x \\
\quad (Pick\ (Route\ x\ y)))\ qs) \rightarrow \\
Route\ x\ y \rightarrow \\
Shortest\ (Route\ x\ y) \rightarrow \\
\quad [\_ :: Pick\ (Route\ x\ y); \\
\quad \ \_ :: Topos\ (Shortest\ (Route\ x\ y));];
\end{array}
$$

This can be read as follows: 'In the context of a question under discussion, involving picking a route, pick the shortest one, and remember why it was picked'. The latter is represented in the last line and alludes to the salient *topos* used for this choice. Note, that here we leave the destination underspecified, and further underspecifications are possible: not only shortest routes might be preferred but also shortest times or sentences.

Following Breitholtz (2020) we treat why-questions as questions asking for a topos, which become the question under discussion ($Q\ USER\ Reason\ t\ (Topos\ t)$) where t is a metavariable representing the reason for choosing the bypass. With a local topos produced by *toposShortest* rule at hand we can apply the standard *produceAnswer* rule, which would elicit a short answer:

$$
\begin{array}{l}
ShortAnswer\ Reason \\
\quad (Shortest\ (Route\ Bypass\ Home)) \\
\quad\quad SYSTEM\ USER
\end{array}
$$

It can be realised as an utterance 'Because it is the shortest', concluding our example (16c).

Our system supports several competing topoi, for instance we can analogously add *toposPrettiest* and *toposCheapest* rules. Assuming random selection of an applicable rule (only one topos is selected), the system will be able to offer a justification for whichever of the routes it chooses, based on the underlying topos.

Taking into consideration the caveats discussed in Section 16.4.2, we see in our implementation the potential to be extended to support handling the rhetorical resources needed for calculating humorous incongruities.

Part V

Conclusions

17 Concluding remarks

17.1 CONTRIBUTIONS

This thesis set out to study laughter in a way that is applicable to spoken dialogue systems. Our main principle was to base our investigations in dialogue theory, so that it would be mutually fruitful for both theoretical and practical purposes. We hope that our studies have laid the ground work for further understanding of laughter and for integrating it into spoken dialogue systems.

We divided this thesis in three main areas of concern and summarise our respective contributions below.

17.1.1 Laughter placement

After drawing attention to different patterns of placing laughter in regard to laughables, we looked at laughter predictability by neural networks. We showed that convolutional neural networks (CNNS), recurrent neural networks (RNNS), and their combinations can effectively predict laughs from dialogue transcripts. Moreover, we gave the same task to humans at Amazon Mechanical Turk (AMT) and demonstrated that neural networks performed significantly better than untrained humans. This might show that people do not have 'metalinguistic' or 'metapragmatic' intuitions regarding why they laugh at particular junctures while nevertheless following regular patterns in the production of such behaviours. It raises a lot of questions about Gricean pragmatics and the role of psychological constructs like intentions in language processing (see e.g., Gregoromichelaki et al., 2011; Stone, 2004; Lepore and Stone, 2014).

Models like ours can be useful for the components of dialogue systems that involve language modelling, such as automatic speech recognition (ASR), natural language understanding (NLU) and natural language generation (NLG). It allows them to predict user

laughter and, if needed, put system laughter in an appropriate place.

For a dialogue system involving an avatar or other multimodal behaviour realisation, such as an embodied conversational agent (ECA), it is important to be aware of laughter placement in connection to other behaviours. Therefore, we designed a study to determine gaze patterns in association with laughter. We showed that gaze is crucial for laughter coordination. In particular, for the case of co-active laughter we observed that gaze is used to synchronise the onsets and offsets of laughter. Another important finding is that different types of laughter are accompanied by different gaze patterns both by the person laughing (the laugher) and their partner. In the case of pleasant incongruity laughter the laugher is more likely to look at their partner before the onset of laughter, whereas after the onset, social incongruity involves gazing at the partner. These findings suggest that future ECAS should not address all laughs in a uniform fashion – they should be sensitive to precise coordination between laughter and gaze that are critically dependent on the types of laughables. It also raises the issue whether one should analyse such behaviours as compositional structures broken down into constituents like laughter, gaze and speech or whether they can be represented as behavioural gestalts (Holler and Levinson, 2019), or 'composite utterances' (Enfield, 2009).

17.1.2 Laughter interpretation

First, we studied how laughter affects the interpretation of an utterance by considering the level of dialogue acts. This is a direct concern of the NLU module of a dialogue system, responsible for the representation of user inputs and intents. Analysing dialogue acts in collocation with laughs again showed that laughter is tightly connected to dialogue information structure and the dialogue acts performed. Moreover, we looked at the computational task of dialogue act recognition (DAR) and demonstrated that laugher can be a valuable cue for it. It was fascinating to see that in DAR, laughter helps disambiguate meaning, including literal and non-literal interpretation of utterances, demonstrated by differentiating wh-

and rhetorical questions. It has been a long standing challenge for natural language processing (NLP) applications to understand non-literal (i.e. figurative) meanings and we deem it useful to make use of laughter for this purpose. We also brought forward the fact that laughter qualifies as a dialogue act in its own right and future dialogue annotation schemes should account for it.

On the other hand, there is long-standing controversy in linguistics as to whether speech acts or dialogue acts are necessarily involved in processing in interaction. Both Relevance Theory and Conversational Analysis (CA) have disputed the validity of speech acts as necessary ingredients of explanatory accounts (Searle, 1992; Schegloff, 1992; Louch, 1966; Nicolle, 2000; Sperber and Wilson, 1995). It might be the case that with fine-grained analyses of all the components that constitute 'multimodal utterances', there is no need for dialogue act classification as as action interpretation can be directly achieved in terms of (perlocutionary) effects.

Next, we laid the groundwork for the central component of a spoken dialogue system, when we presented our own dialogue manager (DM) – the Linear Dialogue Manager (LDM), developed in accord with the KoS framework (Ginzburg, 2012). Our LDM works towards a domain-general treatment of dialogical phenomena using a concise implementation with a manageable set of rules. It adds rudimentary support for grounding and enthymematic reasoning which are crucial for the treatment of laughter and humour. Moreover, it incorporates processing facts like resource-sensitivity within the architecture of the system itself rather than having to explicitly encode them as separate principles.

Finally, we put our LDM to use. Grounding our study on the practical examples of laughter usage in task-oriented spoken dialogue, we drew connections between laughter and levels of action in dialogue. We then offered proposals as to how certain types of laughter can be processed in DM and NLG. We looked at laughter as a negative answer to polar questions and interrogative feedback, and provided a formal realisation of Interactive Communication Management (ICM) moves that may include laughter. This contribution sets an example for the treatment of laughter within DM and

demonstrates the potential for extension to other types of laughter (e.g. positive feedback related to pleasant incongruities).

17.1.3 Laughter and humour

In the last part we set out to bridge the gap between laughter and humour. We emphasised the essential requirement to analyse humour as an interactive phenomenon and how this connects to laughter.

We proposed a theory which we believe is the first formal interactive approach to humour. We explicated our theory on the basis of foundations that are external to existing theories of humour and used in discussions of other linguistic and interactive phenomena. Moreover, we formally explained the process of integration of new information within joke processing in our case study. We believe that our findings can provide insights for future dialogue systems. The principal theoretical implication of this part is that humour is based on dialogical inference mechanisms. Humour is an immensely complex phenomenon and we are fully aware of the paucity of our investigations. Nevertheless, we did not concentrate on humour in general, e.g. differentiating humour from non-humour, rather we were concerned with the potential of our investigations to be extendable to laughter and laughables.

Not unexpectedly, we raised more questions than we had in the beginning and hope to address them in natural progressions of this work.

17.2 FUTURE DIRECTIONS

The work in this thesis can be extended in multiple ways which we have outlined in the discussion sections of corresponding chapters. They contain information about limitations, necessary additional experiments etc. By contrast, here we would like to discuss further *directions* for future work that we consider the most intriguing and potentially fruitful.

Probabilistic reasoning under uncertainty

This thesis was limited by the absence of a probabilistic aspect in the reasoning components discussed in the thesis. This is mainly a con-

cern for the dialogue management components of dialogue systems. This is particularly important due to the fact that dialogue systems typically deal with multiple ASR and NLU hypotheses accompanied by confidence scores, and the DM has to take into account their contextualised ranking. There is the potential to extend our DM framework presented in Chapter 10 by assigning probabilities to rules and axioms, following the approach by Lison (2015) whilst maintaining a rich information state.

Another potentially fruitful direction in this regard is reasoning with limited resources thus emulating human cognitive constraints. A dialogue system which is capable of calculating incongruity in real-time from a range of rhetorical resources has to have some mechanism of salience calculation (see Chapter 16). Some proposals for reasoning under uncertainty exist in general-purpose artificial intelligence systems, e.g. Wang et al. (2018) and Johansson (2019). Further modelling work will have to be conducted in order to understand how laughter (and humour) can arise from numerous knowledge resources given the time and contextual constraints.

Emotion modelling

A natural progression of this thesis should allow the calculation of the salience of implicit assumptions under uncertainty and time constraints, and further appraisal of them in relation to the emotional states of the agent. To model emotional states of the agent, the dialogue system should be equipped with a computational theory of emotion. In principle, the model of emotions should be tightly connected with a dialogue state and it would be feasible to represent it within the same LDM framework. We can use some insights from new developments of KoS (Ginzburg et al., 2020), where an emotional dimension is added to the Dialogue Gameboard. Such a dimension can reflect appraisals and constructions of emotions, and has been used by Ginzburg et al. (2020) to sketch the effect of laughter, along with other non-verbal social signals, on the demonstrated level of pleasure, and deduce or disambiguate laughter functions.

This require taking a closer look at cognitive theories of emo-

tions, frontiers in affective computing and rigourous evaluation in an ECA. Our next point is devoted to the latter.

Fully fledged evaluation in ECA

Greater efforts are needed to evaluate our findings. We believe that this thesis created an opportunity to equip a dialogue system with the capacity for a deeper understanding of laughter. We covered aspects that can be embedded in the components of a dialogue system on the level of context and meaning (and additionally our laughter prediction study is beneficial for language modelling in ASR). Continued efforts are needed to create a full-scale spoken dialogue system, preferably an ECA, where laughter behaviours can be evaluated in interactions with real users. There have been a significant amount of research (see Section 5.4) implementing laughter in ASR, text-to-speech synthesis (TTS) and behaviour realisation components of an ECA.

Precise account for laughter placement

It is unfortunate that this thesis did not provide a formal account concerning the precise placement of laughter in relation to the laughable. We had started looking into this issue in Eshghi et al. (2019) and Maraev, Eshghi et al. (2019), but the results were preliminary and therefore not included in the current thesis. In these papers we build on recent work on Dynamic Syntax and Type Theory with Records (DS-TTR) (Purver et al., 2011; Gregoromichelaki et al., 2020; Howes and Eshghi, 2021; Eshghi et al., 2015), to describe how laughter is processed and integrated incrementally as an utterance unfolds. This was done by treating laughter as anaphoric, with the analysis following the same pattern as pronouns. We also briefly discuss an account of how other-laughter can act as positive evidence of understanding.

We have also not provided an account of how laughter is distributed in conversation. This requires an empirical analysis of the syntactic points at which laughter tends to occur. We hypothesise that just as with patterns of repair, this will vary across languages (Rieger, 2003) because of the specific features of the language (e.g.

English allows self-repairs which repeat the determiner before a noun, but this strategy is not available for languages without determiners, such as Farsi). For example, there will be different patterns of laughter placement in different languages, constrained by the unfolding structure of the linguistic input.

A further study could assess how the grammars of different languages provide different opportunities for laughter using data with precise laughter annotation collected in the DUEL (French, Chinese and German, Hough et al., 2016) and NOMCO (Nordic languages, Navarretta et al., 2012) projects.

Evaluation of laughter placement

In this thesis we proposed a prediction model for laughter relevance and a statistical account for laughter placement in a multimodal setting. Both of the accounts ought to be evaluated in a realistic environment. There are two ways to do this: i) in a spoken dialogue system or an ECA, or ii) using real-time intervention techniques.

The first way is more common. For instance, we can build a dialogue system which inserts laughs in NLG informed by the model of laughter prediction (Chapter 7) at places where laughter is more likely to occur, based on a certain threshold. Then a dialogue system will be evaluated extrinsically in comparison with either a system that does not laugh or a system that inserts laughs at random places. An alternative account can also be driven by certain basic heuristics (e.g., insert laughs at random but only at the end of turn).

Similarly, laughter placement can be assessed with an ECA which produces certain gaze behaviours (Chapter 8) according to laughs that are either generated by the system (in this case the ECA is the laugher) or when originating from the user (in this case ECA is the partner). Such systems can be evaluated using the established subjective evaluation metrics (e.g. Subjective Assessment of Speech System Interfaces (SASSI), Hone and Graham, 2000) or follow previous studies evaluating laughing ECAS, e.g. Mancini et al. (2017).

In addition to dialogue systems, placement can be assessed using real-time interventions. In such a setup participants communicate

with each other using the Dialogue Experimental Toolkit (Healey et al., 2003)[1], where their messages pass through the server which can intercept and modify them experimentally. In previous studies we have used such techniques to assess laughter mimicry and the interrelation between laughter and emotional contagion (Mills et al., 2021). In Maraev, Mazzocconi, Mills et al. (2020) we inserted spoof turns to study reactions to clarification requests related to laughs, e.g. 'lol what?' or 'haha?'. In regard to laughter placement, additional laughs can be inserted at laughter relevance spaces in accord with a prediction model, either a neural network (Chapter 7) or the aforementioned formal model. These interventions can then be assessed according to how disruptive they are for the flow of conversation, analogously to the Poppe et al. (2011) study of appropriateness of backchannels.

A natural extension to this would be to go beyond text-based interventions and extend the intervention techniques to immersive virtual reality (VR) environments. In such environments participants can be embodied as avatars. Their gestures and gaze directions can be tracked and manipulated in real time. This approach would allow checking our findings regarding the associations of gaze patterns with different laughable types (Chapter 8) without needing to develop an ECA.

In sum, this thesis has paved the way for future research on laughter in dialogue systems. As ever, much remains to be done.

[1]https://dialoguetoolkit.github.io/chattool/

Bibliography

Abel, Andreas and Jean-Philippe Bernardy (2020). 'A unified view of modalities in type systems'. In: *Proceedings of the ACM on Programming Languages* 4.ICFP.

Allwood, Jens (1995). *An activity based approach to pragmatics*. Tech. rep. (GPTL) 75. Gothenburg papers in theoretical linguistics, University of Gothenburg.

AMI Project (2005). *Guidelines for Dialogue Act and Addressee Annotation. Augmented Multiparty Interaction Project Document, v. 1.0*.

Andersson, Richard and Olof Sandgren (2016). 'ELAN Analysis Companion (EAC): A Software Tool for Time-course Analysis of ELAN-annotated Data'. In: *Journal of Eye Movement Research* 9.3.

Anscombre, Jean-Claude (1995). 'La Théorie des topoi: Sémantique ou rhétorique?' In: *Hermés* 15, pp. 185–198.

Arkes, Hal R and Peter Ayton (1999). 'The sunk cost and Concorde effects: Are humans less rational than lower animals?' In: *Psychological bulletin* 125.5, p. 591.

Atkey, Robert (2018). 'Syntax and Semantics of Quantitative Type Theory'. In: *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK*, pp. 56–65.

Attardo, Salvatore (2010). *Humorous texts: A semantic and pragmatic analysis*. Vol. 6. Berlin: Walter de Gruyter.

Attardo, Salvatore and Victor Raskin (1991). 'Script theory revis (it) ed: Joke similarity and joke representation model'. In: *Humor-International Journal of Humor Research* 4.3-4, pp. 293–348.

Aucouturier, Jean-Julien, Petter Johansson, Lars Hall, Rodrigo Segnini, Lolita Mercadié and Katsumi Watanabe (2016). 'Covert

digital manipulation of vocal emotion alter speakers’ emotional states in a congruent direction'. In: *Proceedings of the National Academy of Sciences* 113.4, pp. 948–953.
Austin, John Langshaw (1975). *How to do things with words*. Vol. 88. Oxford: Oxford university press.
Bååth, Rasmus (2010). 'ChildFreq: An online tool to explore word frequencies in child language'. In: *Lucs Minor* 16, pp. 1–6.
Bailly, Gérard, Stephan Raidt and Frédéric Elisei (2010). 'Gaze, conversational agents and face-to-face communication'. In: *Speech Communication* 52, pp. 598–612.
Barwise, Jon and John Perry (1981). 'Situations and attitudes'. In: *The Journal of Philosophy* 78.11, pp. 668–691.
Bates, Douglas, Reinhold Kliegl, Shravan Vasishth and Harald Baayen (2015). 'Parsimonious mixed models'. In: *arXiv preprint arXiv: 1506.04967*.
Batliner, Anton, Stefan Steidl, Florian Eyben and Björn Schuller (2019). 'On laughter and speech-laugh, based on observations of child-robot interaction'. In: *arXiv preprint arXiv:1908.11593*.
Bavelas, Janet Beavin and Nicole Chovil (2000). 'Visible acts of meaning: An integrated message model of language in face-to-face dialogue'. In: *Journal of Language and social Psychology* 19.2, pp. 163–194.
Bavelas, Janet Beavin, Linda Coates and Trudy Johnson (2002). 'Listener responses as a collaborative process: The role of gaze'. In: *Journal of Communication* 52.3, pp. 566–580.
Bavelas, Janet Beavin, Jennifer Gerwing, Chantelle Sutton and Danielle Prevost (2008). 'Gesturing on the telephone: Independent effects of dialogue and visibility'. In: *Journal of Memory and Language* 58.2, pp. 495–520.
Beattie, Geoffrey W (1978). 'Sequential temporal patterns of speech and gaze in dialogue'. In: *Semiotica* 23.1-2, pp. 29–52.
— (1981). 'A further investigation of the cognitive interference hypothesis of gaze patterns during conversation'. In: *British Journal of Social Psychology* 20.4, pp. 243–248.
Beaver, David and Henk Zeevat (2007). 'Accommodation'. In: *The Oxford handbook of linguistic interfaces*, pp. 503–536.

Becker-Asano, Christian and Hiroshi Ishiguro (2009). 'Laughter in social robotics-no laughing matter'. In: *Intl. Workshop on Social Intelligence Design*, pp. 287–300.

Bertero, Dario and Pascale Fung (2016). 'A long short-term memory framework for predicting humor in dialogues'. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 130–135.

Billig, Michael (2005). *Laughter and ridicule: Towards a social critique of humour*. London: SAGE.

Binsted, Kim et al. (1995). 'Using humour to make natural language interfaces more friendly'. In: *Proceedings of the AI Alife and Entertainment workshop, Fourteenth International Joint Conference on Artificial Intelligence*. Montreal, Quebec, Canada: IJCAI Organization.

Bird, Steven and Edward Loper (2004). 'NLTK: the natural language toolkit'. In: *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*. Association for Computational Linguistics, p. 31.

Blache, Philippe, Massina Abderrahmane, Stéphane Rauzy, Magalie Ochs and Houda Oufaida (2020). 'Two-level classification for dialogue act recognition in task-oriented dialogues'. In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, pp. 4915–4925.

Bondarenko, Anastasia, Christine Howes and Staffan Larsson (2020). *Directory Enquiries Corpus*.

Bonin, Francesca, Nick Campbell and Carl Vogel (2012). 'Laughter and topic changes: Temporal distribution and information flow'. In: *2012 IEEE 3rd international conference on cognitive infocommunications (CogInfoCom)*. IEEE, pp. 53–58.

Bova, Antonio and Francesco Arcidiacono (2013). 'Investigating children's Why-questions: A study comparing argumentative and explanatory function'. In: *Discourse Studies* 15.6, pp. 713–734.

Bratko, Ivan (1986). *Prolog Programming for Artificial Intelligence*. Reading, Massachusetts: Addison-Wesley.

Breitholtz, Ellen (2014). 'Reasoning with topoi - towards a rhetorical approach to non-monotonicity'. In: *Proceedings of the 50th anniversary convention of the AISB, 1st–4th April*. AISB, pp. 190–198.

— (2020). *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Leiden, The Netherlands: Brill.

Breitholtz, Ellen and Robin Cooper (2011). 'Enthymemes as Rhetorical Resources'. In: *Proceedings of*. Vol. SemDial 2011 (LosAngelogue) Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pp. 149–157.

Breitholtz, Ellen and Christine Howes (2020). 'Communicable Reasons: How Children Learn Topoi Through Dialogue'. In: *Proceedings of the 24th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Virtually at Brandeis, Waltham, New Jersey: SEMDIAL.

Breitholtz, Ellen and Vladislav Maraev (2019). 'How to Put an Elephant in the Title: Modeling Humorous Incongruity with Topoi'. In: *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. London, United Kingdom: SEMDIAL.

Brône, Geert (2020). 'Multimodal (di)stance in interaction. Eye gaze, irony and joint pretence'. In: *Multimodal Communication Seminar Series-Oxford University*.

Brown, Penelope and Steve Levinson (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Brugman, Hennie and Albert Russel (2004). 'Annotating Multimedia/Multi-modal Resources with ELAN'. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*. European Language Resources Association.

Bryant, Gregory A (2016). 'How do laughter and language interact?' In: *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).*

Bryson, Joanna J (2018). 'Patiency is not a virtue: the design of intelligent systems and systems of ethics'. In: *Ethics and Information Technology* 20.1, pp. 15–26.

Bunt, Harry (1981). 'Conversational Principles in Question-Answer Dialogues'. In: *Zur Theorie der Frage. Vorträge des Bad Homburger Kolloquiums, 13.-15. November 1978*. Ed. by Dieter Krallmann and Gerhard Stickel. Gunter Narr Verlag, pp. 119–141.

Burnett, Heather (2019). 'Signalling Games, Sociolinguistic Variation and the Construction of Style'. In: *Linguistics and Philosophy* 42.5, pp. 419–450.

Chapman, Antony J (1983). 'Humor and laughter in social interaction and some implications for humor research'. In: *Handbook of humor research*. Springer, pp. 135–157.

Chen, Lei and Chong Min Lee (2017). 'Predicting Audience's Laughter During Presentations Using Convolutional Neural Network'. In: *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*. Copenhagen, Denmark: Association for Computational Linguistics, pp. 86–90.

Chen, Peng-Yu and Von-Wun Soo (2018). 'Humor Recognition Using Deep Learning'. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, pp. 113–117.

Chomsky, Noam (1965). *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chowdhery, Aakanksha, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann et al. (2022). 'PaLM: Scaling Language Modeling with Pathways'. In: *arXiv preprint arXiv:2204.02311*.

Clark, Herbert H (1996). *Using language*. Cambridge: Cambridge university press.

Clark, Kevin and Christopher D. Manning (2016). 'Improving Coreference Resolution by Learning Entity-Level Distributed Representations'. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics, pp. 643–653.

Cooper, Robin (2005). 'Austinian truth, attitudes and type theory'. In: *Research on Language and Computation* 3.4, pp. 333–362.

— (2011). 'Copredication, Quantification and Frames'. In: *Logical Aspects of Computational Linguistics - 6th International Conference, LACL 2011, Montpellier, France, June 29 - July 1, 2011. Proceedings*. Ed. by Sylvain Pogodalla and Jean-Philippe Prost. Vol. 6736. Lecture Notes in Computer Science. Springer, pp. 64–79.

— (2013). 'Clarification and generalized quantifiers'. In: *Dialogue and Discourse* 4 (1), pp. 1–25.

— (2022). *From Perception to Communication: a Theory of Types for Action and Meaning*. In press. Oxford: Oxford University Press.

Cooper, Robin and Jonathan Ginzburg (2012). 'Negative inquisitiveness and alternatives-based negation'. In: *Logic, Language and Meaning*. Springer, pp. 32–41.

— (2015). 'Type theory with records for natural language semantics'. In: *Handbook of Contemporary Semantic Theory*, pp. 375–407.

Core, Mark G and James F Allen (1997). 'Coding Dialogs with the DAMSL Annotation Scheme'. en. In: *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*. Boston, MA, pp. 28–35.

Dale, Rick, Riccardo Fusaroli, Nicholas D Duran and Daniel C Richardson (2013). 'The self-organization of human interaction'. In: *Psychology of learning and motivation*. Vol. 59. Elsevier, pp. 43–95.

Danaher, John (2020). 'Welcoming robots into the moral circle: a defence of ethical behaviourism'. In: *Science and Engineering Ethics* 26.4, pp. 2023–2049.

Davies, Catherine (1984). 'Joint joking: Improvisational humorous episodes in conversation'. In: *Annual Meeting of the Berkeley Linguistics Society*. Vol. 10, pp. 360–371.

Davies, Christie (2011). *Jokes and targets*. Bloomington: Indiana University Press.

De Ruiter, JP (2005). 'If eye-gaze frequency drops, its relationship with turn-taking disappears'. In: *Poster presented at AMLAP*.

Demuth, Katherine, Jennifer Culbertson and Jennifer Alter (2006). 'Word-minimality, epenthesis and coda licensing in the early acquisition of English'. In: *Language and Speech* 49.2, pp. 137–173.

Devereux, Paul G and Gerald P Ginsburg (2001). 'Sociality effects on the production of laughter'. In: *The Journal of General Psychology* 128.2, pp. 227–240.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2019). 'BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding'. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186.

Ding, Yu, Ken Prepin, Jing Huang, Catherine Pelachaud and Thierry Artières (2014). 'Laughter animation synthesis'. In: *Proc. AAMS 2014*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 773–780.

Dixon, Lucas, Alan Smaill and Tracy Tsang (2009). 'Plans, actions and dialogues using linear logic'. In: *Journal of Logic, Language and Information* 18.2, pp. 251–289.

Dobnik, Simon and Robin Cooper (2017). 'Interfacing language, spatial perception and cognition in Type Theory with Records'. In: *Journal of Language Modelling* 5.2, pp. 273–301.

Doherty-Sneddon, Gwyneth and Fiona G Phelps (2005). 'Gaze aversion: A response to cognitive or social difficulty?' In: *Memory & Cognition* 33.4, pp. 727–733.

Ducrot, Oswald (1980). *Les échelles argumentatives*. Minuit, Paris.

Ducrot, Oswald (1988). 'Topoï et Formes Topique'. In: *Bulletin d'études de la linguistique française* 22, pp. 1–14.
Duncan, Starkey (1972). 'Some signals and rules for taking speaking turns in conversations.' In: *Journal of personality and social psychology* 23.2, p. 283.
Duncan, Starkey and Donald W Fiske (1979). 'Dynamic patterning in conversation: Language, paralinguistic sounds, intonation, facial expressions, and gestures combine to form the detailed structure and strategy of face-to-face interactions'. In: *American Scientist* 67.1, pp. 90–98.
Dybala, Pawel, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki (2008). 'Humor prevails! - Implementing a joke generator into a conversational system'. In: *Australasian Joint Conference on Artificial Intelligence*. Springer, pp. 214–225.
Dybala, Pawel, Michal Ptaszynski, Rafal Rzepka and Kenji Araki (2009). 'Activating Humans with Humor – A Dialogue System That Users Want to Interact with'. In: *IEICE Transactions on Information and Systems* E92.D.12, pp. 2394–2401.
— (2010). 'Extending the chain: humor and emotions in human computer interaction'. In: *International Journal of Computational Linguistics Research* 1.3, pp. 116–125.
Eckert, Penelope (2008). 'Variation and the indexical field'. In: *Journal of sociolinguistics* 12.4, pp. 453–476.
Ekstedt, Erik and Gabriel Skantze (2020). 'TurnGPT: a Transformer-based Language Model for Predicting Turn-taking in Spoken Dialog'. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, pp. 2981–2990.
El Haddad, Kevin, Hüseyin Çakmak, Stéphane Dupont and Thierry Dutoit (2017). 'Amused speech components analysis and classification: Towards an amusement arousal level assessment system'. In: *Computers & Electrical Engineering* 62, pp. 588–600.
Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Language Culture and Cognition. Cambridge: Cambridge University Press.

Eshghi, Arash, Christine Howes, Eleni Gregoromichelaki, Julian Hough and Matthew Purver (2015). 'Feedback in Conversation as Incremental Semantic Update'. In: *Proceedings of the 11th International Conference on Computational Semantics, IWCS 2015, 15-17 April, 2015, Queen Mary University of London, London, UK*. Ed. by Matthew Purver, Mehrnoosh Sadrzadeh and Matthew Stone. The Association for Computer Linguistics, pp. 261–271.

Eshghi, Arash, Vladislav Maraev, Christine Howes, Julian Hough and Chiara Mazzocconi (2019). '"What are you laughing at?" Incremental Processing of Laughter in Interaction'. In: *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*. London, United Kingdom: SEMDIAL.

Falcon, Andrea (2019). 'Aristotle on Causality'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2019. Metaphysics Research Lab, Stanford University.

Fernández, Raquel (2022). 'Dialogue'. In: *The Oxford Handbook of Computational Linguistics 2nd edition*. Oxford: Oxford University Press.

Fernández, Raquel and Jonathan Ginzburg (2002). 'Non-sentential utterances: A corpus study'. In: *Tratement Automatique de Langues* 43.2, pp. 13–42.

Fernández, Raquel, Jonathan Ginzburg and Shalom Lappin (2007). 'Classifying Ellipsis in Dialogue: A Machine Learning Approach'. In: *Computational Linguistics* 33.3, pp. 397–427.

Fresno, Andrea Carrión del, Staffan Larsson and Vladislav Maraev (2022). 'Dialogue strategies for... cómo se dice entrenamiento de vocabulario?' In: *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Dublin, Ireland: SEMDIAL.

Freud, Sigmund (1905). *Jokes and their relarion fo the unconscious (J. Strachey, Trans., 1960)*. London: Hogarth.

Fridlund, Alan J (2014). *Human facial expression: An evolutionary view*. San Diego: Academic press.

Fusaroli, Riccardo and Kristian Tylén (2012). 'Carving language for social coordination: A dynamical approach'. In: *Interaction studies* 13.1, pp. 103–124.

Gibson, James J (1979). *The theory of affordances. The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gillick, Jon, Wesley Deng, Kimiko Ryokai and David Bamman (2021). 'Robust Laughter Detection in Noisy Environments'. In: *Proc. Interspeech 2021*, pp. 2481–2485.

Gilmartin, Emer, Francesca Bonin, Carl Vogel and Nick Campbell (2013). 'Laugher and Topic Transition in Multiparty Conversation'. In: *Proceedings of the SIGDIAL 2013 Conference*. Metz, France: Association for Computational Linguistics, pp. 304–308.

Ginzburg, Jonathan (2012). *The interactive stance*. Oxford: Oxford University Press.

Ginzburg, Jonathan, Ellen Breitholtz, Robin Cooper, Julian Hough and Ye Tian (2015). 'Understanding Laughter'. In: *Proceedings of the 20th Amsterdam Colloquium*, pp. 137–146.

Ginzburg, Jonathan, Raquel Fernández and David Schlangen (2014). 'Disfluencies as intra-utterance dialogue moves'. In: *Semantics and Pragmatics* 7.9, pp. 1–64.

Ginzburg, Jonathan, Chiara Mazzocconi and Ye Tian (2020). 'Laughter as language'. In: *Glossa: a journal of general linguistics* 5.1, pp. 1–51.

Girard, J.-Y. (1995). 'Linear Logic: its syntax and semantics'. In: *Advances in Linear Logic*. London Mathematical Society Lecture Note Series. Cambridge University Press, pp. 1–42.

Gironzetti, Elisa (2017). 'Multimodal and Eye-tracking Evidence in the Negotiation of Pragmatic Intentions in Dyadic Conversations: The Case of Humorous Discourse'. PhD thesis. Texas A&M University–Commerce.

Glas, Dylan F., Takashi Minato, Carlos Toshinori Ishi, Tatsuya Kawahara and Hiroshi Ishiguro (2016). 'ERICA: The ERATO Intelligent Conversational Android'. In: *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016, New York, NY, USA, August 26-31, 2016*. IEEE, pp. 22–29.

Glenberg, Arthur M, Jennifer L Schroeder and David A Robertson (1998). 'Averting the gaze disengages the environment and facilitates remembering'. In: *Memory & Cognition* 26.4, pp. 651–658.

Glenn, Phillip (2003). *Laughter in interaction*. Cambridge: Cambridge University Press.

Glenn, Phillip and Elizabeth Holt (2013). *Studies of laughter in interaction*. London: Bloomsbury Academic.

Goffman, Erving (1955). 'On face-work: An analysis of ritual elements in social interaction'. In: *Psychiatry* 18.3, pp. 213–231.

Goodwin, Charles (1980). 'Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning'. In: *Sociological inquiry* 50.3-4, pp. 272–302.

Goodwin, Marjorie Harness and Charles Goodwin (1986). 'Gesture and coparticipation in the activity of searching for a word'. In: *Semiotica* 62.1-2, pp. 51–76.

Gopalan, Pushkas and Simon T Browning (2002). 'Intractable paroxysmal sneezing'. In: *The Journal of Laryngology & Otology* 116.11, pp. 958–959.

Gray, Colin M, Yubo Kou, Bryan Battles, Joseph Hoggatt and Austin L Toombs (2018). 'The dark (patterns) side of UX design'. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–14.

Gregoromichelaki, Eleni, Ruth Kempson and Christine Howes (2020). 'Actionism in syntax and semantics.' In: pp. 12–27.

Gregoromichelaki, Eleni, Ruth Kempson, Matthew Purver, Gregory J Mills, Ronnie Cann, Wilfried Meyer-Viol and Patrick GT Healey (2011). 'Incrementality and intention-recognition in utterance processing'. In: *Dialogue & Discourse* 2.1, pp. 199–233.

Grice, H. Paul (1975). 'Logic and Conversation'. In: *The Logic of Grammar*. Ed. by Donald Davidson and Gilbert Harman. Encino, California: Dickenson Publishing Co., pp. 64–75.

Günther, Ulrike (2003). 'What's in a laugh?: Humour, jokes and laughter in the conversational corpus of the BNC'. PhD thesis. University of Freiburg.

Haddad, Kevin El, Hüseyin Çakmak, Emer Gilmartin, Stéphane Dupont and Thierry Dutoit (2016). 'Towards a listening agent: a system generating audiovisual laughs and smiles to show interest'. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, pp. 248–255.

Harris, Christine R (1999). 'The mystery of ticklish laughter'. In: *American Scientist* 87.4, p. 344.

Hay, Jennifer (2000). 'Functions of humor in the conversations of men and women'. In: *Journal of pragmatics* 32.6, pp. 709–742.

Healey, P. G. T., Matthew Purver, James King, Jonathan Ginzburg and Greg Mills (2003). 'Experimenting with Clarification in Dialogue'. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Boston, Massachusetts.

Heldner, Mattias, Anna Hjalmarsson and Jens Edlund (2013). 'Backchannel relevance spaces'. In: *Nordic Prosody XI, Tartu, Estonia, 15-17 August, 2012*. Peter Lang Publishing Group, pp. 137–146.

Hempelmann, Christian F (2008). 'Computational humor: Beyond the pun?' In: *The Primer of Humor Research. Humor Research* 8, pp. 333–360.

Hempelmann, Christian F and Salvatore Attardo (2011). 'Resolutions and their incongruities: Further thoughts on logical mechanisms'. In: *Humor-International Journal of Humor Research* 24.2, pp. 125–149.

Hladky, Mirella, Tanja Schneeberger and Patrick Gebhard (2021). 'Understanding Shame Signals: Functions of Smile and Laughter in the Context of Shame'. In: *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 01–07.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). 'Long short-term memory'. In: *Neural computation* 9.8, pp. 1735–1780.

Holler, Judith and Stephen C. Levinson (2019). 'Multimodal Language Processing in Human Communication'. In: *Trends in Cognitive Sciences* 23.8, pp. 639–652.

Holt, Elizabeth (2010). 'The last laugh: Shared laughter and topic termination'. In: *Journal of Pragmatics* 42.6, pp. 1513–1525.

Hone, Kate S. and Robert Graham (2000). 'Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI)'. In: *Nat. Lang. Eng.* 6.3&4, pp. 287–303.

Hough, Julian and Matthew Purver (2012). 'Processing Self-Repairs in an Incremental Type-Theoretic Dialogue System'. In: *Proceedings of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial)*. Paris, France, pp. 136–144.

Hough, Julian, Ye Tian, Laura de Ruiter, Simon Betz, Spyros Kousidis, David Schlangen and Jonathan Ginzburg (2016). 'DUEL: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter'. In: *10th edition of the Language Resources and Evaluation Conference*.

Howes, Christine, Anastasia Bondarenko and Staffan Larsson (2019). 'Good call! Grounding in a Directory Enquiries Corpus'. In: *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. London, United Kingdom: SEMDIAL.

Howes, Christine and Arash Eshghi (2021). 'Feedback Relevance Spaces: Interactional Constraints on Processing Contexts in Dynamic Syntax'. In: *Journal of Logic, Language and Information* 30.2, pp. 331–362.

Howes, Christine, Matthew Purver and Rose McCabe (2014). 'Linguistic indicators of severity and progress in online text-based therapy for depression'. In: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 7–16.

Hu, Jiaxiong, Yun Huang, Xiaozhu Hu and Yingqing Xu (2021). 'Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues'. In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–7.

Huang, Minlie, Xiaoyan Zhu and Jianfeng Gao (2020). 'Challenges in building intelligent open-domain dialog systems'. In: *ACM Transactions on Information Systems (TOIS)* 38.3, pp. 1–32.

Hurley, Matthew M, Daniel Clement Dennett, Reginald B Adams Jr and Reginald B Adams (2011). *Inside jokes: Using humor to reverse-engineer the mind*. MIT press.

Hutto, Clayton J. and Eric Gilbert (2014). 'VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text'. In: *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014*. Ed. by Eytan Adar, Paul Resnick, Munmun De Choudhury, Bernie Hogan and Alice Oh. The AAAI Press.

Jefferson, Gail (1979). 'A technique for inviting laughter and its subsequent acceptance/declination'. In: *Everyday language: Studies in ethnomethodology*, pp. 79–96.

— (1984). 'On the organization of laughter in talk about troubles'. In: *Structures of Social Action: Studies in Conversation Analysis*, pp. 346–369.

— (1989). 'Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation'. In: *Conversation: An Interdisciplinary Perspective*. Ed. by Derek Roger and Peter Bull. Clevedon, England. Chap. 8, pp. 166–196.

Jefferson, Gail, Harvey Sacks and Emanuel A. Schegloff (1987). 'Notes on laughter in the pursuit of intimacy'. In: *Talk and Social Organisation*. Ed. by Graham Button and John R. E. Lee. Clevedon: Multilingual Matters, pp. 152–205.

Johansson, Robert (2019). 'Arbitrarily Applicable Relational Responding'. In: *Artificial General Intelligence*. Ed. by Patrick Hammer, Pulin Agrawal, Ben Goertzel and Matthew Iklé. Cham: Springer International Publishing, pp. 101–110.

Jokinen, Kristiina (2009). *Constructive dialogue modelling: Speech interaction and rational agents*. Vol. 10. USA: Wiley-Interscience.

Jurafsky, Daniel, Liz Shriberg and Debra Biasca (1997a). 'Switchboard Dialog Act Corpus'. In: *International Computer Science Inst. Berkeley CA, Tech. Rep.*

— (1997b). *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual*.

Kantharaju, Reshmashree Bangalore, Caroline Langlet, Mukesh Barange, Chloé Clavel and Catherine Pelachaud (2020). 'Multimodal Analysis of Cohesion in Multi-party Interactions'. In: *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*. European Language Resources Association, pp. 498–507.

Katevas, Kleomenis, Patrick GT Healey and Matthew Tobias Harris (2014). 'Robot stand-up: engineering a comic performance'. In: *Proceedings of the Workshop on Humanoid Robots and Creativity at the IEEE-RAS International Conference on Humanoid Robots Humanoids (Madrid)*.

Kaushik, Lakshmish, Abhijeet Sangwan and John H. L. Hansen (2015). 'Laughter and filler detection in naturalistic audio'. In: *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*. ISCA, pp. 2509–2513.

Kempson, Ruth, Wilfried Meyer-Viol and Dov Gabbay (2001). *Dynamic Syntax: The Flow of Language Understanding*. New Jersey, US: Wiley-Blackwell.

Kendon, Adam (1967). 'Some functions of gaze-direction in social interaction'. In: *Acta Psychologica* 26, pp. 22–63.

Kendrick, Kobin H and Judith Holler (2017). 'Gaze direction signals response preference in conversation'. In: *Research on Language and Social Interaction* 50.1, pp. 12–32.

Kim, Yoon (2014). 'Convolutional Neural Networks for Sentence Classification'. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. Ed. by Alessandro Moschitti, Bo Pang and Walter Daelemans. ACL, pp. 1746–1751.

Kozareva, Zornitsa and Sujith Ravi (2019). 'ProSeqo: Projection Sequence Networks for On-Device Text Classification'. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong,

China: Association for Computational Linguistics, pp. 3894–3903.
Lala, Divesh, Koji Inoue and Tatsuya Kawahara (2020). 'Prediction of Shared Laughter for Human-Robot Dialogue'. In: *Companion Publication of the 2020 International Conference on Multimodal Interaction*. ICMI '20 Companion. Virtual Event, Netherlands: Association for Computing Machinery, pp. 62–66.
Larsson, Staffan (2002). 'Issue-based dialogue management'. PhD thesis. University of Gothenburg.
Lavia, Lisa, Harry J. Witchel, Francesco Aletta, Jochen Steffens, André Fiebig, Jian Kang, Christine Howes and Patrick G. T. Healey (2018). 'Non-participant observation methods for soundscape design and urban planning'. In: *Handbook of Research on Perception-Driven Approaches to Urban Assessment and Design*. Ed. by Francesco Aletta and Jieling Xiao. IGI Global.
Lee, Jina and Stacy Marsella (2006). 'Nonverbal Behavior Generator for Embodied Conversational Agents'. In: *Intelligent Virtual Agents, 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21-23, 2006, Proceedings*. Ed. by Jonathan Gratch, Robert Michael Young, Ruth Aylett, Daniel Ballin and Patrick Olivier. Vol. 4133. Lecture Notes in Computer Science. Springer, pp. 243–255.
Lemon, Oliver (2011). 'Learning what to say and how to say it: Joint optimisation of spoken dialogue management and natural language generation'. In: *Comput. Speech Lang.* 25.2, pp. 210–221.
Lepore, Ernie and Matthew Stone (2014). *Imagination and convention: Distinguishing grammar and inference in language*. Oxford: Oxford University Press.
Levinson, Stephen C. (1983). *Pragmatics*. Cambridge, U.K.: Cambridge University Press.
Lewis, David (1979). 'Scorekeeping in a Language Game'. In: *Journal of Philosophical Logic* 8, pp. 339–359.
Li, Xiujun, Yun-Nung Chen, Lihong Li, Jianfeng Gao and Asli Celikyilmaz (2017). 'End-to-End Task-Completion Neural Dialogue Systems'. In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei,*

Taiwan, November 27 - December 1, 2017 - Volume 1: Long Papers. Ed. by Greg Kondrak and Taro Watanabe. Asian Federation of Natural Language Processing, pp. 733–743.

Lison, Pierre (2015). 'A hybrid approach to dialogue management based on probabilistic rules'. In: *Computer Speech & Language* 34.1, pp. 232–255.

Lison, Pierre and Jorg Tiedemann (2016). 'OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles'. en. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation*, p. 7.

Liu, Bing and Ian R. Lane (2017). 'An End-to-End Trainable Neural Network Model with Belief Tracking for Task-Oriented Dialog'. In: *Interspeech 2017, 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, August 20-24, 2017*. Ed. by Francisco Lacerda. ISCA, pp. 2506–2510.

Louch, Alfred R (1966). *Explanation and human action*. Oakland, CA: University of California Press.

Lücking, Andy (2016). 'Modeling Co-Verbal Gesture Perception in Type Theory with Records'. In: *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, pp. 383–392.

Lücking, Andy, Sebastian Ptock and Kirsten Bergmann (2011). 'Assessing agreement on segmentations by means of Staccato, the Segmentation Agreement Calculator according to Thomann'. In: *International Gesture Workshop*. Springer, pp. 129–138.

Luong, Hieu-Thi and Junichi Yamagishi (2021). 'LaughNet: synthesizing laughter utterances from waveform silhouettes and a single laughter example'. In: *CoRR* abs/2110.04946.

Łupkowski, Paweł and Jonathan Ginzburg (2017). 'Query responses'. In: *Journal of Language Modelling* 4.2, pp. 245–292.

MacHale, D, AP Nilsen, P Derks, P Lewis, AA Berger, L Mintz, DLF Nilsen, C Gruner, E Oring, Willibald Ruch et al. (1997). 'Debate: Humor and political correctness'. In: *HUMOR: International Journal of Humor Research* 10.4, pp. 453–513.

MacWhinney, Brian (2000). *The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs, volume II: The database.*

Maguire, Eimear (2019). 'Enthymemetic Conditionals: Topoi as a guide for acceptability'. In: *Proceedings of the IWCS 2019 Workshop on Computing Semantics with Types, Frames and Related Structures*, pp. 65–74.

Mancini, Maurizio, Beatrice Biancardi, Florian Pecune, Giovanna Varni, Yu Ding, Catherine Pelachaud, Gualtiero Volpe and Antonio Camurri (2017). 'Implementing and evaluating a laughing virtual character'. In: *ACM Transactions on Internet Technology (TOIT)* 17.1, pp. 1–22.

Mansouri, Nadia and Zied Lachiri (2020). 'Laughter synthesis: A comparison between Variational autoencoder and Autoencoder'. In: *5th International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2020, Sousse, Tunisia, September 2-5, 2020*. IEEE, pp. 1–6.

Maraev, Vladislav, Jean-Philippe Bernardy and Jonathan Ginzburg (2020). 'Dialogue management with linear logic: the role of metavariables in questions and clarifications'. In: *Traitement Automatique des Langues (TAL)* 61.3, pp. 43–67.

Maraev, Vladislav, Jean-Philippe Bernardy and Christine Howes (2021). 'Non-humorous Use of Laughter in Spoken Dialogue Systems'. In: *Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2021)*, pp. 33–44.

Maraev, Vladislav, Ellen Breitholtz and Christine Howes (2020). 'How do you make an AI get the joke? Here's what I found on the web'. In: *First AISB Symposium on Conversational AI (SoCAI).*

Maraev, Vladislav, Ellen Breitholtz, Christine Howes and Jean-Philippe Bernardy (2021). 'Why Should I Turn Left? Towards Active Explainability for Spoken Dialogue Systems.' In: *Proceedings of the Reasoning and Interaction Conference (ReInAct 2021).* Gothenburg, Sweden: Association for Computational Linguistics, pp. 58–64.

Maraev, Vladislav, Ellen Breitholtz, Christine Howes, Staffan Larsson and Robin Cooper (2021). 'Something Old, Something

New, Something Borrowed, Something Taboo: Interaction and Creativity in Humour'. In: *Frontiers in Psychology* 12, p. 1443.
Maraev, Vladislav, Arash Eshghi, Julian Hough, Christine Howes and Chiara Mazzocconi (2019). 'Laughter relevance spaces'. In: *The 3rd Dynamic Syntax Conference*.
Maraev, Vladislav, Christine Howes and Jean-Philippe Bernardy (2019). 'Predicting Laughter Relevance Spaces in Dialogue'. In: *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction - 10th International Workshop on Spoken Dialogue Systems, IWSDS 2019, Syracuse, Sicily, Italy, 24-26 April 2019*. Ed. by Erik Marchi, Sabato Marco Siniscalchi, Sandro Cumani, Valerio Mario Salerno and Haizhou Li. Vol. 714. Lecture Notes in Electrical Engineering. Springer, pp. 41–51.
Maraev, Vladislav, Chiara Mazzocconi, Christine Howes and Jonathan Ginzburg (2018). 'Integrating laughter into spoken dialogue systems: preliminary analysis and suggested programme'. In: *Proceedings of the FAIM/ISCA Workshop on Artificial Intelligence for Multimodal Human Robot Interaction*, pp. 9–14.
Maraev, Vladislav, Chiara Mazzocconi, Gregory Mills and Christine Howes (2020). 'LOL what?: Empirical study of laughter in chat based dialogues'. In: *Proceedings of Laughter Workshop 2020*, pp. 55–58.
Maraev, Vladislav, Bill Noble, Chiara Mazzocconi and Christine Howes (2021). 'Dialogue act classification is a laughing matter'. In: *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Potsdam, Germany: SEMDIAL.
Marsella, Stacy C and Jonathan Gratch (2009). 'EMA: A process model of appraisal dynamics'. In: *Cognitive Systems Research* 10.1, pp. 70–90.
Martens, Chris (2015). 'Programming Interactive Worlds with Linear Logic'. PhD thesis. Carnegie Mellon University Pittsburgh, PA.
Mathur, Arunesh, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty and Arvind Narayanan (2019). 'Dark patterns at scale: Findings from a crawl of 11K shop-

ping websites'. In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW, pp. 1–32.

Matsumoto, Kento, Sunao Hara and Masanobu Abe (2020). 'Controlling the Strength of Emotions in Speech-Like Emotional Sound Generated by WaveNet'. In: *Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25-29 October 2020*. Ed. by Helen Meng, Bo Xu and Thomas Fang Zheng. ISCA, pp. 3421–3425.

Mazzocconi, Chiara (2019). 'Laughter in interaction: semantics, pragmatics and child development'. PhD thesis. Université de Paris.

Mazzocconi, Chiara and Jonathan Ginzburg (2022). 'A Longitudinal Characterization of Typical Laughter Development in Mother–Child Interaction from 12 to 36 Months: Formal Features and Reciprocal Responsiveness'. In: *Journal of Nonverbal Behavior*. In press.

Mazzocconi, Chiara, Vladislav Maraev, Vidya Somashekarappa and Christine Howes (2021). 'Looking for Laughs: Gaze Interaction with Laughter Pragmatics and Coordination'. In: *Proceedings of the 2021 International Conference on Multimodal Interaction*. ICMI '21. Montréal, QC, Canada: Association for Computing Machinery, pp. 636–644.

Mazzocconi, Chiara, Ye Tian and Jonathan Ginzburg (2020). 'What's your laughter doing there? A taxonomy of the pragmatic functions of laughter'. In: *IEEE Transactions on Affective Computing*, pp. 1–1.

McTear, Michael (2020). *Conversational AI: Dialogue systems, conversational agents, and chatbots*. Vol. 13. 3. Morgan & Claypool Publishers, pp. 1–251.

McTear, Michael, Zoraida Callejas and David Griol (2016). *The conversational interface*. Vol. 6. 94. Springer.

Mehri, Shikib, Evgeniia Razumovskaia, Tiancheng Zhao and Maxine Eskenazi (2019). 'Pretraining Methods for Dialog Context Representation Learning'. en. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguist-*

ics. Florence, Italy: Association for Computational Linguistics, pp. 3836–3845.
Mehu, Marc (2011). 'Smiling and laughter in naturally occurring dyadic interactions: Relationship to conversation, body contacts, and displacement activities'. In: *Human Ethology Bulletin* 26.1, pp. 10–28.
Meteer, Marie W, Ann A Taylor, Robert MacIntyre and Rukmini Iyer (1995). *Dysfluency annotation stylebook for the switchboard corpus*. University of Pennsylvania Philadelphia, PA.
Mihalcea, Rada and Carlo Strapparava (2005). 'Making computers laugh: Investigations in automatic humor recognition'. In: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 531–538.
Miller, Tristan, Christian Hempelmann and Iryna Gurevych (2017). 'SemEval-2017 Task 7: Detection and Interpretation of English Puns'. In: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada: Association for Computational Linguistics, pp. 58–68.
Mills, Gregory, Eleni Gregoromichelaki, Chris Howes and Vladislav Maraev (2021). 'Influencing laughter with AI-mediated communication'. In: *Interaction Studies* 22.3, pp. 416–463.
Mirenda, Patricia, Anne Donnellan and David Yoder (1984). 'Gaze behavior: A new look at an old problem'. In: *Journal of Autism and Developmental Disorders* 13, pp. 397–409.
Moradlou, Sara, Xiaobei Zheng, Tian Ye and Jonathan Ginzburg (2021). 'Wh-Questions are understood before polar-questions: Evidence from English, German, and Chinese'. In: *Journal of Child Language* 48.1, pp. 157–183.
Morkes, John, Hadyn K. Kernal and Clifford Nass (1999). 'Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory'. In: *Human–Computer Interaction* 14.4, pp. 395–435.
Morreall, John (2011). *Comic relief: A comprehensive philosophy of humor*. Vol. 27. John Wiley & Sons.

Mrkšić, Nikola, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson and Steve Young (2017). 'Neural Belief Tracker: Data-Driven Dialogue State Tracking'. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1777–1788.

Müller, Vincent C. (2021). 'Ethics of Artificial Intelligence and Robotics'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2021. Metaphysics Research Lab, Stanford University.

Nagata, Tomohiro and Hiroki Mori (2020). 'Defining Laughter Context for Laughter Synthesis with Spontaneous Speech Corpus'. In: *IEEE Trans. Affect. Comput.* 11.3, pp. 553–559.

Navarretta, Costanza, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen and Patrizia Paggio (2012). 'Feedback in Nordic First-Encounters: a Comparative Study.' In: *LREC*, pp. 2494–2499.

Nicolle, Steve (2000). 'Communicated and non-communicated acts in relevance theory'. In: *Pragmatics* 10.2, pp. 233–245.

Niewiadomski, Radosław, Elisabetta Bevacqua, Maurizio Mancini and Catherine Pelachaud (2009). 'Greta: an interactive expressive ECA system'. In: *Proc. AAMS 2009*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1399–1400.

Niewiadomski, Radoslaw, Jennifer Hofmann, Jérôme Urbain, Tracey Platt, Johannes Wagner, Bilal Piot, Hüseyin Çakmak, Sathish Pammi, Tobias Baur, Stéphane Dupont, Matthieu Geist, Florian Lingenfelser, Gary McKeown, Olivier Pietquin and Willibald Ruch (2013). 'Laugh-aware virtual agent and its impact on user amusement'. In: *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '13, Saint Paul, MN, USA, May 6-10, 2013*. Ed. by Maria L. Gini, Onn Shehory, Takayuki Ito and Catholijn M. Jonker. IFAAMAS, pp. 619–626.

Nijholt, Anton, Andreea I Niculescu, Valitutti Alessandro and Rafael E Banchs (2017). 'Humor in Human-Computer Interaction: A Short Survey'. In: *Adjunct Conference Proceedings Interact 2017*.

Noble, Bill and Vladislav Maraev (2021). 'Large-scale text pre-training helps with dialogue act recognition, but not without fine-

tuning'. In: *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*. Groningen, The Netherlands (online): Association for Computational Linguistics, pp. 166–172.
Oatley, Keith and P.N. Johnson-Laird (2014). 'Cognitive Approaches To Emotions'. In: *Trends in Cognitive Sciences* 18.3, pp. 134–140.
Ochs, Magalie and Catherine Pelachaud (2013). 'Socially Aware Virtual Characters: The Social Signal of Smiles [Social Sciences]'. In: *IEEE Signal Processing Magazine* 30.2, pp. 128–132.
Oring, Elliott (2016). *Joking Asides: The Theory, Analysis, and Aesthetics of Humor*. University Press of Colorado.
Ozbal, Gözde and Carlo Strapparava (2012). 'Computational humour for creative naming'. In: *Computational Humor 2012*, p. 15.
Pammi, Sathish, Houssemeddine Khemiri and Gérard Chollet (2012). 'Laughter detection using alisp-based N-gram models'. In: *Proc. workshop on laughter and other non-verbal vocalisations. Dublin, Ireland*, pp. 16–17.
Pashler, Harold and Christine R Harris (2012). 'Is the replicability crisis overblown? Three arguments examined'. In: *Perspectives on Psychological Science* 7.6, pp. 531–536.
Passonneau, Rebecca J. and Bob Carpenter (2014). 'The Benefits of a Model of Annotation'. In: *TACL* 2, pp. 311–326.
Patro, Badri N., Mayank Lunayach, Deepankar Srivastava, Sarvesh Sarvesh, Hunar Singh and Vinay P. Namboodiri (2021). 'Multimodal Humor Dataset: Predicting Laughter tracks for Sitcoms'. In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 576–585.
Pennington, Jeffrey, Richard Socher and Christopher D Manning (2014). 'Glove: Global vectors for word representation'. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.
Petitjean, Cécile and Etienne Morel (2017). '"Hahaha": Laughter as a resource to manage WhatsApp conversations'. In: *Journal of Pragmatics* 110, pp. 1–19.
Pickering, Martin J. and Simon Garrod (2004). 'Toward a mechanistic psychology of dialogue'. In: *Behavioral and Brain Sciences* 27.2, pp. 169–190.

Pinheiro, Ana P, Carla Barros, Marcelo Dias and Sonja A Kotz (2017). 'Laughter catches attention!' In: *Biological psychology* 130, pp. 11–21.

Plessner, Helmuth (1970). *Laughing and crying: a study of the limits of human behavior*. Northwestern University Press.

Poppe, Ronald, Khiet P. Truong and Dirk Heylen (2011). 'Backchannels: Quantity, Type and Timing Matters'. In: *Intelligent Virtual Agents - 11th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings*. Ed. by Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella and Kristinn R. Thórisson. Vol. 6895. Lecture Notes in Computer Science. Springer, pp. 228–239.

Poyatos, Fernando (1993). *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sounds*. Vol. 92. John Benjamins Publishing.

Provine, Robert R (1993). 'Laughter punctuates speech: Linguistic, social and gender contexts of laughter'. In: *Ethology* 95.4, pp. 291–298.

— (2004). 'Laughing, tickling, and the evolution of speech and self'. In: *Current Directions in Psychological Science* 13.6, pp. 215–218.

Prusak, Bernard G (2005). 'The science of laughter: Helmuth Plessner's Laughing and Crying revisited'. In: *Continental philosophy review* 38.1, pp. 41–69.

Purandare, Amruta and Diane Litman (2006). 'Humor: Prosody analysis and automatic recognition for f* r* i* e* n* d* s'. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 208–215.

Purver, Matthew (2001). *SCoRE: A Tool for Searching the BNC*. Tech. rep. TR-01-07. Department of Computer Science, King's College London.

— (2006). 'CLARIE: Handling Clarification Requests in a Dialogue System'. In: *Research on Language & Computation* 4.2, pp. 259–288.

Purver, Matthew, Arash Eshghi and Julian Hough (2011). 'Incremental Semantic Construction in a Dialogue System'. In: *Pro-*

ceedings of the 9th International Conference on Computational Semantics. Ed. by J. Bos and S. Pulman. Oxford, UK, pp. 365–369.

Pustejovsky, James (1991). 'The Generative Lexicon'. In: *Comput. Linguist.* 17.4, pp. 409–441.

Raclaw, Joshua and Cecilia E Ford (2017). 'Laughter and the management of divergent positions in peer review interactions'. In: *Journal of Pragmatics* 113, pp. 1–15.

Rajendran, Pavithra, Danushka Bollegala and Simon Parsons (2016). 'Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews'. In: *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pp. 31–39.

Raskin, Victor (1985). *Semantic mechanisms of humor*. Synthese language library, 24. Dordrecht: Reidel.

— (2017). 'Script-based semantic and ontological semantic theories of humor'. In: *The Routledge handbook of language and humor*. Routledge, pp. 109–125.

Raskin, Victor and Salvatore Attardo (1994). 'Non-Literalness and Non-Bona-Fîde in Language: An Approach to Formal and Computational Treatments of Humor'. In: *Pragmatics and Cognition* 2.1, pp. 31–69.

Reddy, Vasudevi, Emma Williams and Amy Vaughan (2002). 'Sharing humour and laughter in autism and Down's syndrome'. In: *British journal of psychology* 93.2, pp. 219–242.

Ribeiro, Eugénio, Ricardo Ribeiro and David Martins de Matos (2019). 'Deep Dialog Act Recognition using Multiple Token, Segment, and Context Information Representations'. In: *J. Artif. Intell. Res.* 66, pp. 861–899.

Rieger, Caroline L (2003). 'Repetitions as self-repair strategies in English and German conversations'. In: *Journal of Pragmatics* 35.1, pp. 47–69.

Rieser, Verena and Oliver Lemon (2011). *Reinforcement Learning for Adaptive Dialogue Systems - A Data-driven Methodology for Dialogue Management and Natural Language Generation*. Theory and Applications of Natural Language Processing. Springer.

Ritchie, Graeme (2004). *The linguistic analysis of jokes*. London: Routledge.
— (2005). 'Computational Mechanisms for Pun Generation'. In: *Proceedings of the Tenth European Workshop on Natural Language Generation, ENLG 2005, Aberdeen, UK, August 8-10, 2005*. Ed. by Graham Wilcock, Kristiina Jokinen, Chris Mellish and Ehud Reiter. ACL.
— (2018). *The Comprehension of Jokes: A Cognitive Science Framework*. London: Routledge.
Roberts, Craige (2012). 'Information structure in discourse: Towards an integrated formal theory of pragmatics'. In: *Semantics and Pragmatics* 5.6, pp. 1–69.
Roller, Stephen, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau and Jason Weston (2021). 'Recipes for Building an Open-Domain Chatbot'. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*. Ed. by Paola Merlo, Jörg Tiedemann and Reut Tsarfaty. Association for Computational Linguistics, pp. 300–325.
Romaniuk, Tanya (2009). 'The 'Clinton Cackle': Hillary Rodham Clinton's Laughter in News Interviews'. In: *Crossroads of Language, Interaction, and Culture* 7, pp. 17–49.
Rossano, Federico (2013). 'Gaze in Conversation'. In: *The Handbook of Conversation Analysis*. Ed. by Jack Sidnell and Tanya Stivers. John Wiley & Sons. Chap. 15, pp. 308–329.
Rossano, Federico, Penelope Brown and Stephen C Levinson (2009). 'Gaze, questioning and culture'. In: *Conversation Analysis: Comparative Perspectives*, pp. 187–249.
Ruiter, Jan P de and Saul Albert (2017). 'An appeal for a methodological fusion of conversation analysis and experimental psychology'. In: *Research on Language and Social Interaction* 50.1, pp. 90–107.
Sacks, Harvey, Emanuel Schegloff and Gail Jefferson (1974). 'A simplest systematics for the organization of turn-taking in Conversation'. In: *Language* 50, pp. 325–345.

Sandgren, Olof, Richard Andersson, Joost van de Weijer, Kristina Hansson and Birgitta Sahlén (2012). 'Timing of gazes in child dialogues: A time-course analysis of requests and back channelling in referential communication'. In: *International Journal of Language & Communication Disorders* 47.4, pp. 373–383.

Schegloff, Emanuel A (1968). 'Sequencing in Conversational Openings'. In: *American Anthropologist* 70.6, pp. 1075–1095.

— (1982). 'Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences'. In: *Analyzing discourse: Text and talk* 71, p. 93.

— (1992). 'To Searle on Conversation: A Note in Return'. In: *(On) Searle on Conversation*. Ed. by Herman Parret and Jef Verschueren. Amsterdam: John Benjamins Publishing Company, pp. 113–128.

Schegloff, Emanuel A and Harvey Sacks (1973). 'Opening up closings'. In: *Semiotica* 8.4, pp. 289–327.

Scherer, Klaus R (2009). 'The dynamic architecture of emotion: Evidence for the component process model'. In: *Cognition and emotion* 23.7, pp. 1307–1351.

Schlangen, David (2003). 'A coherence-based approach to the interpretation of non-sentential utterances in dialogue'. PhD thesis. University of Edinburgh. College of Science and Engineering.

Schlöder, Julian, Ellen Breitholtz and Raquel Fernández (2016). 'Why?' In: *Proceedings of JerSem*, pp. 5–14.

Schneier, Franklin R, Thomas L Rodebaugh, Carlos Blanco, Hillary Lewin and Michael R Liebowitz (2011). 'Fear and avoidance of eye contact in social anxiety disorder'. In: *Comprehensive psychiatry* 52.1, pp. 81–87.

Scott, Sophie K, Nadine Lavan, Sinead Chen and Carolyn McGettigan (2014). 'The social life of laughter'. In: *Trends in cognitive sciences* 18.12, pp. 618–620.

Searle, John R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, England: Cambridge Univesity Press.

— (1992). 'Conversation'. In: *(On) Searle on Conversation*. Ed. by Herman Parret and Jef Verschueren. Amsterdam: John Benjamins Publishing Company, pp. 1–34.

Shah, Huma and Kevin Warwick (2017). 'Machine humour: examples from Turing test experiments'. In: *AI & SOCIETY* 32.4, pp. 553–561.

Shalyminov, Igor, Arash Eshghi and Oliver Lemon (2017). 'Challenging Neural Dialogue Models with Natural Data: Memory Networks Fail on Incremental Phenomena'. In: *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Saarbrücken, Germany: SEMDIAL.

Skantze, Gabriel (2017). 'Towards a General, Continuous Model of Turn-taking in Spoken Dialogue using LSTM Recurrent Neural Networks'. In: *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Saarbrücken, Germany, August 15-17, 2017*. Ed. by Kristiina Jokinen, Manfred Stede, David DeVault and Annie Louis. Association for Computational Linguistics, pp. 220–230.

— (2021). 'Turn-taking in Conversational Systems and Human-Robot Interaction: A Review'. In: *Computer Speech & Language* 67, pp. 101–178.

Somashekarappa, Vidya, Christine Howes and Asad Sayeed (2020). 'An Annotation Approach for Social and Referential Gaze in Dialogue'. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 759–765.

Soury, Mariette and Laurence Devillers (2014). 'Smile and Laughter in Human-Machine Interaction: a study of engagement'. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA).

Speer, Robyn, Joshua Chin and Catherine Havasi (2017). 'ConceptNet 5.5: An Open Multilingual Graph of General Knowledge'. In: pp. 4444–4451.

Sperber, Dan and Deirdre Wilson (1995). *Relevance: communication and cognition*. 2nd ed. Oxford: Blackwell.

Stalnaker, Robert C (1978). 'Assertion'. In: *Pragmatics*. Brill, pp. 315–332.

Stanley, Gordon and Donald S Martin (1968). 'Eye-contact and the recall of material involving competitive and noncompetitive associations'. In: *Psychonomic Science* 13.6, pp. 337–338.

Stevens, Catherine J., Bronwyn Pinchbeck, Trent Lewis, Martin Luerssen, Darius Pfitzner, David M. W. Powers, Arman Abrahamyan, Yvonne Leung and Guillaume Gibert (2016). 'Mimicry and expressiveness of an ECA in human-agent interaction: familiarity breeds content!' In: *Computational Cognitive Science* 2.1.

Stevenson, Marguerite B, James N Ver Hoeve, Mary A Roach and Lewis A Leavitt (1986). 'The beginning of conversation: Early patterns of mother-infant vocal responsiveness'. In: *Infant behavior and Development* 9.4, pp. 423–440.

Stock, Oliviero and Carlo Strapparava (2005). 'Hahacronym: A computational humor system'. In: *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*. Association for Computational Linguistics, pp. 113–116.

Stolcke, Andreas, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema and Marie Meteer (2000). 'Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech'. en. In: *Computational Linguistics* 26.3, pp. 339–373.

Stone, Matthew (2004). 'Communicative Intentions and Conversational Processes in Human-Human and Human-Computer Dialogue'. In: *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*. Ed. by John Truesdell and Michael Tannenhaus. Cambridge, Massachusetts: The MIT Press.

Suls, Jerry M (1972). 'A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis'. In: *The psychology of humor: Theoretical perspectives and empirical issues* 1, pp. 81–100.

Tahon, Marie and Laurence Devillers (2015). 'Laughter detection for on-line human-robot interaction'. In: *Interdisciplinary Workshop on Laughter and Non-verbal Vocalisations in Speech*.

Taylor, Julia M (2009). 'Computational detection of humor: A dream or a nightmare? the ontological semantics approach'. In: *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. Vol. 3. IEEE, pp. 429–432.

Taylor, Julia M and Lawrence J Mazlack (2004). 'Computationally recognizing wordplay in jokes'. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 26. 26.

Tepperman, Joseph, David Traum and Shrikanth Narayanan (2006). '"Yeah Right": Sarcasm Recognition for Spoken Dialogue Systems'. In: *Ninth International Conference on Spoken Language Processing*.

Thagard, Paul (2005). *Mind: Introduction to cognitive science*. MIT press.

Tian, Ye, Chiara Mazzocconi and Jonathan Ginzburg (2016). 'When do we laugh?' In: *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 360–369.

Tibballs, Geoff (2000). *The mammoth book of humor*. Running Press.

Tits, Noé, Kevin El Haddad and Thierry Dutoit (2020). 'Laughter Synthesis: Combining Seq2seq Modeling with Transfer Learning'. In: *Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25-29 October 2020*. Ed. by Helen Meng, Bo Xu and Thomas Fang Zheng. ISCA, pp. 3401–3405.

Torres, Obed, Justine Cassell and Scott Prevost (1997). 'Modeling gaze behavior as a function of discourse structure'. In: *First International Workshop on Human-Computer Conversation*.

Truong, Khiet P and David A Van Leeuwen (2007). 'Automatic discrimination between laughter and speech'. In: *Speech Communication* 49.2, pp. 144–158.

Tsujita, Hitomi and Jun Rekimoto (2011). 'HappinessCounter: smile-encouraging appliance to increase positive mood'. In: *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume, Vancouver, BC, Canada, May 7-12, 2011*. Ed. by Desney S. Tan, Sale-

ema Amershi, Bo Begole, Wendy A. Kellogg and Manas Tungare. ACM, pp. 117–126.
Turing, A. M. (1950). 'Computing Machinery and Intelligence'. In: *Mind*. New Series 59.236, pp. 433–460.
Urbain, Jérôme and Thierry Dutoit (2011). 'A phonetic analysis of natural laughter, for use in automatic laughter processing systems'. In: *International Conference on Affective Computing and Intelligent Interaction*. Springer, pp. 397–406.
Vettin, Julia and Dietmar Todt (2004). 'Laughter in conversation: Features of occurrence and acoustic structure'. In: *Journal of Nonverbal Behavior* 28.2, pp. 93–115.
Wang, Pei, Xiang Li and Patrick Hammer (2018). 'Self in NARS, an AGI System'. In: *Frontiers in Robotics and AI* 5.
Warren, Caleb and A Peter McGraw (2016). 'Differentiating what is humorous from what is not.' In: *Journal of Personality and Social Psychology* 110.3, p. 407.
Wen, Tsung-Hsien, David Vandyke, Nikola Mrksic, Milica Gasic, Lina Maria Rojas-Barahona, Pei-Hao Su, Stefan Ultes and Steve J. Young (2017). 'A Network-based End-to-End Trainable Task-oriented Dialogue System'. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*. Ed. by Mirella Lapata, Phil Blunsom and Alexander Koller. Association for Computational Linguistics, pp. 438–449.
Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz and Jamie Brew (2019). 'HuggingFace's Transformers: State-of-the-Art Natural Language Processing'. In: *arXiv:1910.03771 [cs]*.
Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey et al. (2016). 'Google's neural machine translation system: Bridging the gap between human and machine translation'. In: *arXiv preprint arXiv:1609.08144*.

Xu, Kaibin, Junpei Zhong and Kristiina Jokinen (2021). 'It is Time to Laugh: Discovering Specific Contexts for Laughter with Attention Mechanism'. In: *2021 IEEE 4th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 211–215.

Yablo, Stephen (2014). *Aboutness*. Princeton, New Jersey: Princeton University Press.

Yang, Diyi, Alon Lavie, Chris Dyer and Eduard Hovy (2015). 'Humor recognition and humor anchor extraction'. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2367–2376.

Young, Steve, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson and Kai Yu (2010). 'The hidden information state model: A practical framework for POMDP-based spoken dialogue management'. In: *Computer Speech & Language* 24.2, pp. 150–174.

Yus, Francisco (2003). 'Humor and the search for relevance'. In: *Journal of pragmatics* 35.9, pp. 1295–1331.

Zhang, Renxian and Naishi Liu (2014). 'Recognizing humor on twitter'. In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, pp. 889–898.

Index

A

accommodation, 181, 198
action ladder, 29, 139, 145–149
affective computing, 74
affordance, 24
agency, 12
Agenda, 134–136
Amazon
    Alexa, 9
    Mechanical Turk, 60–61, 69
ambiguity, 52, 89, 115, 126, 131
AMT, *see* Amazon, Mechanical Turk
answer
    dispreferred, 88, 112
    integration, 137
    short, 132, 138
apology, 29, 103, 105, 109–110, 145, 149, 152
appraisal, 23, 173, 189, 199
arousal, 17, 78, 93
assertion, 133, 138
awkwardness, 152, 154
axiom, 127

B

backchannel, 20, 28, 69, 121–122, 234
    relevance space, 48
Bayesian networks, 221
behaviour
    linguistic, 10
    manipulation, 10
benevolence induction, 108, 109
BERT, *see* Transformer
bias, 11, 12
BNC, *see* corpus, British national
bonding, 3, 15

C

CA, *see* conversational, Analysis
change, *see* update
chatbot, 35
clarification request, 17, 31, 33, 36, 39, 125, 126, 138, 147, 233
CNN, *see* neural network, convolutional
cognitive
    load, 74
    science, 10, 33
collaborative constructions, 27
commitment, 147

communicative intent, *see* dialogue act
competence & performance, 27, 163, 165
Computational-Representational Understanding of Mind, 33
Concorde fallacy, 10
confidence score, 139, 147, 151
confusion matrix, 115
constant, 128
content
  modification, *see* update
  propositional, 5, 23, 87
conversational
  agent
    embodied, 5, 36, 37, 39, 45, 74, 92, 232, 233
  Analysis, 20, 32, 41, 165, 229
  relevance, 31
corpus
  AMI-DA, 114
  British national, 3, 49, 51, 165, 183, 218
  CHILDES, 218
  Directory Enquiries, 143, 215
  Good Housekeeping Institute, 76
  OpenSubtitles, 119
  studies, 32, 34
  Switchboard dialogue act, 3, 101, 114
criticism, 3, 19, 112, 189

D

DA, *see* dialogue act
DAR, *see* dialogue act recognition
dark pattern, 10
deep
  fake, 10
  learning, 10, 11
DGB, *see* dialogue, GameBoard
dialogue
  act, 3, 5, 28, 70, 71, 87, 99–101
    corpus, *see* corpus, Switchboard dialogue act
    Markup in Several Layers, 100
    modification, 105
    non-verbal, 120
    recognition, 100, 112–123
  coherence, 27, 125
  Experimental Toolkit, *see* intervention
  GameBoard, 31, 134, 139, 205, 208, 210
  manager, 5, 38–39, 126, 128
    linear, 133, 221
  move, *see* move
  multi-party, 41
  state, *see* information state
    tracking, 154
  systems
    defined, 35

end-to-end, 35, 36, 39
evaluation, 157
disambiguation, *see* ambiguity
discomfort, *see* incongruity, social
disfluency, 56, 70, 114
DM, *see* dialogue manager
downplayer, 29, 103, 105, 109–110, 155
Dynamic Syntax, 33, 126, 232

E

ECA, *see* conversational agent, embodied
ELAN, 76, 80
Analysis Companion, 80
elephant, *see Cover 1–77*
joke, *see* jke203
emotion
modelling, 8, 231
positive, 13
theories of, 23, 231
emotional
appraisal, *see* appraisal
context, 18
intelligence, 13
empathy, 13
enthymeme, 19, 161, 165, 170
defined, 167
mining, 218
under discussion, 206, 209
exaggeration, 177
explainability, 11, 36

F

face
-saving tactic, 110
-threatening act, 110, 149
social, 18, 110–112, 149
feedback, 90, 139, 147, 152, 155
negative, 145
positive, 140
relevance space, 48, 69
forced reinterpretation, 181

G

gaze, 4, 71–93, 234
aversion, 74, 75, 88
elicitation, 75
generation, 74
mutual, 73, 89
window, 73, 91
genre, 135
GHI, *see* corpus, Good Housekeeping Institute
Google Duplex, 10
grounding, 5, 29–30, 36, 139–142, 144
GTVH, *see* humour theories

H

humour
blue, 190
competence, 163, 165
computational, 213–215
generation, 214
in dialogue systems, 215
Jewish, 200
political, 19, 181
recognition, 214
sexist, 24

theories, 164–166
GTVH, 164
OSTH, 219
SSTH, 164

I
identity construction, 20
implicit
assumption, *see* topos
impropriety, 24, 108, 191, 192, 199
incongruity, 19, 23, 161, 188
pleasant, 21, 75, 78, 89–91, 216
pragmatic, 21
resolution, 165
social, 21, 71, 75, 78, 90, 109, 112, 216
theory, 164
types of, 21, 76, 78, 80–84, 88–91
incrementality, 185
inference
rule, 127, 129, 169, 198, 207
information state, 5, 30, 128, 133, 205
update, 30, 126
initiative, 36
intent, *see* dialogue act
inter-annotator agreement, 29, 49, 60, 77–79
interactive communication management, 30, 139, 150
intersubjectivity, 20
intervention, 32, 69, 233
intrusion, 144, 149–150
irony, 52, 170

J
joke, *see also* humour
bustle, 168
canned, 4, 6, 165
chicken, 179, 180
elephant, 187, 203
enhancers, 197
flu, 180
knock-knock, 183
lockdown, 180, 185
maximisers, 197
outdated, 180

K
KoS, 31–33

L
ladder, *see* action ladder
language
model, 100, 154, 232
tutoring, 13, 155–156
laughable, 5, 6, 19, 21, 23, 47–53, 78, 87, 88, 152, 170
defined, 19
laughter
antiphonal, *see* laughter coordination
as a response, 108, 122, 148
as an answer, 150–152
as disbelief, 148
bouts, 22
causes of, 15–25
clarification request, 233

coactive, *see* laughter coordination
coordination, 5, 72, 76, 78, 80–81, 84–87, 91
detection, 40
elicitation, 70, 75, 91
ethics, 9
humorous, 6, 163, 215
involuntary, 15
isolated, *see* laughter coordination
non-humorous, 5, 20, 150
prediction, 41, 55–70, 233
profile, 103–105
reciprocation, 5, 91
relevance space, 48, 55–70, 233
synthesis, 39–40
taxonomy, 21, 87
to get attention, 75
voluntary, 15
LDM, *see* dialogue manager, linear
linear
arrow, 128
hypothesis, 129
logic, *see* logic, linear
rule, 129
logic
first-order, 127
intuitionistic, 127
linear, 126, 128

M

metavariable, 126, 128, 130
misinterpretation, 17
moral status, 12
move, 31, 99, 134

N

natural language
generation, 5, 45, 133, 136, 233
understanding, 5, 45, 133, 135, 139, 146
negation, 32
neural network
convolutional, 63, 113
deep, *see* deep, learning
recurrent, 62, 113
NLG, *see* natural language, generation
NLU, *see* natural language, understanding
non-sentential utterance, *see* utterance
non-verbal
behaviour, 5, 75, 92, 122
dialogue act, *see* dialogue, act
social signal, 3, 7

O

ontology, 219
over-answering, 30

P

performative equivalence, 12
perspective-taking, 163, 166
phrase boundaries, 47
Politeness theory, 110–112, 149, 192
Prolog, 126, 128

proof search, 127
propositional content, *see* content
prosody, 70

Q

QUD, *see* question, under discussion
question
  -answering, 126, 130
  polar, 130
    negative, 130
  rejection, 152
  rhetorical, 112, 115, 131
  tag-, 111
  under discussion, 31, 131, 135, 137, 151, 223
    incrementation, 137
  wh-, 131
  why, 218

R

reasoning
  enthymematic, *see* enthymeme
  probabilistic, 230
  under uncertainty, 230
Relevance Theory, 229
repair, 33
  self-, 125
rhetorical
  question, *see* question
  resources, 12, 173, *see also* topos, 211, 217
RNN, *see* neural network, recurrent
robot
  for elderly, 16
  rights, 12
  social, 36, 75
  Sophia, 12
rule
  of inference, *see* inference

S

sarcasm, 16, 101, 198
script
  opposition, 164
  overlap, 164
self-attention, *see* Transformer
sentiment analysis, 59–60, 69
seriousness, 115
sincerity, 115
singular value decomposition, 105
smoothing, 17, 19, 105, 108, 144, 152, 154
sneezing, 15
social
  face, *see* face
  hierarchy, 18
  incongruity, *see* incongruity, social
  robot, *see* robot, social
  situation, 18
  stress, 74, 75, 88
specification, 207
speech
  act, 28, 73
  recognition, 29, 37, 45, 139, 146, 232

synthesis, 37, 39, 232
speech-laughter, 16, 22
SSTH, *see* humour theories
Staccato algorithm, 77
state update, *see* information state
destructive, 128
sunk cost fallacy, *see* Concorde fallacy
surprise, 153
Surströmming, 197
SVD, *see* singular value decomposition
SWDA, *see* corpus, Switchboard dialogue act
syntactic structure, 5

T

taboo, 12, 190
Theory of Types with Records, 32, 127, 204–205
tickling, 15
time-saving, 154
topic change, 36, 41
topos, 19, 161, 165, 167–170, 219
accommodation, 182, 189
aquisition, 178, 197
decay, 180, 197
defined, 167
linking, 179
mapping, 174
novelty, 179
salience, 174, 220, 223
Transformer, 100, 101, 114, 119
transition-relevant place, 48, 65, 68, 69
TRP, *see* transition-relevant place
TTR, *see* Theory of Types with Records
Turing test, 9, 35
turn-taking, 20, 28, 68, 73, 155

U

uncertainty, 126, 139
unlaughter, 24, 192
update, 31, 33, 181
user
engagement, 41
preference, 221
utterance, 5, 27, 101
non-sentential, 27, 31, 39, 125

V

voice user interface, 35

W

word
embedding, 113
piece tokeniser, 114