ASPECTS OF SEMANTICS AND PRAGMATICS OF DIALOGUE

SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Edited by
Paweł Łupkowski and Matthew Purver

POLISH SOCIETY
FOR COGNITIVE SCIENCE
POZNAŃ 2010

ASPECTS OF SEMANTICS AND PRAGMATICS OF DIALOGUE

SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Edited by
Paweł Łupkowski and Matthew Purver

POLISH SOCIETY
FOR COGNITIVE SCIENCE
POZNAŃ 2010

ISBN 978-83-930915-0-8

This book contains papers presented at PozDial, the 14th edition of SemDial, which took place in Poznań, Poland, 16-18 June, 2010.
The SemDial Workshops aim at bringing together researchers working on the semantics and pragmatics of dialogue in fields such as formal semantics and pragmatics, artificial intelligence, computational linguistics, psychology, and neural science.
SemDial Workshop Series homepage: http://www.illc.uva.nl/semdial/

Acknowledgements

PROGRAMME CHAIRS
Matthew Purver (Queen Mary University of London)
Andrzej Wiśniewski (Adam Mickiewicz University in Poznań)

INVITED SPEAKERS
Dale Barr (University of Glasgow)
Jonathan Ginzburg (King's College, London)
Jeroen Groenendijk (University of Amsterdam)
Henry Prakken (Utrecht University, The University of Groningen)

PROGRAMME COMMITTEE
Ron Artstein (USC Institute for Creative Technologies), Nicholas Asher (CNRS Laboratoire IRIT), Luciana Benotti (INRIA Nancy Grand Est), Anton Benz (Syddansk University), Johan Bos (Università di Roma "La Sapienza"), Harry Bunt (Tilburg University), Donna Byron (The Ohio State University), Robin Cooper (Göteborg University), Paul Dekker (ILLC/University of Amsterdam), David DeVault (USC Institute for Creative Technologies), Jens Edlund (Royal Technical Institute (KTH)), Raquel Fernández (University of Amsterdam), Mary Ellen Foster (Heriot-Watt University), Claire Gardent (CNRS/LORIA), Jonathan Ginzburg (King's College, London), Eleni Gregoromichelaki (King's College London), Joakim Gustafson (Royal Technical Institute (KTH)), Peter Heeman (Oregon Health & Science University), Pat Healey (Queen Mary University of London), Anna Hjalmarsson (Centre for Speech Technology, KTH), Ruth Kempson (King's College London), Alistair Knott (University of Otago), Alexander Koller (Saarland University), Ivana Kruijff-Korbayova (Saarland University), Nicolas Maudet (LAMSADE, Univ. Paris-Dauphine), John Niekrasz (School of Informatics, University of Edinburgh), Stanley Peters (Stanford University), Ron Petrick (School of Informatics, University of Edinburgh), Martin Pickering (University of Edinburgh), Paul Piwek (The Open University), Massimo Poesio (University of Essex), Hannes Rieser (Bielefeld University), David Schlangen (University of Potsdam), Gabriel Skantze (Royal Technical Institute (KTH)), Mark Steedman (University of Edinburgh), Amanda Stent (Stony Brook University), Matthew Stone (Rutgers University), David Traum (ICT USC), Marilyn Walker(University of Sheffield)

LOCAL ORGANIZING COMMITTEE
Katarzyna Budzyńska-Nowacka (Cardinal Stefan Wyszyński University in Warsaw)
Maria Golka (Adam Mickiewicz University in Poznań)
Dorota Leszczyńska-Jasion (Adam Mickiewicz University in Poznań)
Paweł Łupkowski (Adam Mickiewicz University in Poznań)
Jerzy Pogonowski (Adam Mickiewicz University in Poznań)
Joanna Szwabe (Adam Mickiewicz University in Poznań)
Mariusz Urbański (Adam Mickiewicz University in Poznań)

ENDORSED BY
SIGSEM
SIGdial

Printed from camera-ready manuscripts submitted by the authors.

Preface

The papers collected in this book cover a range of topics in semantics and pragmatics of dialogue. All these papers were presented at SemDial 2010, the 14th Workshop on the Semantics and Pragmatics of Dialogue. This 14th edition in the SemDial series, also known as PozDial, took place in Poznań (Poland) in June 2010, and was organized by the Chair of Logic and Cognitive Science (Institute of Psychology, Adam Mickiewicz University).

From over 30 submissions overall, 14 were accepted as full papers for plenary presentation at the workshop, and all are included in this book. In addition, 10 were accepted as posters, and are included here as 2-4 page short papers. Finally, we also include abstracts from our keynote speakers.

We hope that the ideas gathered in this book will be a valuable source of up-to-date achievements in the field, and will become a valuable inspiration for new ones.

We would like to express our thanks to all those who submitted to and participated in SemDial 2010, especially the invited speakers: Dale Barr (University of Glasgow), Jonathan Ginzburg (King's College London), Jeroen Groenendijk (University of Amsterdam) and Henry Prakken (Utrecht University, The University of Groningen).

Last but not least, we would like to thank everybody engaged in the workshop organization – the chairs, the local organizing committee for their hard work in Poznań, and the programme committee members for their thorough and helpful reviews.

Editors

Table of contents

PAPERS

Dialogue act models

Volha Petukhova, Harry Bunt and Andrei Malchanau
Empirical and theoretical constraints on dialogue act combinations 1

Alexander Koller, Andrew Gargett and Konstantina Garoufi
A scalable model of planning perlocutionary acts 9

Coordination

Gregory Mills and Eleni Gregoromichelaki
Establishing coherence in dialogue: adjacency, intentions and negotiation 17

David Schlangen
Practices in Dialogue 25

Incrementality

Okko Buss and David Schlangen
Modelling Sub-Utterance Phenomena in Spoken Dialogue Systems 33

Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol and Ronnie Cann
Splitting the 'I's and Crossing the 'You's: Context, Speech Acts and Grammar 43

Rational agents

Matthew Stone and Alex Lascarides
Coherence and Rationality in Grounding 51

Katarzyna Budzynska and Kamila Dębowska
Dialogues with conflict resolution: goals and effects 59

Erotetic logic

Mariusz Urbański and Paweł Łupkowski
Erotetic Search Scenarios: Revealing Interrogator's Hidden Agenda 67

Paweł Łupkowski
Cooperative answering and Inferential Erotetic Logic 75

Semantics

Staffan Larsson
Accommodating innovative meaning in dialogue 83

Robin Cooper
Generalized quantifiers and clarification content 91

Reference

Florian Hahn and Hannes Rieser
Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology 99

Raquel Fernandez
Early Interpretation by Implicature in Definite Referential Descriptions 111

Table of contents

INVITED SPEAKERS

Dale Barr
On the distributed nature of mutual understanding 119

Jonathan Ginzburg
Relevance for Dialogue 121

Jeroen Groenendijk
Radical Inquisitive Semantics 131

Henry Prakken
Argumentation in Artificial Intelligence 133

SHORT PAPERS

Ellen Breitholtz
Clarification requests as enthymeme elicitors 135

Jenny Brusk
A Computational Model for Gossip Initiation 139

Nina Dethlefs, Heriberto Cuayahuitl, Kai-Florian Richter, Elena Andonova and John Bateman
Evaluating Task Success in a Dialogue System for Indoor Navigation 143

Marta Gatius and Meritxell Gonzalez
Guiding the User When Searching Information on the Web 147

Maria H. Golka
Semantics and pragmatics of negative polar questions 149

Volha Petukhova and Harry Bunt
Context-driven dialogue act generation 151

Joanna Szwabe and Anna Brzezińska
The Communicative Style of the Physically Disabled – a Corpus Study 153

Thora Tenbrink and Elena Andonova
Communicating routes to older and younger addressees 155

Marcin Włodarczak, Harry Bunt and Volha Petukhova
Entailed feedback: evidence from a ranking experiment 159

Michael Wunder and Matthew Stone
Statistical Evaluation of Intention in Group Decision-making Dialogue 163

Empirical and theoretical constraints on dialogue act combinations

Volha Petukhova and **Harry Bunt** and **Andrei Malchanau**
Tilburg Center for Creative Computing
Tilburg University, The Netherlands
{v.petukhova,harry.bunt}@uvt.nl; a.malchanau@concepts.nl

Abstract

This paper presents an empirical study and analytical examination of the actual and possible co-occurrence of dialogue acts in dialogue units of various sorts. We formulate semantic and pragmatic constraints on dialogue act combinations for various types of dialogue unit.

1 Introduction

One of the reasons why people can communicate efficiently is because they use linguistic and nonverbal means to address several aspects of the communication at the same time. Consider, for example, the following dialogue fragment[1]:

(1) *U1: What is RSI?*
S1: RSI stands for Repetitive Strain Injury
U2: Yes but what is it?
S2: Repetitive Strain Injury is an infliction where...

Utterance (U2) in 1 indicates that (1) the user interpreted the system's previous utterance (S1) successfully (signalled by 'Yes'); (2) the system did not interpret utterance (U1) as intended (signalled by 'but'); and (3) the user requests information about the task domain. If the system does not recognize all three functions, it will most likely resolve the anaphoric pronoun 'it' as coreferential with 'RSI' and interpret (U2) as a repetition of (U1), and thus not be able to react properly.

This example shows that the multifunctionality of utterances must be taken into account in order to avoid errors and misunderstandings, and to support a dialogue that is effective and efficient.

While the multifunctionality of dialogue utterances has been widely recognised (Allwood, 2000; Bunt, 2000; Popescu-Belis, 2005), computationally oriented approaches to dialogue generally see multifunctionality as a problem, both for the development of annotation schemes and for the design of dialogue systems (Traum, 2000). Information that may be obtained through a multifunctional analysis is often sacrificed for simplicity in computational modelling. As a consequence, the actual multifunctionality of dialogue utterances are still understudied (though see Bunt, 2010).

The present study is concerned with the forms of multifunctionality that occur in natural dialogue and the relations between the communicative functions of a multifunctional dialogue units (Section 3). In Section 4 we formulate the semantic and pragmatic constraints on the multifunctionality of dialogue units. Section 5 ends with conclusions and prospects for future research.

2 Semantic framework

We used the semantic framework of Dynamic Interpretation Theory (DIT, Bunt, 2000), which takes a multidimensional view on dialogue in the sense that participation in a dialogue is viewed as performing several activities in parallel, such as pursuing the dialogue task, providing and eliciting feedback, and taking turns. The activities in these various 'dimensions' are called *dialogue acts* and are formally interpreted as update operations on the information states of the dialogue participants and have two main components: a *semantic content* which is to be inserted into, to be extracted from, or to be checked against the current information state; and a *communicative function*, which specifies more precisely how an addressee updates his information state with the semantic content when he understands the corresponding aspect of the meaning of a dialogue utterance.

A communicative function captures beliefs and intentions of the speaker. For instance, the preconditions to perform an Answer are: (1) Speaker (S) believes that Addressee (A) wants to have some information, and (2) S believes that the informa-

[1] From a dialogue with the IMIX system translated from Dutch - see (Keizer & Bunt, 2007).

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

tion is true. Applying this to a particular semantic content type, e.g. Auto-Feedback, gives the following: (1) S believes that A wants to know about S's processing state, and (2) S believes that the information about S's processing state is true.

The DIT taxonomy of communicative functions distinguishes 10 dimensions, addressing information about the task or domain (*Task*), speaker's processing of the previous utterance(s) (*Auto-feedback*) or this of the addressee (*Allo-feedback*), difficulties in the speaker's contributions (*Own-Communication Management - OCM*) or those of the addressee (*Partner Communication Management- PCM*), the speaker's need for time (*Time Management*), maintaining contact (*Contact Management*), allocation of speaker role (*Turn Management*), future structure of dialogue (*Dialogue Structuring - DS*), and social constraints (*Social Obligations Management- SOM*).

Some communicative functions can be combined with only one particular type of information, such as Turn Grabbing, which is concerned with the allocation of the speaker role. Being specific for a particular dimension, these functions are called *dimension-specific*. Other functions are not specifically related to any dimension, e.g. one can request the performance of any type of action (such as 'Please close the door' or 'Could you please repeat that'). Question, Answer, Request, Offer, Inform, and many other 'classical' functions are applicable to a wide range of semantic content types. These communicative functions are called *general-purpose* functions.

3 Forms of multifunctionality

To examine the forms of multifunctionality that occur in natural dialogue we performed a corpus analysis, using human-human multi-party interactions (AMI-meetings[2]). Three scenario-based meetings were selected containing 17335 words. Dialogue contributions were segmented at turn level (776 turns); at utterance level (2,620 utterances); and at the finer level of functional segments (see below; 3,897 functional segments). The data was annotated according to the DIT dialogue annotation scheme (DIT^{++} tagset[3]).

[2]Augmented Multi-party Interaction (`http://www.amiproject.org/`).

[3]For more information about the tagset, please visit: `http://dit.uvt.nl/`

3.1 Relations between communicative functions

The DIT^{++} tagset has been designed in such a way that two communicative functions which can be applied in the same dimension either (1) are *mutually exclusive*, or (2) one *entails* the other. Consider, for example, the Time Management dimension. The speaker may suspend the dialogue for one of several reasons and signal that he is going to resume it after a minor or a prolonged delay (Stalling or Pause, respectively). Evidently, stalling and pausing acts are mutually exclusive: they both cannot apply to one and the same segment. In the case of an entailment relation, a functional segment has a communicative function, characterized by a set of preconditions which logically imply those of a dialogue act with the same semantic content and with the entailed communicative function. For instance, more specific functions entail less specific ones, such as Agreement, Disagreement entailing Inform, and Confirm and Disconfirm entailing Propositional Answer. This intra-dimensional entailment relation is called *functional subsumption* (Bunt, 2010).

A communicative function in one dimension may also entail a function in another dimension. This inter-dimensional entailment relation occurs between responsive acts in non-feedback dimensions on the one hand and auto- and allo-feedback acts on the other. For example, accepting or rejecting an offer, suggestion, invitation or request, answering a question, responding to a greeting and accepting apology entail positive Auto-Feedback.

A functional segment may have multiple functions by virtue of its observable surface features (called *independent* multifunctionality), like wording, prosodic and acoustic features or accompanying nonverbal signals. For example, 'yes' and 'okay', said with an intonation that first falls and subsequently rises, express positive feedback and give the turn back to the previous speaker.

A functional segment may also have multiple communicative functions due to the occurrence of conversational implicatures. *Implicated* functions correspond semantically to an additional context update operation and are an important source of multifunctionality. For example, a shift to a relevant new discussion topic implicates positive feedback about the preceding discussion. In DIT^{++}, five processing levels in Auto- and Allo-Feedback also have logical relations that turn up as impli-

Table 1: *Co-occurrences of communicative functions across dimensions in one functional segment, expressed in relative frequency in %, implied functions (implicated and entailed) excluded and included.*

segments in / *have function in*	form	Task	Auto-F.	Allo-F.	Turn M.	Time M.	DS	Contact M.	OCM	PCM	SOM
Task	independent	0	1.1	0	2.2	0.1	19.6	0	3.8	0	0
	implied	49.8	47.9	24.9	97.5	2.4	31.5	0.4	69.6	0.1	0.7
Auto-F.	independent	0.7	0	0	11.0	0.6	1.9	11.1	0.8	0	0
	implied	38.9	100	0	88.7	11.4	11.2	20.2	11.7	65.0	8.7
Allo-F.	independent	0	0	0	0.1	0	0	0	0	0	0
	implied	24.9	0	100	94.8	35.7	2.1	1.2	7.9	0.7	0.3
Turn M.	independent	3.4	26.9	6.7	0	28.6	12.4	7.4	4.8	18.2	6.7
	implied	76.0	66.2	19.4	0	42.9	14.6	13.8	99.6	27.3	10.5
Time M.	independent	0.1	0.7	0	44.9	0	4.7	0	1.3	0	0
	implied	28.2	11.3	7.8	98.6	0	1.7	0	83.2	0.5	0
DS	independent	0.1	0.4	0	0.3	0	0	0.9	0	0	6.7
	implied	3.2	58.3	29.1	87.5	4.9	4.6	25.0	3.7	0	12.5
Contact M.	independent	1.7	0.3	0	3.6	0.5	3.7	0	0	0	1.3
	implied	2.4	97.1	1.6	98.8	0.5	2.4	0	0.3	0	3.7
OCM	independent	1.2	0.4	0	2.8	0.5	0	0	0	0	6.7
	implied	82.2	2.8	2.5	96.9	7.8	3.9	13.5	0	0.9	7.6
PCM	independent	0	0	0	0.3	0	0	0	0	0	0
	implied	11.8	65.0	11.8	79.1	12.2	0	0	0	0	0
SOM	independent	0	0	0	0.2	0	0	2.7	0.3	0	0
	implied	0.7	80.0	10.0	90.0	0	30.0	3.9	2.0	0	0

cations between feedback acts at different levels:

(2) *attention* $<$ *perception* $<$ *understanding* $<$ *evaluation* $<$ *execution*

The implication relations between feedback at different levels are either entailments or implicatures. In the case of positive feedback, an act at level L_i *entails* positive feedback at all levels L_j where $i > j$; positive feedback at execution level therefore entails positive feedback at all other levels. Positive feedback at level L_i *implicates* negative feedback at all levels L_j where $i < j$; for instance, a signal of successful perception implicates negative understanding. This is, however, not a logical necessity, but rather a pragmatic matter. For negative feedback the entailment relations work in the opposite direction. For allo-feedback the same relations hold as for auto-feedback.

3.2 Relations between dialogue units

Dialogues can be decomposed into *turns*, defined as stretches of speech produced by one speaker, bounded by periods of silence of that speaker. Turns consist of one or more *utterances*, linguistically defined stretches of communicative behaviour that have a communicative function. The stretches of behaviour that are relevant for interpretation as dialogue acts often coincide with utterances in this sense, but they may be discontinuous, may overlap, and may even contain parts of more than one turn. They therefore do not always correspond to utterances, which is why we have introduced the notion of a *functional segment* as a minimal stretch of communicative behaviour that has a communicative function (and possibly more than one)[4]. Thus, the units of dialogue that our analysis will be concerned with, are turns and functional segments.

There are different forms of multifunctionality. Allwood in (1992) claims that if an utterance is multifunctional, 'its multifunctionality can be sequential and simultaneous'. Bunt (2010) examines this claim using empirical data from several dialogue annotation experiments and concludes that sequential multifunctionality disappears if we take sufficiently fine-grained dialogue units into account ('functional segments' rather than turns). It was shown that even if we consider fine-grained units of communicative behaviour we do not get rid of simultaneous multifunctionality. The minimum number of functions that one segment has in dialogue is 1.3 on average and this number increases when entailed and implicated functions are taken into account.

3.2.1 Multifunctionality in segments

Our observations show that different functions in different dimensions may address the same span in the communicative channel. This what is called *simultaneous* multifunctionality. Segments may have two or more communicative functions in different dimensions. For example:

(3) *B1: Any of you anything to add to that at all?*
A1: No
D1: I'll add it later in my presentation

[4]These stretches are 'minimal' in sense of not being unnecessarily long.

Table 2: *Co-occurrences of communicative functions across dimensions in overlapping segments, expressed in relative frequency in %.*

have function in \ *segments in*	Task	Auto-F.	Allo-F.	Turn M.	Time M.	Contact M.	DS	OCM	PCM	SOM
Task	0	40.8	23.4	42.4	38.2	0	28.2	65.4	22.9	18.2
Auto-F.	10.5	6.7	16.9	16.9	19.1	18.8	19.1	14.2	54.8	9.5
Allo-F.	1.5	4.2	1.3	4.3	12.1	18.8	12.1	5.4	16.2	9.1
TurnM.	14.1	31.4	45.9	0	14.6	25.0	14.6	76.0	25.8	4.9
TimeM.	2.9	7.7	20.2	12.8	0	0	0.8	3.4	16.1	3.2
ContactM.	0.3	0.2	1.8	0.1	0	0	5.6	0	0	2.9
DS	2.1	6.9	11.4	0.2	3.9	37.5	0	5.6	0	8.2
OCM	4.6	3.8	5.8	4.4	2.3	0	2.2	0	0	1.6
PCM	0	0.9	0.9	1.2	0.7	0	0.7	0	0	0
SOM	0	0.1	1.3	2.1	0.3	23.3	0.3	0.2	0	0

In utterance B1 the speaker's intention is to elicit feedback, and the utterance also has an explicitly expressed ('any of you') turn releasing function. In utterance A1 the speaker provides an answer to B1. The speaker in utterance D1 gives no answer to B1, instead he indicates that he will provide the requested information later in the dialogue (negative Auto-Feedback act combined with Discourse Structuring act). A segment may have one or more functions by virtue of its observable features and one or more functions by implication. For example:

(4) *B1: Just to wrap up the meeting*
D1: Can we just go over the functionality again?

Utterance D1 in (3) is a request to shift the topic back to what was already discussed before. This utterance by implication has a function of negative feedback about B1, disagreeing to close dialogue as announced in B1.

Table 1 gives an overview of co-occurrences of communicative functions across dimensions for one and the same stretch of communicative behaviour simultaneously as observed in features of this behaviour, and when entailed or implicated functions occur[5]. It can be observed that functions which address the same dimension never co-occur, except for Auto- and Allo-Feedback where functions are not mutually exclusive but entail or implicate each other, and some general-purpose functions addressing different dimensions (in our data Task and Discourse Structuring) that are not mutually exclusive but a specialization of the other as discussed in Section 3.1.

Some combinations of functions are relatively frequent, e.g. time- and turn management acts often co-occur. A speaker who wants to win some time to gather his thoughts and wants to continue in the sender role, may intend his stalling behaviour to signal the latter as well (i.e., to be interpreted as a Turn Keeping act). But stalling behaviour does not *always* have that function; especially an extensive amount of stallings accompanied by relatively long pauses may be intended to elicit support for completing an utterance.

Co-occurrence scores are higher when entailed and implicated functions are taken into account (see also Bunt, 2010). An *implicated* function is for instance the positive feedback (on understanding and evaluating the preceding addressee's utterance(s)) that is implicated by an expression of thanks; examples of *entailed* functions are the positive feedback on the preceding utterance that is implied by answering a question or by accepting an invitation. Questions, which mostly belong to the Task dimension, much of the time have an accompanying Turn Management function, either releasing the turn or assigning it to another participant, allowing the question to be answered. This implicature, however, may be cancelled or suspended when the speaker does not stop speaking after asking a question. Similarly, when accepting a request the speaker needs to have the turn, so communicative functions like Accept Request will often be accompanied by function like Turn Accept. Such cases contribute to the co-occurrence score between the Turn Management and other dimensions.

3.2.2 Multifunctionality in segment sequences

Participants do not limit their dialogue contributions to functional segments; their goal is to produce coherent utterances. Utterances may be *discontinuous*, where smaller segments can be inside larger functional segments. For example, the speaker of the utterance in (5) interrupts his Inform with a Set-Question:

[5]Tables 1, 2 and 3 should be read as follows: from all identified segments addressing dimension in column, these segments have also a communicative function in dimension listed in rows.

Table 3: *Co-occurrences of communicative functions across dimensions in a sequence of two functional segments in one turn, expressed in relative frequency in %.*

segments in / *have function in*	Task	Auto-F.	Allo-F.	Turn M.	Time M.	DS	Contact M.	OCM	PCM	SOM
Task	26.5	36.5	33.3	33.5	42.4	0	15.4	21.6	20.0	46.7
Auto-F.	15.9	24.8	9.9	16.7	17.2	33.3	19.2	8.0	30.0	13.3
Allo-F.	0.4	1.1	6.6	0.6	0.6	0	0	0.5	0	0
TurnM.	59.7	38.1	36.7	53.0	44.2	15.3	61.5	69.9	50.0	33.3
TimeM.	27.9	20.4	20.0	30.9	18.8	0	15.4	55.4	0	26.7
ContactM.	0	0.1	0	0.1	0	34.2	0	0	0	54.6
DS	0.5	1.2	0	0.6	0.6	15.0	7.6	0.5	0	0
OCM	9.9	8.0	6.7	11.3	13.9	0	7.7	9.5	0	0
PCM	0.4	0.42	0	0.1	0.1	0	0	0.3	0	0
SOM	0.2	0.6	0	0.3	0.1	33.3	0	0.5	0	6.7

(5) *Twenty five Euros for a remote... **how much is that locally in pounds?** is too much to buy a new one*

Segments with different functions may *overlap* (see Table 2). For example:

(6) *B1: I think we're aiming for the under sixty five*
*D1: **Under sixty five** is a good constraint*

Utterance D1 is positive feedback about B1 at the level of evaluation, whereas the bold marked part is an explicit feedback signal at the level of perception. Such a co-occurrence is possible because higher levels of positive feedback entail lower levels of positive feedback.

The most important sources of overlapping multifunctionality are entailed functions, but here they are expressed explicitly by means of certain utterance features. For instance, as mentioned above answers entail that the previous question was successfully processed. Answers often overlap with explicitly expressed positive feedback, e.g. when the speaker repeats (positive perception) or paraphrases the partner's previous (part of) utterance (positive interpretation) in a segment within his utterance. Discourse markers may also be used for this purpose signalling that higher processing levels are reached (i.e. evaluation or execution). For example:

(7) *D1: Which is the clunky one on the left or on the right?*
*C1: **The clunky one** is the one on the right*

The speaker of C1 could have said *'on the right'* which would be a perfectly acceptable answer to the question D1. Instead, he repeats part of the question and thereby signals that his perception was successful. In the same way, Accept and Reject Offer, Suggestion and Request, but in fact any responsive, which entail positive auto-feedback, may overlap with such segments.

Another source of overlapping is pragmatic implicatures. It is often possible to add explicitly what is implicated without being redundant. For example, positive feedback implicated by shifting to a new topic, related to the previous one, may be expressed explicitly and happens very often by means of discourse markers, such as 'and then', 'okay then', 'next', etc. (see Petukhova&Bunt, 2009). More generally, any relevant continuation of the dialogue implicates positive feedback, such as question that moves the dialogue forward. But this may also be expressed by repeating or paraphrasing parts of previous utterances, or using discourse markers like 'then'. For example:

(8) *D1: This idea focuses on the twenty five age group*
*B1: Are we aiming at **a fairly young market then**?*

Functional segments following each other within a turn give rise to *sequential* multifunctionality at turn level. We analysed sequences of a length of 2 functional segments for the most frequently occurring patterns of communicative function combinations (see Table 3). It was observed that the co-occurrence scores for Turn Management, Task and Auto-Feedback with other dimensions are relatively high. This means that Task functional segments are frequently preceded or followed by Turn Management or Auto-Feedback segments or segments that have functions in these two dimensions simultaneously. For instance, a frequent pattern for constructing a turn is first performing a turn-initial act (e.g. Turn Take, Accept or Grab) combined with or followed by an Auto-Feedback act and one or more segments in another dimension, and closing up the turn with a turn-final act. This pattern occurs in about 49.9% of all turns. For example:

(9) *B1: well (Neg.Auto-Feedback Evaluation + Turn Take)*
B2: Twenty five euro is about eighteen pounds, isn't it? (Auto-Feedback Check Question)
D1: um (Turn Take+Stalling)
D2: Yep (Allo-Feedback Confirm)

Dialogue participants make their contributions consistent. To perform a task act and then to ex-

plicitly take the turn would not be a logical thing to do, because by starting speaking one already implicitly indicates that one wants to occupy the sender role. Similarly, to reject a request and then to accept it would be very unfortunate, unless the first act is performed by mistake or the speaker changes his mind and withdraws the first act.

We often observed sequences where the speaker performed a certain act and subsequently tried to justify this by elaborating or explaining what he just said. For example:

(10) *A1: it ties you on in terms of the technologies*
A2: ***like for example*** *voice recognition*
A3: ***because*** *you need to power a microphone*
A4: ***so*** *thats one constraint there*

In example (10) discourse markers are used by the speaker to indicate the steps in a sequence of arguments: he makes a statement (Inform); then provides an example for this statement (Inform Exemplify); justifies his choice (Inform Justification); and draws a conclusion (Inform Conclude).

4 Constraints on dialogue act combinations

A good understanding of the nature of the relations among the various multiple functions that a segment may have, and how these segments relate to other units in dialogue, opens the way for defining a computational update semantics for the interpretation and generation of dialogue utterances. In order to develop such a semantics, it is necessary to investigate forms of multifunctionality that occur in natural dialogue and the relations between the communicative functions of a multifunctional utterance. Moreover, no corpus is big enough to examine all possible function co-occurrences; corpus-based observations call for an additional analytical examination of the conditions for performing a certain dialogue act.

The DIT^{++} set of 10 dimensions is *orthogonal* (see Petukhova & Bunt (2009)), thus, theoretically it is possible that a segment has a communicative function in each dimension (thus, 10 tags per segment). There are, however, certain constraints on the use of functions within a dimension. The following should be taken into account: (1) that there's at most one (most specific) applicable function per dimension, and (2) the total number of functions available per dimension. DIT^{++} tagset has 44 general-purpose functions and 56 dimension specific functions. Distribution of function across dimensions is, therefore, as follows: Task dimension has 44 functions; Auto-Feedback - 54; Allo-Feedback - 59; Turn Management - 50; Time Management - 46; Contact Management - 46; DS - 50; OCM - 47; PCM - 46; and SOM - 54. A function, however, can be assigned not in each dimension. The total number of possible combinations is the sum of the possible number of 10 tags, the number of 9 tags, the number of 8 tags, ... the number of single tags. The number of possible combinations of 10 tags is $44 \times 54 \times 59 \times 50 \times 46 \times 46 \times 50 \times 47 \times 46 \times 54 = 8.66 \times 10^{16}$; adding the number of possible combinations of nine tags or less gives a total of 8.82×10^{16}.

In practice, it has been shown that 2 functions per segment is a realistic number when we count functions expressed by virtue of utterance features and implicated functions (see Bunt, 2010). This gives us $(D_1 \times D_2 + D_1 \times D_3 + D_1 \times D_4 + ...) = 110,605$ possible dialogue act combinations.

We analysed these function combinations and determine whether there are additional constraints on their combinations and what nature they have: do they have a logical or a pragmatic origin. For each dialogue act we calculated logical entailments and generated dialogue act pairs, in search of logical conflicts between them. Entailments between dialogue acts are defined by logical implications between their preconditions. Calculating the entailment relations among dialogue acts through their preconditions ensures completeness in the sense of finding all entailments between dialogue acts. While entailments depend solely on the definitions of communicative functions in terms of their preconditions, implicatures are pragmatic relations between a dialogue act and a condition that may be a precondition of another dialogue act, as will be illustrated below, and are a matter of empirical observation.

4.1 Logical constraints

From a logical point of view, two communicative functions cannot be applied to one and the same semantic content if they have logical conflicts in their preconditions or/and entailments. We analysed functional consistency pairwise between (1) preconditions of F_1 and F_2; (2) entailments of F_1 and F_2; (3) entailments of F_1 and preconditions of F_2 and vice versa.

The use of two functions (F_1 and F_2) applied to the same semantic content p is logically inconsis-

tent if there is a proposition q which can be derived from the set of preconditions P_1 of F_1, while $\neg q$ can be derived from the preconditions P_2 of F_2. This is for instance the case when we deal with alternative end-nodes in the tagset hierarchy. For example, one cannot accept and reject an offer in one functional segment: Accept Offer requires that $believes(S, will\_do\_action(A,a))$; $believes(S, can\_do(A,a))$; $believes(S, wants(A, believes(S, will\_do\_action(A,a))))$ and $wants(S, plan\_do\_action(A,p))$; for Reject Offer the same preconditions hold except for the last one which is $\neg wants(S, plan\_do\_action(A,a))$.

Similarly, F_1 and F_2 applied to the same semantic content p are logically conflicting if F_1 has an entailed condition q and F_2 has the entailment $\neg A$. For example, the entailments of an answer to a question expressed by utterance u ($wants(S, knows(A, Interpreted(S,u)))$) are in conflict with entailments of negative Auto-Feedback at the level of perception and lower (e.g. $wants(S, knows(A, \neg Perceived(S,u)))$ entails $wants(S, knows(A, \neg Interpreted(S,u)))$).

Two acts are also in conflict if the entailments of one are in logical conflict with preconditions of the other. The most obvious case is that of responsive dialogue acts and negative Auto-Feedback at all processing levels. For example, in order to provide a correction the speaker needs to have paid attention, perceived and understood the relevant previous utterance.

Note that the combination of two acts in one functional segment that share the same semantic content are not necessarily in conflict if they refer to different segments or acts in the previous discourse, i.e. if they have different *functional* or *feedback dependency relations*, see Bunt (2010).

4.2 Pragmatic constraints

Pragmatically speaking, two acts A_1 and A_2 are inconsistent in the following to cases:

(11) (1) an implicated condition q_1 of A_1 blocks the performance of A_2;
(2) an implicated condition q_1 of A_1 is in conflict with implicated condition q_2 of A_2.

An example of the first type of pragmatic inconsistency is the combination of direct and conditional (indirect) variants of the same act. For instance, a direct request like *Please tell me where Harry's office is* has the precondition that the addressee is able to perform the requested action: $believes(S, can\_do\_action(A,a))$, whereas a conditional request (like *Can you tell me where Harry's office is?*) does not have this preconditions; instead, it implicates that the speaker wants to know whether the addressee is able to perform the action ($wants(S, knowsif(S, can\_do\_action(A,a)))$).

Similarly, questions and requests implicate that the speaker wants the addressee to have the next turn, hence the speaker does not want to have the next turn himself: ($\neg wants(S, Turn\_Allocation(S))$), whereas such acts as Stallings or Pausing, but also acts like Self-Correction, Error Signalling and Retraction, implicate that the speaker wants to keep the turn himself: ($wants(S, Turn\_Allocation(S))$).

Two dialogue acts cannot be combined in one segment if an implicature of one act makes the performance of another act impossible. For example, positive auto-feedback acts at the level of perception and lower do not satisfy the conditions for the speaker to be able, for example, to assist the addressee by providing a completion or a correction of the addressee's mistakes, because for being able to offer a completion or a correction it is not sufficient to pay attention and hear what was said, but understanding and evaluation are required, and positive perception implicates negative feedback at these higher processing levels.

As noted in (11), two acts cannot be combined in one segment if implicatures of one are in conflict with implicatures of another. For instance, Contact Check carries an implicature of negative perception of partner's linguistic or non-verbal behaviour, whereas, for example, Opening carries an implicature of positive perception of partner's behaviour. Similarly, Partner Communication Management acts are pragmatically inconsistent with dialogue acts like Opening, Self-Introduction, Greeting or Contact Check, because PCM acts are performed in reaction to certain linguistic behaviour of the dialogue partner, and therefore implicate higher levels of successful processing of such behaviour, whereas dialogue initiating acts implicate lower processing levels like attention or perception, or elicit them. PCM acts can be combined with responsive acts in these dimensions although we do not find examples of this in our corpus data.

4.3 Constraints for segment sequences

We discussed above logical and pragmatic constraints for simultaneous multifunctionality. Since overlapping multifunctionality is a special case

of simultaneous multifunctionality; the constraints discussed above apply in this case as well.

For sequential multifunctionality within turns there are fewer and softer constraints on dialogue act combinations than for simultaneous multifunctionality. For example, the combination of two mutually exclusive acts in a sequence is in principle possible. A speaker who wants to constructed turn coherent and logically consistent turns should not combine logically or pragmatically conflicting dialogue acts associated with segments within the same turn. However, such combinations cannot be excluded entirely, since a speaker can perform a dialogue act by mistake and subsequently correct himself. Hence we may expect sequences of the following kind:

(12) 1. dialogue act A_1
2. retraction of A_1
3. dialogue act A_2

where A_1 and A_2 are conflicting.

5 Discussion and conclusions

The main conclusion from this study is that in order to define a multidimensional computational update semantics for dialogue interpretation it is important to understand the nature of the relations among the various multiple functions that a segment may have and how these segments relate to other units in dialogue. We investigated the forms of multifunctionality that occur in natural dialogue and analysed the obtained functions co-occurrence matrices across dimensions. Additionally, analytical examination of act preconditions, entailments implication relations was performed. General constraints on the use of dialogue act combinations were formulated. These constraints are also general in a sense that they are not only applicable when using the DIT^{++} dialogue act set but also other multidimensional tagsets such as DAMSL (Allen and Core, 1997), MRDA (Dhillon et al., 2004) and Coconut (Di Eugenio et al., 1998). These constraints are important for efficient computational modelling of dialogue and dialogue context, as well as for automatic dialogue act tagging, in that it could facilitate the effective computations and reduce the search space significantly.

The results of this study do not only have consequences for the semantic interpretation of dialogue contributions, but also for their generation. Our future work will be concerned with the automatic generation of sets of dialogue acts for contribution planning; the formulation of rules assigning priorities among alternative admissible dialogue acts; and formulating linguistic constraints on possible combinations of dialogue acts in a segment, an utterance, and a turn.

Acknowledgments

This research was conducted within the project 'Multidimensional Dialogue Modelling', sponsored by the Netherlands Organisation for Scientific Research (NWO), under grant reference 017.003.090. We are also very thankful to anonymous reviewers for their valuable comments.

References

Allen, J., Core, M. 1997. *Draft of DAMSL: Dialog Act Markup in Several Layers*.

Jens Allwood. 1992. *On dialogue cohesion*. Gothenburg Papers in Theoretical Linguistics 65. Gothenburg University.

Jens Allwood. 2000. *An activity-based approach to pragmatics*. In H. Bunt and W. Black (eds.), Abduction, Belief and Context in Dialogue. Amsterdam: Benjamin, pp. 47-81.

Harry Bunt. 2000. *Dialogue pragmatics and context specification*. H. Bunt and W. Black (eds.), Abduction, Belief and Context in Dialogue. Amsterdam: Benjamins, pp.81-150.

Harry Bunt. forthcoming 2010. *Multifunctionality in dialogue*. Computer, Speech and Language, Special issue on dialogue modeling.

Rajdip Dhillon, Sonali Bhagat, Hannah Carvey, and Elizabeth Schriberg. 2004. *Meeting recorder project: dialogue labelling guide*. ICSI Technical Report TR-04-002.

Barbara Di Eugenio, Pamela W. Jordan, and Liina Pylkkaenen. 1998. *The COCONUT project: dialogue annotation manual*. ISP Technical Report 98-1.

Simon Keizer and Harry Bunt. 2007. *Evaluating combinations of dialogue acts for generation*. In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp, pp. 158-165.

Volha Petukhova and Harry Bunt. 2009. *The independence of dimensions in multidimensional dialogue act annotation*. Proceedings NAACL HLT Conference, Boulder, Colorado.

Andrei Popescu-Belis. 2005. *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper 62, ISSCO, Geneva.

David Traum. 2000. *20 questions on dialogue act taxonomies*. Journal of Semantics, 17(1): 7-30.

A scalable model of planning perlocutionary acts

Alexander Koller and **Andrew Gargett** and **Konstantina Garoufi**
Cluster of Excellence "Multimodal Computing and Interaction"
Saarland University, Saarbrücken, Germany
{koller|gargett|garoufi}@mmci.uni-saarland.de

Abstract

We propose a new model of perlocutionary acts, in which perlocutionary effects of communicative actions are simply effects of operators in a planning problem. A plan using such operators can be computed efficiently, under the assumption that all perlocutionary effects come true as intended; the speaker then monitors the plan execution to detect it when they don't. By scaling the complexity of the execution monitor up or down, we can reconstruct previous approaches to speech act planning and grounding, or build an instruction generation system with real-time performance.

1 Introduction

The reason why people say things is the same as why they perform physical actions: because they want to achieve some goal by doing it. This is most obvious when the communicative action is an instruction which asks an interlocutor to perform a certain physical action; but it is still true for utterances of declarative sentences, which are intended to change the hearer's mental state in some way. The goals which an utterance achieves, or is meant to achieve, are called *perlocutionary effects* by Austin (1962).

However, relatively little work has been done on precise formal and computational models of perlocutionary effects, and in particular on the goal-directed use of communicative actions for their perlocutionary impact. Mainstream approaches such as Perrault and Allen (1980), which rely on modeling complex inferences in the hearer's mind and use non-standard planning formalisms, have never been demonstrated to be computationally efficient enough for practical use. On the other hand, issues of grounding (Clark, 1996) are highly relevant for the problem of modeling perlocutionary effects: If an utterance has not been understood, it cannot be expected to have its intended effect.

In this paper, we propose a new, general model of perlocutionary effects based on AI planning. In this model, the speaker computes a plan of communicative actions, each of which may have perlocutionary effects, under the assumption that all intended perlocutionary effects come true. That is, we model the effect of uttering "please open the window" as changing the world state such that the window becomes open. Because communicative actions can fail to have the intended effects (perhaps the hearer misunderstood, or is uncooperative), the speaker then observes the hearer's behavior to monitor whether the communicative plan has the intended effects. If the speaker notices that something goes wrong, they can react by diagnosing and repairing the problem.

This model makes it possible to deliberately compute a sequence of communicative actions that is fit to achieve a certain perlocutionary effect. Because the assumption that perlocutionary effects come true makes the planning easier, we can compute communicative plans efficiently; furthermore, our model can subsume communicative and physical actions within the same framework quite naturally. By scaling the execution monitoring module, we can trade off the precision with which the hearer's state is modeled against the inefficiency and model complexity this involves, according to the needs of the application. We show how a number of existing approaches to speech act planning and grounding can be reconstructed in this way, and illustrate the use of our model for the situated real-time generation of instructions in a small but fully implemented example.

Plan of the paper. We introduce our model in Section 2, connect it to the earlier literature in Section 3, and show its application to instruction generation in Section 4. Section 5 concludes.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

2 A new model of speech acts

We start by describing our model of speech acts, which combines communicative action planning with monitoring of the hearer's actions.

2.1 Communicative planning

The fundamental idea of our approach is to model a communicative act simply as an action in some planning problem. We take a communicative action to be some act of uttering a string of words. Ultimately, an agent performs such actions in order to achieve a (perlocutionary) goal which is external to language. This could be a physical goal (the light is now on), a goal regarding the mental state of the hearer (the hearer now believes that I have a cat), or something else. In this sense, communicative actions are exactly the same as physical actions: activities that are performed because they seem suitable for reaching a goal.

We model perlocutionary effects as effects of communicative actions in a planning problem. While we use classical planning throughout this paper, the basic idea applies to more expressive formalisms as well. A planning problem consists of a set of planning operators with preconditions and effects; an instance of an operator can be applied in a given planning state if its preconditions are satisfied, and then updates the state according to its effects. *Planning* (see e.g. Nau et al. (2004)) is the problem of finding a sequence of actions (i.e. operator instances) which transforms a given initial state into one that satisfies a given goal.

Consider the following example to illustrate this. An agent A is in a room with a light l_1 and two buttons b_1 and b_2; b_1 will turn on l_1, while b_2 is a dummy, and pressing it has no effect. We can encode this by taking an initial state which contains the atoms $\mathsf{agent}(A)$, $\mathsf{ltswitch}(b_1, l_1)$ (i.e. b_1 is the light switch for l_1), and $\mathsf{state}(l_1, \mathsf{off})$ (i.e. the light is off). Let's also include in the initial state that A is at location p_1 and b_1 is at a (different) location p_2, via atoms $\mathsf{at}(A, p_1)$, $\mathsf{at}(b_1, p_2)$ and $\mathsf{near}(p_1, p_2)$. Finally, let's assume that A wants l_1 turned on. We express this desire by means of a goal $\mathsf{state}(l_1, \mathsf{on})$. Then one valid plan will be to execute instances of the operators in Fig. 1, which encode physical actions performed by the agent. Specifically, **moveto**(A, p_1, p_2) and then **press**(b_1, A, p_2, l_1) will achieve the goal: The first action moves A to p_2, establishing the preconditions for the second action and turning on the light.

moveto(x, y_1, y_2):
Precond: $\mathsf{agent}(x), \mathsf{at}(x, y_1), \mathsf{near}(y_1, y_2)$
Effect: $\neg\mathsf{at}(x, y_1), \mathsf{at}(x, y_2)$

press(w, x, y, z):
Precond: $\mathsf{agent}(x), \mathsf{ltswitch}(w, z), \mathsf{at}(x, y), \mathsf{at}(w, y), \mathsf{state}(z, \mathsf{off})$
Effect: $\neg\mathsf{state}(z, \mathsf{off}), \mathsf{state}(z, \mathsf{on})$

Figure 1: Physical actions for turning on the light.

"press"(w, z):
Precond: $\mathsf{ltswitch}(w, z), \mathsf{state}(z, \mathsf{off})$
Effect: $\neg\mathsf{state}(z, \mathsf{off}), \mathsf{state}(z, \mathsf{on}),$
$\forall w'.w' \neq w \rightarrow \mathsf{distractor}(w')$
"the light switch"(w):
Precond: $\exists z.\mathsf{ltswitch}(w, z)$
Effect: $\forall w'.(\neg\exists z.\mathsf{ltswitch}(w', z) \rightarrow \neg\mathsf{distractor}(w'))$

Figure 2: Communicative actions for turning on the light.

If there is a second agent B in the room, then A can alternatively achieve the goal of switching l_1 on by asking B to do it, using communicative actions along the lines of those in Fig. 2. Here we add a further formula $\forall x.\neg\mathsf{distractor}(x)$ to the goal in order to require that the hearer can resolve all referring expressions uniquely. A valid plan is **"press"**(b_1, l_1) and then **"the light switch"**(b_1); this corresponds to uttering the sentence "press the light switch". (We write the names of communicative actions in quotes in order to distinguish them from physical actions.) The first action already achieves A's goal, $\mathsf{state}(l_1, \mathsf{on})$, but also introduces the atom $\mathsf{distractor}(b_2)$ into the planning state, indicating that the hearer won't be able to tell which button to press after hearing only "press ...". Since b_1 is the only light switch, this atom is easily removed by the action **"the light switch"**, which brings us into a goal state.

These two plans involve completely different kinds of actions: One uses physical actions performed by A, the other communicative actions performed by A intended to make B perform appropriate physical actions. Nevertheless, both plans are equally capable of achieving A's goal. We claim that communication is generally a goal-directed activity of this kind, and can be usefully modeled in terms of planning.

2.2 Plan execution monitoring

One crucial feature of this model is that the **"press"** operator, which encodes the action of uttering the word "press", has the effect that the light

is on. At first, this seems surprising, as if simply saying "press ..." could magically operate the light switch. This effect can be understood in the following way. Assume that the hearer of an utterance containing "press ..." which is complete in the sense that it is grammatically correct and all referring expressions can be resolved uniquely understands this utterance. Assume also that the hearer is cooperative and follows our request, and that they manage to achieve the goal we have set for them. Then the communicative plan underlying the utterance will indeed have the effect of switching on the light, through the physical actions of our cooperative hearer.

Our communicative planning operators directly contain the perlocutionary effects that the utterance will have if everything goes as the speaker intended. This makes it possible for a perlocutionary effect of one action in the plan to establish the precondition of another, and thus to form communicative plans that are longer than a single utterance; we will present an example where this is crucial in Section 4. But of course, we must account for the possibility that the hearer misunderstood the utterance, or is unwilling or unable to respond in the way the speaker intended; that is, that an action may not have the intended effect.

Here, too, communicative planning is no different from ordinary planning of physical actions. It is reasonable to assume for planning purposes that the operators in the physical plan of Subsection 2.1 have the intended effects, but the plan may fail if A is not able to reach the light switch, or if she made wrong assumptions about the world state, perhaps because the power was down. Inferring whether a plan is being carried out successfully is a common problem in planning for robots, and is called *plan execution monitoring* (Washington et al., 2000; Kvarnström et al., 2008) in that context. Although there is no commonly accepted domain-independent approach, domain-dependent methods typically involve observing the effects of an agent's actions as they are being carried out, and inferring the world state from these observations. Because there is usually some uncertainty about the true world state, which tends not to be directly observable, this can be a hard problem.

A speaker who detects a problem with the execution of their communicative plan has the opportunity to diagnose and repair it. Imagine that after hearing the utterance "press the light switch" in the earlier example, the hearer moves to a point where they can see both b_1 and b_2, and then hesitates. In this case, a hesitation of sufficient duration is evidence that the hearer may not execute the instruction, i.e. that the plan execution didn't have the intended perlocutionary effect. The speaker can now analyze what went wrong, and in the example might conclude that the hearer didn't know that b_2 isn't a light switch. This particular problem could be repaired by supplying more information to help the hearer remove distractors, e.g. by uttering "it's the left one". Deciding when and how to repair is an interesting avenue for future research.

2.3 A scalable model

Putting these modules together, we arrive at a novel model of perlocutionary acts: The speaker computes a plan of communicative actions that is designed to reach a certain goal; executes this plan by performing an utterance; and then observes the hearer's actions to monitor whether the intended perlocutionary effects of the plan are coming to pass. If not, the speaker repairs the plan.

By making optimistic assumptions about the success of perlocutionary effects, this model can get away with planning formalisms that are much simpler than one might expect; in the example, we use ordinary classical planning and move all reasoning about the hearer into the execution monitor. Among other things, this allows us to use fast off-the-shelf planners for the communicative planning itself. As we will see below, even relatively complex systems can be captured by making the execution monitor smart, and even shallow execution monitors can already support useful performances in implemented systems.

2.4 Limitations and extensions

The model proposed above is simplified in a number of ways. First, we have dramatically simplified the planning operators in Fig. 2 for easier presentation. At least, they should distinguish between the knowledge states of A and B and perhaps their common ground; for instance, in the effects of **"the light switch"**, only objects of which the *hearer* knows that they are not light switches should be excluded from the set of distractors. Koller and Stone (2007) show how to extend a planning-based model to make such a distinction.

Although we have only discussed instruction-giving dialogues above, we claim that the model is not limited to such dialogues. On the one hand,

declarative utterances affect the hearer through their perlocutionary effects just like imperative utterances do: They alter the hearer's mental state, e.g. by making a certain referent salient, or introducing a new belief. The role of the truth conditions of a declarative sentence is then to specify what perlocutionary effect on a hearer's belief state an utterance of this sentence can bring about.

On the other hand, we believe that other types of dialogue are just as goal-directed as instruction-giving dialogues are. In an argumentative dialogue, for instance, each participant pursues a goal of convincing their partner of something, and chooses communicative actions that are designed to bring this goal about. The role of execution monitoring in this context is to keep track of the partner's mental state and revise the communicative plan as needed. Because both partners' goals may conflict, this is reminiscent of a game-theoretic view of dialogue. It is conceivable that certain types of dialogue are best modeled with more powerful planning formalisms (e.g., information-seeking dialogues by planning with sensing), but all of our points are applicable to such settings as well. In particular, even in more complex settings the planning problem might be simplified by moving some of the workload into the execution monitor.

Finally, we have focused on plans which only contain *either* physical (Fig. 1) *or* communicative (Fig. 2) actions. However, since we have blended the physical and communicative contributions of those acts together (as e.g. with the communicative act **"press"** of Fig. 2), we can also compute plans which combine both types of action. This would allow us, for instance, to interleave communicative actions with gestures. In this way, our proposal could pave the way for a future unified theory which integrates the various kinds of communicative and physical actions.

3 Speech act planning and grounding

We will now discuss how our model relates to earlier models of speech act planning and grounding.

The most obvious point of comparison for our model is the family of speech act planning approaches around Perrault and Allen (1980) (henceforth, P&A), which are characterized by modeling speech act planning as a complex planning problem involving reasoning about the beliefs, desires, and intentions (BDI) of the interlocutors. P&A model the perlocutionary effect of a speech act REQUEST(P) as causing the hearer to intend to do P. However, this effect has to be justified during the planning process by inferences about the hearer's mental state, in which the hearer first recognizes the speaker's intention to request P and then accepts P as their own intention. Although we agree with the fundamental perspective, we find this approach problematic in two respects. First, the perlocutionary effect of REQUEST is modeled as limited to the hearer: it is not that P happens, but only that the hearer wants P to happen. This makes it impossible to compute communicative plans in which a subsequent utterance relies on the intended perlocutionary effects of an earlier utterance, as e.g. in Section 4 below. Second, even if we limit ourselves to the effect on the hearer's mental state, the formal approach to planning that P&A take is so complex that computing plans of nontrivial length is infeasible.

Our model solves the first problem by modeling the intended physical and mental effects directly as effects of the operator, and it solves the second problem by using simple planning formalisms. Compared to P&A, it takes a more optimistic stance in that the default assumption is that perlocutionary effects happen as intended. Any reasoning about the hearer's BDI state can happen during the execution monitoring phase, in which we can compute the expected step-by-step effects of the utterance on the hearer's state (intention recognition, goal uptake, etc.) as P&A do, and then try to establish through observations whether one of them fails to come true. This allows us to compute very simple plans without sacrificing linguistic correctness. We believe that similar comments hold for other recent planning-based models, such as (Steedman and Petrick, 2007; Brenner and Kruijff-Korbayová, 2008; Benotti, 2009).

We share our focus on modeling uncertainty about the effects of communicative actions with recent approaches to modeling dialogue in terms of POMDPs (Frampton and Lemon, 2009; Thomson and Young, 2009). POMDPs are a type of probabilistic planning problem in which the effects of actions only come true with certain probabilities, and in which the true current world state is uncertain and only accessible indirectly through observations; the analogue of a plan is a *policy*, which specifies what action to take given certain observations. This makes POMDPs a very pow-

erful and explicit tool for modeling uncertainty about effects, which is however limited to very simple reasoning about observations. Although our planning model is not probabilistic, we believe that the two approaches may be more compatible than they seem: Many recent approaches to probabilistic planning (including the RFF system, which won the most recent probabilistic planning competition for MDPs (Teichteil-Koenigsbuch et al., 2008)) transform the probabilistic planning problem into a deterministic planning problem in which probable effects are assumed to come true, monitor the execution of the plan, and replan if the original plan fails. This is a connection that we would like to explore further in future work.

Grounding – the process by which interlocutors arrive at the belief that they mutually understood each other – falls out naturally as a special case of our model. A speaker will continue to monitor the hearer's behavior until they are sufficiently convinced that their communicative action was successful. This typically presupposes that the speaker believes that the hearer understood them; traditional classes of devices for achieving grounding, such as backchannels and clarification requests, are among the observations considered in the monitoring. Conversely, the speaker can stop monitoring once they believe their perlocutionary goal has been achieved; that is, when their degree of belief in mutual understanding is "sufficient for current purposes" (Clark and Schaefer, 1989), i.e. the current perlocutionary goal. Our prediction and tracking of expected perlocutionary effects is reminiscent of the treatment of grounding in information state update models, in which utterances introduce *ungrounded discourse units* (Matheson et al., 2000) into the conversational record, which must be later grounded by the interlocutors. In our approach, the first step could be implemented by introducing the ungrounded unit as an effect and then verifying that grounding actually happened in the execution monitor.

In its reliance on planning, our approach is somewhat in contrast to Clark (1996), who fundamentally criticizes planning as an inappropriate model of communication because "people ... don't know in advance what they will actually do [because] they cannot get anything done without the others joining them, and they cannot know in advance what the others will do". We claim that this ignorance of speakers about what is going to happen need not keep them from forming a communicative plan and attempting a promising speech act; after all, if the hearer does unexpected things, the speaker will be able to recognize this and react appropriately. In our perspective, communication is not primarily a collaborative activity, but is driven by each individual agent's goals, except insofar as collaboration is necessary to achieve these goals (which it often is). This seems in line with recent psycholinguistic findings indicating that a speaker's willingness to select an utterance that is optimal for the partner is limited (Shintel and Keysar, 2009; Wardlow Lane and Ferreira, in press).

We deliberately keep details about the execution monitoring process open, thereby subsuming approaches where the speaker explicitly models the hearer's mental state (Poesio and Rieser, 2010), or only does this if necessary (Purver, 2006), or which emphasize inferring success from directly accessible observations (Skantze, 2007; Frampton and Lemon, 2009). In this sense, the model we propose is scalable to different modeling needs.

4 Communicative planning in practice

At this point, we have argued that very expressive execution monitors can in principle be used to reconstruct a number of approaches from the literature. We will now demonstrate that even a very *in*expressive execution monitor can be useful in a concrete application. The example on which we illustrate this is the SCRISP system (Garoufi and Koller, 2010), which extends the CRISP NLG system (Koller and Stone, 2007) to situated communication. CRISP, in turn, is a planning-based reimplementation of the SPUD system (Stone et al., 2003) for integrated NLG with tree-adjoining grammars (TAG, (Joshi and Schabes, 1997)).

SCRISP generates real-time navigation and action instructions in a virtual 3D environment. The overall scenario is taken from the GIVE-1 Challenge (Byron et al., 2009): A human instruction follower (IF) must move around in a virtual world as in Fig. 3, which is presented to them in 3D as in Fig. 4. The NLG system receives as input a *domain plan*, which specifies the (simulated) physical actions in the world that the IF should execute, and must compute appropriate *communicative plans* to make the IF execute those physical actions. Thus the perlocutionary effects that the NLG system needs to achieve are individual ac-

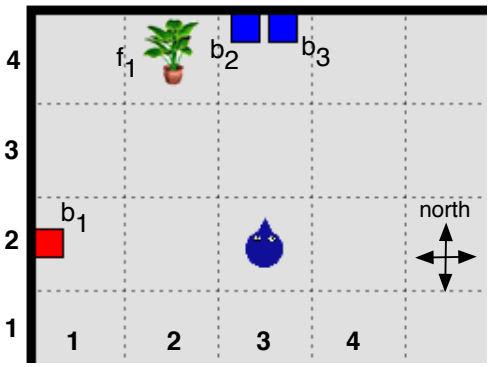


Figure 3: An example map for instruction giving.

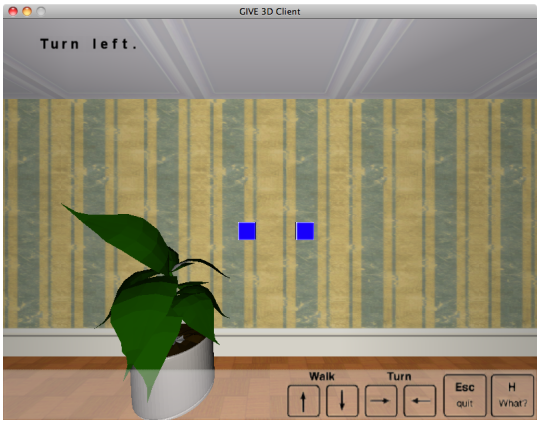


Figure 4: The IF's view of the scene in Fig. 3, as rendered by the GIVE client.

tions in the domain plan. In the example of Fig. 3, one action of the domain plan is $\mathsf{push}(b_1)$, i.e. the act of the IF pressing b_1. A sequence of communicative actions that has a good chance of achieving this is to utter "turn left and push the button".

SCRISP can compute such a communicative act sequence using planning, and can monitor the execution of this communicative plan. The average time it takes to compute and present a plan, on an original GIVE-1 evaluation world (as represented by a knowledge base of approx. 1500 facts and a grammar of approx. 30 lexicon entries), is about one second on a 3 GHz CPU. The plans are computed using the FF planner (Hoffmann and Nebel, 2001; Koller and Hoffmann, 2010). This shows that the approach to speech act planning we propose here can achieve real-time performance.

4.1 Situated CRISP

SCRISP assumes a TAG lexicon in which each elementary tree has been equipped with pragmatic preconditions and effects next to its syntactic and semantic ones (see Fig. 5). Each of these is a set of atoms over constants, free variables, and argument names such as obj, which encode the individuals in the domain to which the nodes of the elementary tree refer. These atoms determine the preconditions and effects of communicative actions.

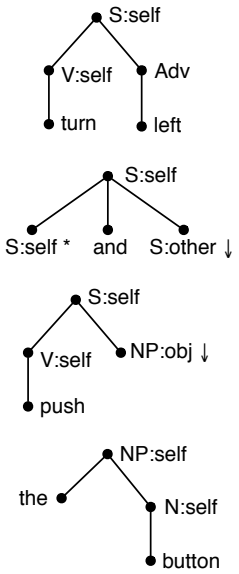
For instance, the lexicon in Fig. 5 specifies that uttering "push X" has the perlocutionary effect that the IF presses X. It also says that we may only felicitously say "push X" if X is visible from the IF's current position and orientation. The position and orientation of the IF, and with them the currently visible objects, can be modified by first uttering "turn left". The two utterances can be chained together by sentence coordination ("and"). Finally, introducing the noun phrase "the button" as the object of "push" makes the sentence grammatically complete.

In order to generate such a sequence, SCRISP converts the lexicon and the perlocutionary goal that is to be achieved into a planning problem. It then runs an off-the-shelf planner to compute a plan, and decodes it into sentences that can be presented to the hearer. The operators of the planning problem for the example lexicon of Fig. 5 are shown in simplified form in Fig. 6, which can be seen as an extended and more explicit version of those in Fig. 2. We do not have the space here to explain the operators in full detail (see Garoufi and Koller (2010)). However, notice that they have both grammar-internal preconditions and effects (e.g., subst specifies open substitution nodes, ref connects syntax nodes to the semantic individuals to which they refer, and canadjoin indicates the possibility of an auxiliary tree adjoining the node) and perlocutionary ones. In particular, the **"push"** action has a perlocutionary effect $\mathsf{push}(x_1)$.

4.2 Planning and monitoring perlocutionary acts with SCRISP

Now let's see how SCRISP generates instructions that can achieve this perlocutionary effect.

First, we encode the state of the world as depicted in Fig. 3 in an initial state which contains, among others, the atoms $\mathsf{player\text{–}pos}(\mathsf{pos}_{3,2})$, $\mathsf{player\text{–}ori}(\mathsf{north})$, $\mathsf{next\text{–}ori\text{–}left}(\mathsf{north}, \mathsf{west})$, $\mathsf{visible}(\mathsf{pos}_{3,2}, \mathsf{west}, b_1)$, etc. As the goal for the planning problem, we take our perlocutionary goal, $\mathsf{push}(b_1)$, along with linguistic constraints including $\forall A \forall u. \neg \mathsf{subst}(A, u)$ (encoding syntactic completeness) and $\forall u \forall x. \neg \mathsf{distractor}(u, x)$



cond: player–ori(o_1),
next–ori–left(o_1, o_2)
eff: ¬player–ori(o_1),
player–ori(o_2),
turnleft

cond: player–pos(p),
player–ori(o),
visible(p, o, obj)
eff: push(obj)

Figure 5: A simplified example of a SCRISP lexicon, focusing on pragmatic conditions and effects.

"turnleft"(u, x, o_1, o_2):
Precond: subst(S, u), ref(u, x), player–ori(o_1), next–ori–left(o_1, o_2), ...
Effect: ¬subst(S, u), ¬player–ori(o_1), player–ori(o_2), turnleft, ...

"and"(u, u_1, u_n, e_1, e_2):
Precond: canadjoin(S, u), ref(u, e_1), nextref(e_1, e_2), ...
Effect: subst(S, u_1), ref(u_1, e_2), ...

"push"$(u, u_1, u_n, x, x_1, p, o)$:
Precond: subst(S, u), ref(u, x), player–pos(p), player–ori(o), visible(p, o, x_1), ...
Effect: ¬subst(S, u), subst(NP, u_1), ref(u_1, x_1), $\forall y.(y \neq x_1 \wedge$ visible$(p, o, y) \rightarrow$ distractor$(u_1, y))$, push(x_1), canadjoin(S, u), ...

"the–button"(u, x):
Precond: subst(NP, u), ref(u, x), button(x)
Effect: $\forall y.(\neg$button$(y) \rightarrow \neg$distractor$(u, y))$, ¬subst(NP, u), ...

Figure 6: Simplified SCRISP planning operators for the lexicon in Fig. 5.

(encoding unique reference).

The planner can then apply the action **"turnleft"**(root, e, north, west) to the initial state. This action makes player–ori(west) true and subst(root, e) false. The new state does not contain any subst atoms, but we can continue the sentence by adjoining "and", i.e. by applying the action **"and"**(root, $\mathsf{n}_1, \mathsf{n}_2, e, e_1$). This produces a new atom subst(S, e_1), which satisfies one precondition of **"push"**($\mathsf{n}_1, \mathsf{n}_2, \mathsf{n}_3, e_1, b_1, \mathsf{pos}_{3,2}$, west). Because **"turnleft"** changed the player orientation, the visible precondition of **"push"** is now satisfied too (unlike in the initial state). Applying the action **"push"** introduces the need to substitute a noun phrase for the object, which we can eliminate with an application of **"the button"**(n_2, b_1). We are thus brought into a goal state, in which the planner terminates.

The final state of this four-step plan contains the atom push(b_1), indicating that if everything goes as intended, the hearer will push b_1 upon hearing the instructions. Crucially, we were only able to compute the plan because we make the optimistic assumption that communicative acts have the intended perlocutionary effects: For instance, it is only because we assumed that uttering "turn left" would make the IF change their orientation in space that this action was able to establish the precondition of the "push" action. A REQUEST operator as in P&A, which only makes the IF *want* to turn left, would not have achieved the same.

Having uttered these instructions, SCRISP observes the IF's behavior in the virtual world to determine whether the intended effects actually come to pass. We achieve this through three simple submodules of the execution monitor. The "inactivity" submodule tracks whether the user has not moved or acted in a certain period of time, and resends the previous instruction when this happens. Fresh instructions from the user's current location are issued when the "distance" submodule observes that the user is moving away from the location at which they need to perform the next physical action. Finally, the "danger" submodule monitors whether the user comes close to a trap in the world, and warns them away from it. It is clear that these three modules only allow the system a very limited view into the hearer's mental state. Nevertheless, they already allow SCRISP to achieve competitive performance in the GIVE-1 task (Garoufi and Koller, 2010).

5 Conclusion

In this paper, we have proposed to model communicative actions as planning operators and their intended perlocutionary effects as effects of these operators; we further proposed that after uttering something, the speaker then observes the hearer's behavior to infer whether the utterance had the intended effect. This moves the main complexity of communication into the plan execution monitoring module, where it can be handled with as much effort as the application requires, while keeping the

planning itself simple and fast.

We see the most interesting task for the future in working out some of the connections we sketched here in more detail, particularly a full model of the P&A approach and an extension to declarative utterances. In addition, it would be exciting to see how the notions of utility and uncertainty from POMDPs can be generalized in cases where a belief about the state is formed from observations using more than just a probability distribution, while retaining efficient planning.

Acknowledgments. We are indebted to Luciana Benotti, Oliver Lemon, Ron Petrick, Matt Purver, and David Schlangen for challenging and constructive discussions on the topic of this paper. We thank our reviewers for their insightful comments.

References

J. Austin. 1962. *How to do things with words*. Oxford University Press.

L. Benotti. 2009. Clarification potential of instructions. In *Proc. 10th SIGDIAL*.

M. Brenner and I. Kruijff-Korbayová. 2008. A continual multiagent planning approach to situated dialogue. In *Proc. LonDial*.

D. Byron, A. Koller, K. Striegnitz, J. Cassell, R. Dale, J. Moore, and J. Oberlander. 2009. Report on the First NLG Challenge on Generating Instructions in Virtual Environments. In *Proc. 12th ENLG*.

H. Clark and E. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13(2):259–294.

H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, England.

M. Frampton and O. Lemon. 2009. Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, 24(4):375–408.

K. Garoufi and A. Koller. 2010. Automated planning for situated natural language generation. In *Proc. 48th ACL*.

J. Hoffmann and B. Nebel. 2001. The FF planning system: Fast plan generation through heuristic search. *J. Artificial Intelligence Research*, 14:253–302.

A. Joshi and Y. Schabes. 1997. Tree-Adjoining Grammars. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3, pages 69–123. Springer.

A. Koller and J. Hoffmann. 2010. Waking up a sleeping rabbit: On natural-language sentence generation with FF. In *Proc. 20th ICAPS*.

A. Koller and M. Stone. 2007. Sentence generation as planning. In *Proc. 45th ACL*.

J. Kvarnström, F. Heintz, and P. Doherty. 2008. A temporal logic-based planning and execution monitoring system. In *Proc. 18th ICAPS*.

C. Matheson, M. Poesio, and D. Traum. 2000. Modelling grounding and discourse obligations using update rules. In *Proc. 6th ANLP*.

D. Nau, M. Ghallab, and P. Traverso. 2004. *Automated Planning: Theory and Practice*. Morgan Kaufmann.

C. R. Perrault and J. Allen. 1980. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3–4):167–182.

M. Poesio and H. Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1(1).

M. Purver. 2006. CLARIE: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288.

H. Shintel and B. Keysar. 2009. Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, 1(2):260–273.

G. Skantze. 2007. *Error Handling in Spoken Dialogue Systems: Managing Uncertainty, Grounding and Miscommunication*. Ph.D. thesis, KTH, Stockholm, Sweden.

M. Steedman and R. Petrick. 2007. Planning dialog actions. In *Proc. 8th SIGdial*.

M. Stone, C. Doran, B. Webber, T. Bleam, and M. Palmer. 2003. Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, 19(4):311–381.

F. Teichteil-Koenigsbuch, G. Infantes, and U. Kuter. 2008. RFF: A robust, FF-based MDP planning algorithm for generating policies with low probability of failure. In *Proc. 6th IPC at ICAPS*.

B. Thomson and S. Young. 2009. Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems. *Computer Speech and Language*.

L. Wardlow Lane and V. S. Ferreira. in press. Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

R. Washington, K. Golden, and J. Bresina. 2000. Plan execution, monitoring, and adaptation for planetary rovers. *Electronic Transactions on Artificial Intelligence*, 4(A):3–21.

Establishing coherence in dialogue: sequentiality, intentions and negotiation

Gregory Mills
Department of Psychology
Stanford University
gjmills@stanford.edu

Eleni Gregoromichelaki
Department of Philosophy
King's College London
eleni.gregor@kcl.ac.uk

Abstract

Communication in everyday conversation requires coordination both of content and of process. While the former has been studied extensively, there has been a paucity of studies on the latter. This paper addresses the question of how sequences of talk are established and sustained. We present evidence from a series of maze-game experiments that raise fundamental questions concerning the basic coordination mechanisms that are involved in this process.

1 Introduction

Dialogue, the primary site of language use, is fundamentally part, and constitutive of everyday social settings. Interaction situated in these settings is underpinned by constraints on expected and permissible contributions which provide the foundation of interlocutors' meaningful exchange and effective coordination. An important component of this foundation consists of the *sequential organisation* of interleaved utterances and actions over multiple turns by different speakers.

Sequential organization of interaction is perhaps most apparent in formal settings that require strict adherence to a protocol of prescribed steps, e.g. religious ceremonies or court proceedings (Atkinson and Drew 1979). However, dialogue research has revealed that sequential organisation is a principal means of achieving coordination in everyday conversation too, e.g. making requests, or entering and exiting phone conversations, (Schegloff 1979).

A common thread running through these studies is the recognition that the unit of analysis required for examining these phenomena must be sensitive to the relationship *between* turns (Drew 1997). In Conversation Analysis (CA), the role of turn-taking is emphasized as a locally managed interactional system employed by interlocutors to coordinate their conversation. CA analyses show that the sequential organisation of turn-taking is the basis for establishing two types o*f coherence*: (a) it manages the orderly distribution of contributions, but also (b) it coordinates content: the sequential relationship between turns provides important constraints on utterance interpretation. In Clark's *grounding model* (1996), this approach is extended: a key concept is the *joint proje*ct, which embeds verbal interaction within the more general concept of *joint activities*, thus deriving the constraints on the sequential organisation of interlocutors' contributions from the analytic structure of the common goals that are pursued.

However, there are two main issues that arise in these approaches: (a) Despite these studies' close analysis of how particular dialogue trajectories unfold during interaction, there has been a paucity of studies that directly address how *ad-hoc sequential organization* emerges, becomes established and is maintained (b) Existing models of dialogue provide conflicting accounts of which mechanisms are employed by interlocutors to achieve the development of sequential organisation.

1.1 Development of sequentiality

In empirical investigations within CA, coordination is established by interlocutors' turn-by-turn displays of their construals of each other's utterances. The structure underpinning such displays is the conversation analytic notion of *adjacency pairs*. CA investigations show how interlocutors' utterances demonstrate acute sensitivity to the unfolding sequential nature of the dialogue (Schegloff 1972; Drew 1997). Here the emphasis is placed on the coherence relationships among turns which is based on *conditional relevances* set up among contributions: production of the first part of an adjacency pair creates an expectation that the second half will occur. Any response will then be interpreted as pertaining to the second half through a system of inferential significances set up by this expectation. This locally managed turn-coherence system results in global coherence through the hierarchical inter-

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

leaving of embedded sequences that take care of local problems and disturbances by means of, e.g., clarification and elaboration (Levinson 1983; Clark 1996).

However, as CA's main objects of study have been single stretches of talk, such investigations seem to have neglected the interleaving of talk with action and the conditional relevances established in the domain of joint activities in general (Clark 1996). Moreover, as a methodological premise, CA analyses concern "naturally occurring" dialogue as their primary data, rejecting experimental manipulation to probe specific predictions (Schegloff 1992). As a result, CA research has typically treated adjacency pairs as static objects, already shared by interlocutors, and, hence, has not been led to any systematic investigation of how they might develop during conversation.

1.2 Coordination mechanisms

Existing accounts of dialogue differ in their emphasis on which mechanisms are involved in coordination and how their systematic deployment affects the course of dialogue. Formal approaches to dialogue that operate under standard Gricean assumptions (e.g. Grosz and Sidner 1986; Cohen et al 1990; Poesio and Traum 1997; Poesio and Rieser 2010) see the formulation of determinate intentions/plans and their full recognition as the main causal mechanism underlying dialogue comprehension and production. Performance in joint tasks then relies on the coordination of (joint) intentions/plans through negotiation and grounding that establish mutual beliefs and common ground. Similar prominence to explicit negotiation is also given by the grounding model. Here, the role of *coordination devices,* the basis for interlocutors' mutual expectations of each other's individual actions, (Clark 1996; Shelling 1960) is emphasised. Alterman (2001) also argues for the value of explicit negotiation in the achievement of coordination.

In contrast, Garrod and Anderson (1987) observe that explicit negotiation is neither a preferential nor an effective means of coordination. If it occurs at all, it usually happens after participants have already developed some familiarity with the task; even when a particular approach to the task is explicitly negotiated and agreed by the participants they do not seem to persevere with it for long. The Interactive Alignment model developed by Pickering and Garrod (2004) emphasizes the importance of *tacit co-ordination* and *implicit common ground* achieved via the psychological mechanism of priming. However, this model does not appear ideally suited to account for the development of sequential organisation: the establishment of *routines*[1] and the significance of repair as externalised inference are noted by Pickering and Garrod but it is unclear how these mechanisms could be extended straightforwardly to capture sequential structures that span multiple turns and organise the performance of the whole joint activity.[2]

To address these issues concerning the mechanisms that are implicated in the emergence of sequential structure and the significance of the coordination devices interlocutors employ, we draw on the results of a series of maze-game experiments. In all of these experiments, participants collaboratively develop sequences of steps to solve the mazes, thereby providing a helpful means for analysing how sequential structure becomes established.

2 Methods

The experiments employ a modified version of the "Maze Game" devised by Garrod and Anderson (1987). This task creates a recurrent need for pairs of participants to co-ordinate on procedures for solving the mazes.

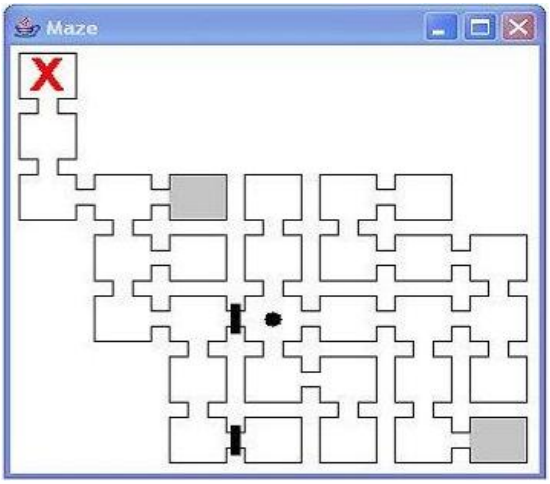


Fig 1: Example maze configuration. The black circle shows the player's current position, the cross represents the goal point that the player must reach, solid bars the gates and grey squares the switch points.

[1]However, see Poesio and Rieser (2010) for an attempt to reformulate their intentional account in terms of routinisation.

[2]Garrod and Anderson's (1987) *principle of input/output co-ordination*, even though intended as a simple heuristic to replace explicit negotiation, also does not seem to account for the development of sequential organisation. This is because its primary focus is on how single conventions, typically a referring expression or descriptive scheme, become established through successive use. It is unclear how this could be extended to capture sequential structures that span multiple turns.

2.1 Materials

2.1.1 Maze Game

The maze application displays a simple maze consisting of a configuration of nodes that are connected by paths to form grid-like mazes (see Fig 1). The mazes are based on a 7x7 grid and are selected to provide both grid-like and asymmetric instances. Participants can move their location markers from one node to another via the paths. Each move is recorded and timestamped by the server. The game requires both subjects to move their location markers from a starting location to a goal. Although the maze topology is the same for both subjects, each subject has a different starting location and goal, neither of which are visible to the other subject. They are also not able to see each other's location markers. Movement through the maze is impeded by gates that block some of the paths between nodes. These gates can be opened by the use of switches (grey coloured nodes). The locations of switches and gates are different on each maze and are not visible to the other participant. Whenever a participant moves to a node that is marked as a switch on the other's screen, all of the other participant's gates open. All the gates subsequently close when they move off the switch.

This constraint forces participants to collaborate: in order for participant A to open their gates, A has to guide B onto a node that corresponds to a switch that is only visible on A's screen. Solving the mazes (i.e. when both participants are on their respective goal positions) requires participants to develop procedures for requesting, describing and traversing switches, gates and goals.

2.1.2 Chat tool interface

Participants communicate with each other via the use of a novel text-based experimental chat tool (Healey and Mills, 2006). All turns generated by the participants pass through a server that allows for the introduction of artificial turns that appear, to participants, to originate from each other.

2.2 Participants and Design

56 pairs of native English speakers participants were recruited from undergraduate students and were assigned to one of three conditions (*baseline*, *clarification requests*, *reduced sequentiality*). In all conditions, dyads played 12 randomly generated mazes.

Baseline: 11 dyads served as control group. No experimental interventions were performed.

Dual window (reduced sequentiality): 18 dyads were assigned to this condition. Participants used a variation of the chat-tool design similar to Anderson et al. (2000). This version directly interferes with the sequential coherence of the unfolding dialogue by separating the chat-text into two separate windows that only display chat-text from a single participant, thereby prohibiting any interleaving between turns (see Fig 2, below).

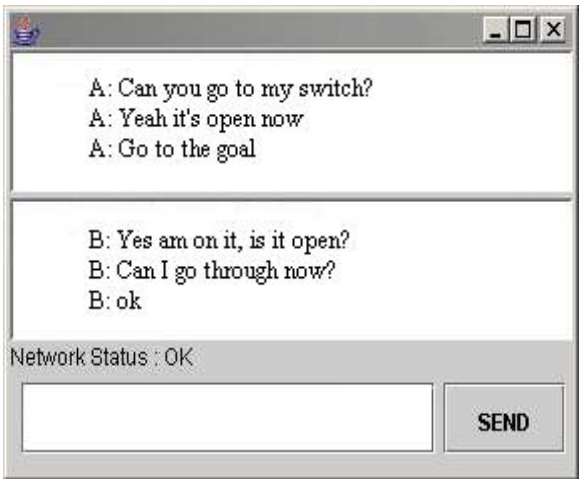


Figure 2: Dual window chat tool. Each half only displays text from one participant.

Artificial clarification requests: 26 dyads were assigned to this condition. Every (+/-) 35 turns, the server randomly generated artificial clarification questions (probe CRs) that appeared, to participants, to originate from each other. Overall, 219 clarification requests were generated. Participants' responses to probe CRs were recorded for analysis and were not relayed to the other participant. After receiving a response to the CR, the server sends an artificial acknowledgement turn ("ok", "ok right") to the recipient and resumes relaying subsequent turns as normal.

The clarification questions were of two types: Reprise Fragments ('**Frags**') that query a constituent of the target turn by echoing it and '**Whats**' (e.g., "what?", "sorry?", "Ehh?", "uhh?") that query the turn as a whole. The excerpt below illustrates a typical fragment clarification sequence:

A:	I'm at the top	*Target turn*
B:	top?	*Artificial turn by server*
A:	yes	*Response by A*
B:	thanks	*Artificial ack. by server*

3 Results

3.1 Sequentiality

To test whether reduced sequential coherence makes the emergence of procedural co-ordination more problematic, all turns in the baseline and reduced sequentiality conditions were classified with respect to procedural '**co-ordination points**' (Alterman and Garland 2001). These are defined as points in a joint activity when participants are interacting with each other in order to develop a common course of action. Co-ordination points here where coded if they mentioned "switch", "gate" or "goal". To provide a measure of the difficulty of establishing procedural co-ordination, the log files were used to calculate the typing speed (characters per second) and the number of edits (deletes per character) at these co-ordination points (See Fig 3 below).

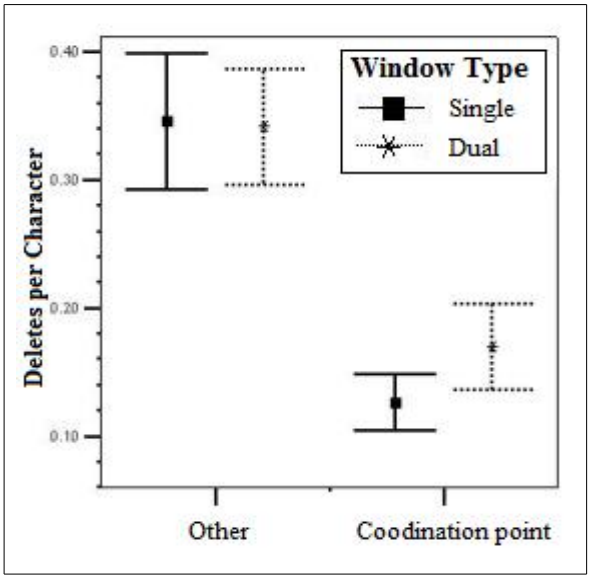


Figure 3: Turn formulation difficulty (Deletes per character) for turns that mention "switches", "gates", "goals" (Coordination point).

Focusing on co-ordination points, subjecting the number of edits to a one way ANOVA, with Window type (dual/single) as between-subjects factor yielded a significant main effect of decreased sequential coherence on number of edits ($F=4.69$, $p<0.05$) and also on turn formulation time ($F= 4.01$, $p < 0.05$). Interfering with sequential coherence increases turn formulation time (11.6 vs. 10.5 characters per second) and results in more self-editing (0.12 vs. 0.17 edits per character).

By contrast, for turns that did not explicitly mention "switch", "gate", "goal", no effect of reduced sequentiality was found on typing speed ($F=0.13$, $p=0.17$), or edits ($F=0.012$, $p=0.91$).

3.2 Intention recognition

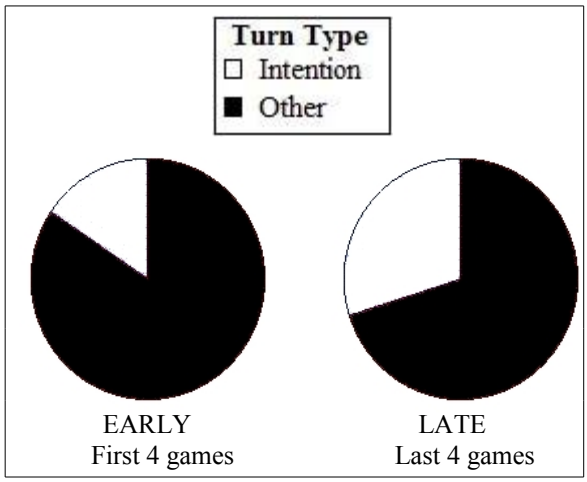


Figure 4. Proportion of CR responses that clarify intentions in first 4 games and last 4 games.

To test the necessity of plan/intention recognition for grounding utterances, participants' responses to clarification requests (CRs) were coded for whether they were construed as concerning the underlying intention of the queried turn, e.g., "you have to go there to open my switch" or "so that my gate opens". All things being equal, if plan/intention recognition is essential to establishing coordination, participants should clarify plans and intentions more frequently in the early stages of the game than in the late stages when the plans have become more established. Comparing participants' responses in the first 4 games with responses from the last 4 games yields the opposite: as task experience increases, participants' responses to clarification requests use significantly more explicit disambiguations concerning "intentions", (chi^2 (1, $N=219$) = 6.3, $p<0.01$) (see Fig 4 above).

3.3 Explicit negotiation

To examine the role played by explicit negotiation, the first mazes in all conditions were coded for attempts to explicitly negotiate the sequence, for example, "first you go through the gate, then I get on the switch and then you tell me when you're through". Examining these attempts yielded no dyad that was able to establish such a sequence through explicit negotiation alone.

4 Discussion

The results outlined above[3] present a problem for existing accounts of dialogue that rely on intention-recognition/planning: At the start of the experiment, participants must interact with each other on how to develop procedures for solving the mazes. However, examination of the data suggests that this interaction does not involve straightforwardly articulated negotiation in the form of plans and intentions (see also Garrod and Anderson 1987). Instead, participants seem to develop gradually and spontaneously a structured solution to the coordination problem the maze game presents, a solution, moreover, which is consistent across pairs of participants.

To illustrate how participants become progressively more coordinated without relying on explicit negotiation, we focus on how their exchanges become progressively structured. For reasons of space, we will examine in parallel what effects this structuring has on coordination as far as interpretation of turns is concerned.

4.1 Embedding moves in sequences

4.1.1 Effects on interpretation

In all three maze game conditions, as observed in previous maze game studies, interlocutors frequently take more than 120 turns to solve each maze (Mills, 2007). However, by the 12th maze, interlocutors develop *sequences* of highly formulaic, elliptical, fragmentary utterances, with very complex context-dependent interpretations (see e.g. fragments 1,2 and 4,5 in Excerpt 1 below). For example, consider a typical full exchange from one of the later games:

1)	**A:**	1,2 2,6 1,4
2)	**A:**	5,6
3)	**B:**	4,5 3,4 7,1
4)	**B:**	1,4
5)	**A:**	4,5
6)	**B:**	1,2
7)	**A:**	4,5

Excerpt 1: Highly elliptical dialogue from late stages of maze game

However, participants are able to unproblematically deal with these fragments, as in these later stages of the experiment, they have developed adequate procedural expertise which allows them to navigate the task with minimal conversational exchange. Of great interest at this stage is the absence of any explicit requests, confirmations or feedback.

In contrast, at a slightly earlier stage, the dialogues are less elliptical:

1)	**A:**	I have switches at 1,2 2,6 1,4
2)	**B:**	Where's your goal?
3)	**A:**	5,6
4)	**B:**	mine are at 4,5 3,4 7,1 can you get to any of them?
5)	**A:**	I can get to 4,5, can you get to any of mine?
6)	**B:**	I can get to 1,2
7)	**A:**	I'm on 4,5 you can go through now..
8)	**A:**	go to your goal
9)	**B:**	Done

Excerpt 2: Less elliptical dialogue from late stages of maze game

Here it is much more apparent what the sequential import of each turn is: in addition to referring to spatial locations, the interlocutors have developed a highly structured joint project for solving the mazes, consisting of the following sub-projects:

1. State (and request) location of switches
2. State (and request) location of goal
3. Signal accessibility
4. Move onto switch
5. Other moves through gate
6. Move onto goal

In the light of this structure we can now interpret Excerpt 1 as demonstrating how the sequence of steps has become sufficiently established (*routinised*) to obviate the need for any explicit indication of what each turn is "doing". Depending on an utterance's sequential location within the joint project, the Cartesian co-ordinate fragments in Excerpt 1 will be taken to refer to a switch, a goal, or to a participant's current location in the maze. Further, the utterance's sequential position also determines whether the mentioned co-ordinate is a request for the other participant to move to that location, a confirmation of having moved to a location, a statement where they are currently located, or where they are blocked (see Excerpt 3 below).

[3] Note that despite the difference in modality, results obtained using the chat-tool interface have been shown to replicate results from the original verbal task (see e.g. Healey & Mills 2006).

4.1.2 How structure emerges

The sequence in Excerpt 2 appears deceptively simple: it is a straightforward exchange consisting of spatial descriptions, requests, confirmations. Despite this apparent simplicity, the results show that the interlocutors are unable to arrive at this sequence through explicit negotiation.[4] Instead, it is arrived at via a series of longer, less elliptical, sequences of turns. Excerpt 3 below shows a typical immediately prior stage in the development of the joint project:

1)	**A:**	I have a switch at 1,2
2)	**B:**	I can't get there
3)	**A:**	how about 2,6?
4)	**A:**	If you go to any of them they will open all of my gates
5)	**B:**	ok, can get to 2,6
6)	**A:**	Go
7)	**B:**	Is your gate open?
8)	**A:**	yeah...thanks..where are your switches?
		(similar to sequence above, but for opening B's switches)
21)	**A:**	My goal is at 2,3
22)	**A:**	I can get to 4,5
23)	**B:**	ok. can you go there to open my switch?
24)	**A:**	yeah..am on it now..is your gate open?
25)	**B:**	yeah, are you through?
26)	**A:**	yeah
27)	**B:**	can you get to your goal?
28)	**A:**	yep..am on goal
29)	**B:**	I'm stuck on 2,3..go back...I need to get through a gate that is between me and my goal..
30)	**A:**	where do you need me to go?
31)	**B:**	can you get to 4,5 and then reach your goal?
32)	**A:**	yeah, shall I go there?
33)	**B:**	go
35)	**B:**	done

Excerpt 3. Less elliptical maze game dialogue.

Space considerations preclude showing the preceding exchanges, since for the first few mazes games frequently take over 150 turns. Nevertheless, global inspection of the data shows that there is no "shortcut" -interlocutors must pass through stages of progressively refining and shortening the sequence they develop.

Importantly, this phenomenon of "telescoping" of sequences of actions is orthogonal to the contraction of referring expressions on repeated use (Krauss and Weinheimer 1975; Brennan and Clark 1996). This certainly does occur, e.g. (1,2) is a contraction of "1 across 2 along", which in turn is a contraction of "1 across from the left and 2 down from the right".[5] However, in the excerpts shown above, the progressive contraction of the sequences (telescoping) is a global pattern of developing *coherence* that operates over whole spates of talk that amount to the resolution of a single maze. Further, the data strongly indicate that this contraction is not simply the verbal exchange becoming shorter and more structured. Instead, it can be seen that by the end of the round of games played, all turns have become moves in a game involving both utterances and actions. Each move projects a next move, and implicitly confirms completion of a prior move (i.e. *conditional relevances* have developed). This, as explained below, can be discerned by the participants' responses to the fake CRs generated through the experimental manipulation.

4.2 Effects of high vs. low co-ordination

Closer examination of the data collected reveals a differential pattern in CR responses in early vs. late games. During the first few mazes, when the participants are relatively inexperienced in the task, CRs are interpreted in familiar ways (Purver 2004; Ginzburg and Cooper 2004; Schlangen 2004) as querying the referential import of the constituent concerned:

1)	**A:**	5, 6
2)	**Server:**	what? / 5?
3)	**A:**	5 across, 6 along counting from the left hand side of the maze

Excerpt 4: Clarifying referential import

At late stages of the interaction though, CRs are more likely to be interpreted as questioning what the target-turn as a whole "is doing" in the sequence (see Drew 1997). Here both fragment

[4] Note that the setting differs from Foster et al (2009), hence the distinct results: both participants here, even though instructed of the general goal of the task, do not initially know what the problem involves and what strategy they should develop to deal efficiently with it.

[5] See also Garrod and Anderson (1987) and Healey (1997) for an account of the semantic development of spatial descriptions in the maze game.

and "What" CRs are interpreted significantly more frequently as concerning the intention or plan behind the target utterance:

1)	**A:**	4th square along
2)	**Server:**	what? / 4th?
3)	**A:**	because you've got to go there / you asked me to go there.

Excerpt 5: Clarification request interpreted as querying intention/plan

1)	**A:**	5, 6
2)	**Server:**	what? / 5?
3)	**A:**	that's where I'm blocked, can you get me out of there?

Excerpt 6: Clarification request interpreted as querying intention/plan

Responses to these CRs reveal the participants' projected analysis of the structure of the joint project into "moves" (sub-projects) through their explicit identification of the purpose of the target turn. This shows that at these late stages of the interaction when sufficient coordination has emerged, participants are able to formulate explicitly the plans and intentions underlying their actions and utterances. This contrasts with earlier stages, where participants seem unable to formulate such plans as the underlying causes of their actions, compounded by the observation that, even when participants attempt to co-ordinate in this way, more often than not, these attempts prove unsuccessful. This result is compatible with Ginzburg's (2003) observations that show, on the basis of corpus research, that recognition of underlying plans and intentions is not necessary for grounding .

Looking then at the aetiology behind the development of the joint project (as illustrated by Excerpts 1, 2, 3), these results seem to suggest that such intentions and plans emerged from interaction, and did not precede it. This is substantiated by the observation that explicit negotiation over developing a sequence of steps for solving the maze is more likely to impede and be ignored in the initial stages:

1)	**A:**	OK, first you've got to tell me where to go and then I can go through
2)	**B:**	where are your switches?
3)	**A:**	tell me where to go so I can get through
4)	**B:**	I'm blocked by the gate in front of me

Excerpt 7: Failure at explicit co-ordination

4.3 Effects of sequentiality on coherence

The data strongly suggest that in both early and late stages in the development of joint projects and their associated moves, *sequentiality*, that is, the possibility of setting up conditional relevances between utterance pairs, plays a key role as the progenitor of co-ordination.

On the one hand, in the later stages of the maze game, highly elliptical exchanges (telescoping) as illustrated by Excerpt 1 are fully reliant on sequential position for their interpretation due to each move in the fully-formed joint project projecting completion of some prior and also projecting the relevant next move (*coherence*). On the other hand, in the early stages of the maze game, while the nascent joint project is still vaguely defined, interlocutors may still rely on explicit mentions of *coordination devices* like "switch", "gate", "goal". As shown by the results obtained in the reduced sequentiality condition (Dual window), at these stages, disruption of the sequential organisation can be seen to have detrimental effects on coordination: interfering with adjacency of turns has an adverse effects on negotiating these "co-ordination points" (see Figure 3).

4.4 Conclusions

These results appear to undermine accounts of co-ordination that rely on an a priori notion of (joint) intentions and plans (see also Clark 1996) and also accounts which rely on some kind of strategic negotiation to mediate coordination (e.g. Alterman 2001 who claims that explicit negotiation is the "co-ordination mechanism" par excellence to deal with coordination problems). Instead, according to the the data, participants, at initial stages of the task, employ a minimal amount of explicit attempts at coordination, and these attempts are either ignored or fail. This and the pattern of differential CR interpretations (early vs late) strongly suggest that planning and intention-recognition are mechanisms only successfully employed at late stages, once interlocutors are sufficiently coordinated. These observations seem consonant with an alternative approach to planning and intention-recognition according to which forming and recognising such constructs is an activity subordinated to the more basic processes that underlie people's performance (see e.g. Agre and Chapman 1990; Suchman 1987/2007). Taken cumulatively with the finding that reducing sequential coherence makes procedural coordination more problematic, the results here strongly suggest that the tacit co-ordination mechanisms of turn-by-turn feed-

back in dialogue provide a richer set of resources than those possible in attempts to describe and resolve co-ordination problems explicitly.

Acknowledgments

This work is supported by an individual Marie Curie IOF fellowship (2010-2012) Grant no.: PI-OF-GA-2009-236632-ERIS, DiET (EPSRC /D057426/1), DynDial (ESRC-RES-062-23-0962). We gratefully acknowledge advice and support from Pat Healey, Ruth Kempson, Matt Purver, Chris Howes, Arash Eshghi.

References

Agre P. E., Chapman D. (1990) What are plans for? In: Maes P. (ed) *Designing autonomous agents: theory and practice from biology to engineering and back*. MIT Press, Cambridge, pp 17–34

Alterman, R. and Garland, A. (2001). Convention in Joint Activity. Cognitive Science, 25:611-657.

Anderson, J. F., Beard, F. K., and Walther, J. B. (2000). The local management of computer mediated conversation. In: Herring (ed.) Interactional coherence in CMC.

Atkinson, J. Maxwell, and Paul Drew. 1979. *Order in court.* Oxford Socio-Legal Studies. Atlantic Highlands, NJ: Humanities.

Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge, UK.

Cohen, P.R., Morgan, J., and Pollack, M.E. (eds.) (1990). *Intentions in Communication*, MIT Press.

Drew, P. (1997). 'Open' class repair initiators in response to sequential sources of troubles. Journal of Pragmatics, 28:69-101.

Foster, M. Giuliani, M, Isard, A, Matheson, C., Oberlander, J. and Knoll, A. (2009). Evaluating Description and Reference Strategies in a Cooperative Human-Robot Dialogue System". In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI-09)"

Garrod, S. and Anderson, A. (1987). Saying what you mean in dialogue. Cognition, 27:181218.

Ginzburg, J. (2003) "Disentangling public from private meaning", In: R. Smith and J. van Kuppevelt, editors, *New Directions in Discourse and Dialogue*, Kluwer. Grosz, B. and Sidner C. (1986). "Attention, Intentions, and the Structure of Discourse." Computational Linguistics, 12,3:175-204.

Healey, P.G.T. (1997). "Expertise or expert-ese: The emergence of task-oriented sub-languages." *Proceedings* o the Nineteenth Annual Conference of The *Cognitive Science Society.* Stanford University, CA. p. 301-306.

Healey, P.G.T. and Mills, G. (2006). Participation, precedence and co-ordination. In Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada.

Krauss, R. M. and Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology* 4:343-346.

Levinson, S. (1983). *Pragmatics*. Cambridge University Press.

Mills, G. J. (2007). The development of semantic co-ordination in dialogue: the role of direct interaction. Unpublished PhD Thesis.

Pickering, M. J, and Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences,* 27(2):169-190.

Poesio, M. and Rieser, H. (2010) Completions, co-ordination, and alignment in dialogue Dialogue and Discourse, 1.

Poesio, M. and Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3) :1-45.

Schegloff E. A. (1972) "Notes on a Conversational Practice: Formulating Place," in D. N. Sudnow (ed.), Studies in Social Interaction. New York: MacMillan, The Free Press, 75-119.

Schegloff E. A. (1979) "Identification and Recognition in Recognition in Telephone Conversation Openings" In G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology* New York: Irvington Publishers, Inc., 1979 23-78.

Schegloff E. A. (1992) Repair after next turn *American Journal of Sociology* 97(5):1295-1345.

Schlangen, D. (2004) Causes and Strategies for Requesting Clarification in Dialogue. In Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue (SIGDIAL 04), Boston, USA, April.

Schelling, T.C. (1960). *The strategy of conflict*. Cambridge MA: Harvard University Press.

Suchman, L. (1987/2007) *Human-Machine Reconfigurations: Plans and situated actions*. Cambridge University Press, New York.

Practices in Dialogue

David Schlangen
Department of Linguistics
University of Potsdam, Germany
david.schlangen@uni-potsdam.de

Abstract

We explore the idea that conversational episodes not only ground *facts*, but also establish *practices* (know how). We apply this idea to the well-studied phenomenon of lexical entrainment. We provide a formal model of situations giving rise to lexical entrainment, and use it to make precise implications of extant attempts to explain this phenomenon (the lexical pact model, and the interactive alignment model) and to make precise where our attempt differs: it is an automatic and non-strategic, but goal-driven account, which has a place for partner-specificity and group forming. We define a learning mechanism for practices, and test it in simulation. We close with a discussion of further implications of the model and possible extensions.

1 Introduction

Agnes and Bert are playing a game. Agnes is describing a figure printed on a card to Bert, and Bert is trying to find a copy of the card among several similar cards. They do this for a while, and begin to establish names for the figures, which they stick to during the course of the game.

This of course is the well-known reference game first described by (Krauss and Weinheimer, 1964), and made famous by (Clark and Wilkes-Gibbs, 1986; Brennan and Clark, 1996). The latter analyse this phenomenon as one of negotiating and forming a *pact*, where this metaphor is justified by observing that this 'pact' has parties for which it holds (in our example, Agnes and Bert, but not Agnes and Claire, who wasn't party to the conversation), and that a party that assumes she has such a pact reacts to it being 'broken' by their partner (Metzing and Brennan, 2003).

A well-known alternative proposal, that of Pickering and Garrod (2004), would assume that Agnes and Bert become, in some real sense, more like each other while playing the game: the representations that represent world and language to them become more alike. Unlike in Clark *et al.*'s proposal, the assumption here is that this is an automatic process that doesn't require reasoning about the partner.

In this paper, we explore an idea that to some extent is a synthesis of elements of both these proposals, namely the idea that rounds in the reference game, and more generally, all kinds of interactional episodes, establish *practices* (a notion we take from anthropology, e.g. (Ortner, 1984; Bourdieu, 1990)), a form of know-how; they, automatically and unreflectedly, establish *a way of doing* something *with someone*, i.e., a way that is specific to a (type of) partner. We make precise what this can mean for the reference game, and speculate more generally on how this can complement models of interaction.

2 Lexical Entrainment in the Card Matching Game

By *lexical entrainment* (a term coined by (Garrod and Anderson, 1987)) we mean here the phenomenon that pairs of interlocutors tend to settle over the course of a conversation on a single description for a given object for which several descriptions would possible (understandable, allowable by the language). This seems to be a very stable phenomenon, observed in symmetric conversations between players (e.g., (Krauss and Weinheimer, 1964; Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Brennan and Clark, 1996)), in conversations between novices and experts (where the experts' terms come to dominate; (Isaacs and Clark, 1987)), in "conversations" between humans and computers (where the strength of the effect depends on the human's beliefs about the capabilities of the computer; (Branigan et al., forthcoming)), and in conversations between 'nor-

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

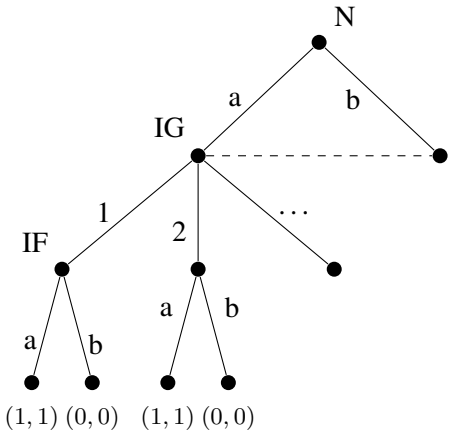


Figure 1: The game in extensive form

mal' dialogue participants and participants suffering from severe amnesia (Duff et al., 2006).

In this section, we provide a formal model of a situation in which the phenomenon has often been studied, and note some characteristics of extant explanation attempts.

2.1 The Game

Nature picks one card from set $\mathcal{O}$. Player A, in role IG (instruction giver), observes this, but player B, in role IF (instruction follower), does not. IG plays from set $\mathcal{V}_A$, her vocabulary. (Note that this set is indexed by player, not role.) IF observes this, and as reaction picks one card out of set $\mathcal{O}$ as well. If IF's pick is equal to nature's pick, both players win. Otherwise, both lose. Players are told whether they have won or lost, and the game is repeated.

The similarity in the formulation above to signalling games (sequential games of incomplete information) studied in game theory (e.g., (Shoham and Leyton-Brown, 2009)), is of course not accidental. Figure 1 shows the game in extensive form, for $\mathcal{O} = \{a, b\}$, and $\mathcal{V}_A = \{1, 2, \ldots\}$ (it doesn't matter here what exactly is in this set). The setting is slightly different, though, and we are interested in slightly different questions. First, unlike in signalling games, where the question is how stable signalling systems evolve, we assume here that there are already conventions between A and B in place. Specifically, we assume that A and B, by virtue of belonging to the same language community (which they know they do), mutually believe that 1 and 2 are good labels for a, and not so good labels for b (and so on, for other combinations of pairs of labels and entities). Second, the question that we are interested in is why it is, if they play this game repeatedly, possibly switching roles between episodes, that when nature picks an element of $\mathcal{O}$ they've encountered before, IG becomes more likely to chose that element from her vocabulary that has been chosen in previous successful episodes—and not any other one which she has reason to believe could be successful as well. (Remember that they assume all normal language conventions hold.)

2.2 Lexical Pacts

The model of Brennan and Clark (1996) is characterised mostly by a list of features: it is *historical*, meaning that previous choices influence the current choice, mediated by *recency* and *frequency of use*, and also influenced by *provisionality*, that is the requirement of lexical choices to be sanctioned by the other party. An aspect that the authors stress is that the model also assumes *partner specificity*: speakers choose their wording "for the specific addressees they are now talking to" (Brennan and Clark, 1996, p.1484).

The authors do not say much about how exactly a model with these characteristics might be realised, apart from "[the results imply that] long-term memory representations are involved" (p.1486), and, from a follow-up paper: "entrainment is supported by an underlying episodic representation that associates a referent, a referring expression (and the perspective it encodes), and other relevant information about the context of use (such as who a partner is)" (Metzing and Brennan, 2003, p.203).

The following is our attempt at providing a formalisation of a model that at least is compatible with the general tone of the description in the papers cited above. It assumes that the episodic representations are explicitly encoded, perhaps along the lines of the reference diaries from (Clark and Marshall, 1978), and that the rule "reuse of labels maximises success" is explicitly represented and used in making the decision. (That explicit inference is involved in the model at least is a charge that opponents seem to make often, e.g. "on their account, alignment is therefore the result of a process of negotiation that is specialized to dialogue and involves inference." (Garrod and Pickering, 2007, p.2).)

If there is reasoning involved, it must be a form

of practical reasoning, as the outcome is a way to act. The following formulates a schema in the form of an Aristotelic practical syllogism which could be assumed to underlie that reasoning.

a. I want to refer to object x for you
b. I believe that honouring our pact regarding the use of α as a label for x is the best way to do (a).
or
I believe that our last successful reference to x was with label α, and that re-using the last most recent successful label is the best way to do a reference.
c. Therefore, I honour our pact, and use α.

This, then, is in this (interpretation of the) model the type of reasoning that IG must perform when it comes to choosing an action, and IF must follow the corresponding version of this scheme governing interpretation. In the course of playing several episodes of the game, the set of beliefs of the partners changes, to include beliefs about which terms were used successfully when.

2.3 Interactive Alignment

The *interactive alignment model* of Pickering and Garrod (2004), which is positioned as a counter-proposal to the one discussed above, is similarly only characterised indirectly and not provided with a formalisation. Its main features are nicely summarised in (Garrod and Pickering, 2007, p.2): "Our claim is that a major reason for this alignment is that the comprehension of *chef* (or alternatively *cook*) activates representations corresponding to this stimulus in B's mind (roughly corresponding to a lexical entry). These representations remain active, so that when B comes to speak, it is more likely that he will utter *chef* (or *cook*). [...] We assume that this tendency to align is automatic."

Applied to our formalisation of the game, it seems that this model assumes that IG bases her choice on whatever is most strongly activated in her mind, and, if understood at all, this representation will increase in activation in IF, making him more likely to use the label as well at the next opportunity. Hence, what changes here over the course of playing the game are the vocabularies of the players (or rather, the way they are represented in the minds of the players); beliefs with propositional content and reasoning do not enter in the description of the phenomenon. There does not seem to be room for partner-specificity in this model.

We now turn to our model, which, as we've mentioned, combines elements of both of these approaches.

3 A Learning Mechanism for Practices, Applied to the Card Matching Game

3.1 Modelling the Players

We assume that there is an additional structure that explicitly represents an agent's associations between objects and labels. Formally, for each player there is a *map* $\mathcal{M}$, which is a matrix where the rows represent the objects from $\mathcal{O}$ and the columns represent the labels from $\mathcal{V}$. Each cell $\mathcal{M}_{o,v}$ encodes the strength of association between object o and label v. Together with a strategy for using such associations, the maps determine the interpretation of an observed label (looking at the respective column), or the label to chose when wanting to refer to an object (looking at the row). Later, we will make the maps relative to *partner* and *situation types*, e.g. $\mathcal{M}^E$.

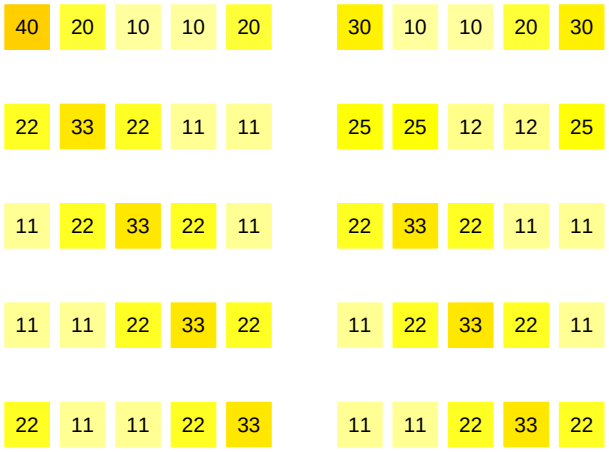


Figure 2: Two Maps (rows are objects, columns are labels, numbers and colours represent association strength)

Figure 2 shows two maps $\mathcal{M}_A$ and $\mathcal{M}_B$ (values are shown as number and are also coded in colour). It is read as follows: for object a (first row), both players rate label 1 (first column) highly; A prefers it, while B thinks label 5 is just as appropriate. *All* labels are somewhat acceptable to refer to a (and in fact, all labels are acceptable for any object). For objects c, d, e (rows 3 to 5), A and B have different preferred labels, and if their partner picks their preferred label to refer to any of these objects,

they would not pick out the same object, if they go for their strongest interpretation preference. E.g., if A says 3 with the intention to refer to object c, B will understand d, if both go for their strongest preference. (This is a fairly extreme example, with lots of ambiguity, and differences in preferences; we will use this below to show that our learning method can still lead to agreement even in such a situation.)

3.2 Amended Rules of the Game, and Learning

We define two versions of the game. In the simple version, the game is exactly as stated above in Section 2.1. At each decision point in an episode, the players choose *greedily*, that is, they go for the appropriate cell with the highest value. E.g., if IG observes nature picking object a, she checks what the highest value is in the corresponding row, and in this way finds the label she wants to use (as the one corresponding to the column of the highest value). IF now, on observing IG using the label, checks in the column corresponding to the label which row has the highest value, and takes the object corresponding to this row as his interpretation, which in this version of the game means that he solves and picks this object.

The episode then closes with *rewards* being distributed to the players, as follows. The choice made by IG is always reinforced (the weight of the cell used in deciding which label to use is increased by some factor), even if IF made the wrong choice, and for *both* players. The idea here is that the mapping IG has used is, once the game has been revealed, mutual knowledge, and so IG should be motivated to use it again, as should IF. Additionally, if IF made a wrong choice, this wrong choice is punished (the weight of the cell used in deciding which label to use is decreased by some factor).[1]

There is an interesting aspect of the game that we have left implicit so far: when players switch roles, they will use the map which in the previous episode they used for generation (finding a label, given an object) or interpretation (finding an object, given a label) for interpretation or generation, respectively. This ensures, without further stipulations, that there is symmetry between interpretation and production: when they have used a label successfully with a certain interpretation, they will become more likely to use this interpretation when they hear this label.

[1]This game is very similar to the one analysed by (Argiento et al., 2009), who present a similar urn-based (i.e., numerical weight-based) learning scheme; our game, however, diverges in two important aspects: we assume that players start with non-uniform distributions (they already have preferences for labels/object combinations), and unlike (Argiento et al., 2009) we do not reward or punish both players equally. These authors are concerned with analysing whether a signal system can be learned this way, we are concerned with whether an existing signal system can be adapted.

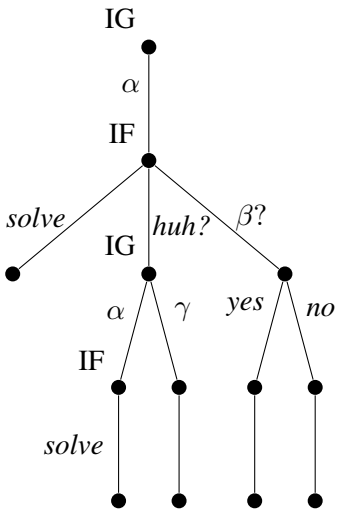


Figure 3: Extended action space for variant of game with Clarification Requests

The second version of the game is somewhat more complicated, as it gives the players more options for acting. (The extended action space is shown graphically in Figure 3, for one branch after IG's first action.) After IG has played (has provided a label α for object o chosen by Nature), IF can decide whether to solve (pick an object), or make a clarification request (CR). It uses the following rules to decide what to do: if there are "distractor objects" for the label IG used, that is, if the highest weight in the appropriate column is not sufficiently far away from other weights in the column, then IF rejects, and ask a CR that could be paraphrased as "huh?". If this is not the case, but there is another label that IF deems more appropriate for the object that is the best interpretation given this label (in the row in which the highest weight of the column corresponding to IF's label is found, there is another cell with a higher weight), then IF asks a reformulation CR, proposing this other label.

IG now reacts as follows to clarification questions: In reply to a reject-CR ("huh?"), IG will use the second best label for the object that is to be

named, if there is one; otherwise, it will just repeat the previous label. In reply to a reformulation-CR, it will say "yes" if the proposed alternative label is within a certain range of the original label (for object o), "no" otherwise. Back to IF: on hearing "yes", IF will solve with the mapping it used for the reformulation. If IG said "no", IF picks the best interpretation for the original label (even though for that object IF thinks there is a better name). If IG offers a reformulation, IF tries to find the object that is best for both labels, but if there is none, he picks the first one.

Rewarding the players after an episode is somewhat more involved in this variant of the game. We now potentially have several different cells involved in one episode (for making the initial proposal, for formulating a CR, for answering it). Rewards are now computed according to formula (1), where c_n denotes the cell used at step n, counting backwards and starting with 0, so step 0 is the last utterance of the player being rewarded. We use hyperbolic discounting on the reward, with some discount factor δ; the rationale here is that the exchange towards the end of the episode is more important and impresses itself more on the players.

$$c_n \leftarrow c_n + \frac{1}{(1+\delta * n)} \quad (1)$$

We again follow the principle that IG is always rewarded, and IG's choices—they are now transparent for IF, since IF knows what IG's goal was—are also reinforced for IF. IF's own choices are rewarded or punished depending on the success of the episode.

Before we turn to the experiments we performed with simulations of these games, two more remarks on the second variant of the game. First, what is the idea behind these rules (which seem positively baroque compared to the austere rules of Game Theory games)? They are meant to be a relatively plausible model of how clarification actually works (Purver et al., 2001; Purver, 2004; Schlangen, 2004).[2] They have the effect of letting the participants more quickly explore their semantic maps, as we will see below. (This perhaps incidentally offers corroborating evidence for the thesis that clarification behaviour supports language construction (Ginzburg and Macura, 2006).)

[2]Of course, ideally, these rules would be learned as well; though not in the same game. We are assuming adult-level language competence here. Learning clarification strategies, perhaps modelling (Matthews et al., 2007), is an orthogonal problem.

As a second remark, the reader is advised to note that the players at no point keep a model of their partner. They always only keep their own map; a feature of this model that brings it closer to the interactive alignment model than to the lexical pacts model.

We now turn to simulation experiments with this game and their results.[3]

3.3 Experiments, and Results

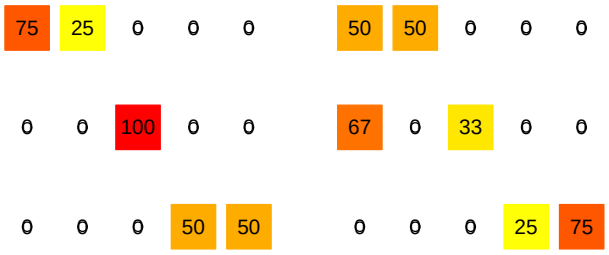


Figure 4: Two maps with less ambiguity. Three objects, five labels.

We implemented the variants of the game in a computer program and ran experiments with two different map sets, the one from Figure 2, and a smaller map set with less ambiguity (Figure 4). As evaluation measures we use *task success*, and *average map distance*, which we define as the mean root squared distance between cell values: $ad(\mathcal{M}^\alpha, \mathcal{M}^\beta) = \sum_i \sum_k \sqrt{(\mathcal{M}^\alpha_{i,k} - \mathcal{M}^\beta_{i,k})^2} * \frac{1}{i*k}$

Figure 5 shows a plot of the average map distance in Experiment 1 (with the maps from Figure 2), for the variant with clarification requests (solid lines) and that without (dashed lines). Superimposed are the successful episodes (green circles) and the unsuccessful ones (blue boxes). We see that in both variants the distance between maps decreases steadily (faster for the variant without CRs), and after some episodes, successes become much more likely than failures (much later for the variant without CRs). What may the reason be for the differences? In the variant with CRs it takes the players longer to align their maps, because there is more room for misunderstanding, as there is more exploration. The flipside of this is that they are faster to explore the whole map, and so have heard all relevant labels used earlier. In the variant without CRs, each episode only teaches them

[3]We would like to, but cannot at the moment offer a formal analysis of properties of this learning method, and so can only look at how it fares on the task.

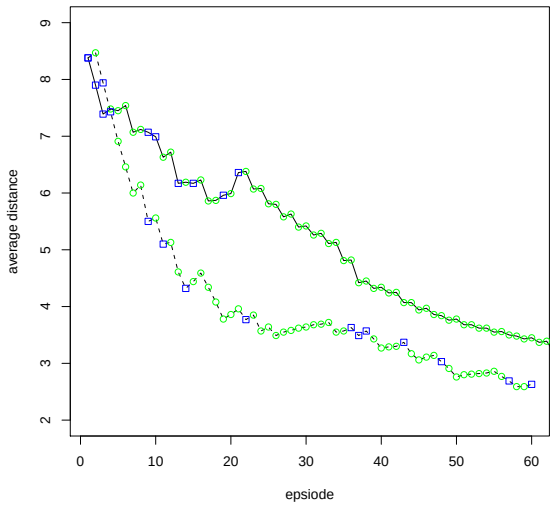


Figure 5: Results of Experiment 1

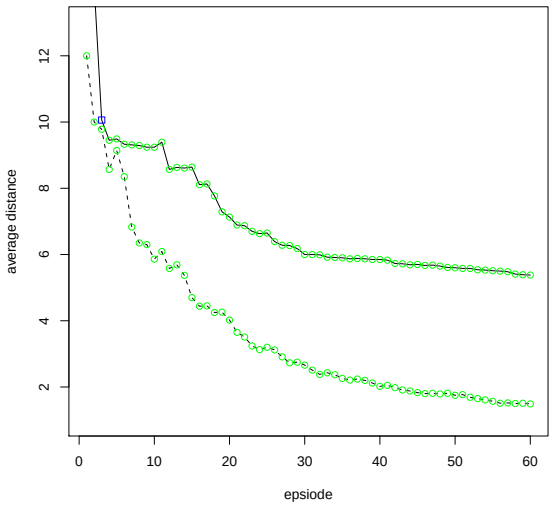


Figure 7: Results of Experiment 2

about one mapping, and so it can happen even after many episodes that they hit a pairing that they haven't sufficiently learned.

64	10	5	5	15	54	3	7	10	26
19	43	19	9	9	21	37	10	10	21
7	13	60	13	7	15	23	46	8	8
6	6	11	66	11	7	14	14	59	7
18	9	9	18	46	8	8	16	24	45

Figure 6: The maps after 20 episodes

Figure 6 shows the two maps from Figure 2 after 20 episodes. The figure shows nicely how similar the maps have become. Note that the players have aligned on player A's map; this seems to be due simply to the fact that A was IG first, and so had a slight, but decisive, advantage in spreading her preferences.

Figure 7 shows the results for Experiment 2 with the simpler maps. We see the same tendency for the variant with CR to decrease less quickly; as the maps are so simple (but arguably much more representative of normal Matching Game settings), the partners very quickly stop making any mistakes.

We take these results as evidence that the learning mechanism does indeed lead to "lexical entrainment" between the partners: they come to be much more likely to re-use successful labels, and this indeed increases their success. So, from the looks of it they do indeed form "conceptual pacts", but without any model of their partner, just by accidentally becoming more like each other.

"Is this not just priming, then?", one may be tempted to ask. If we understand the idea behind priming correctly, then that is a mechanism driven by occurrence, not by success. In our model, reward is given (to IF) only when success is reached. In that sense, our model is goal-driven, whereas priming is not. Alternatively, our model may be seen as a way to spell out what priming is meant to achieve—we are not aware of much work on computational models of priming with respect to this phenomenon. (There is work by Reitter (2008), but that is modelling corpora and not providing simulations. More closely related is (Buschmeier et al., 2009), where alignment and priming is modelled by activation functions. However, these do not take into account the role of clarification requests and of success in reaching a communicative goal.)

In any case, however, what we have done so far only provides half of the story. What will happen when a player encounter a new, naive agent,

or when a previous partner starts to 'misbehave'? This part we have not explored with experiments yet. As the setup is at the moment, players will react in the same way to a new partner using different terms and to their old partner suddenly using different terms. This contradicts the findings of Metzing and Brennan (2003) discussed above. We discuss in the next section how the model could be made to account for this phenomenon as well.

4 Necessary Extensions

We list in this section extensions to the model that are necessary to bring it towards more fully capturing the phenomenon (and, more ambitiously, to making sense of the label *practice*).

The first of those extensions will need to capture the partner-specificity observed by Metzing and Brennan (2003) (reactions to previous partner switching terms) and Branigan et al. (forthcoming) (different strength of alignment depending on beliefs about capabilities of (artificial) partner). At the moment, we can only outline the likely shape of this part of the model, and mention some requirements that we would like to see met.

First, we widen the scope of the discussion. A different view of the maps from the previous section: We have introduced them as linking objects and labels, but, viewed in a more general way, what they link is a way of performing an act of referring (or an act of understanding) to the goal of making the partner pick out an object (or the goal of picking out the object the partner wants me to pick out). The maps were a special case of what more generally we will call a *policy*, which is, then, a structure that links goals and actions.

The idea now is that what is being modified during the course of an activity is a "local" policy, particularised to the current situation. The policy from which one started remains stable, and a new copy is branched off for the current conversation. Some process of abstraction over situations then needs to categorise and generalise: will the local policy be of use again, do the changes need to merged with the starting policy? In such inconsequential situations as the matching game, the policy (the map) might be associated in memory with a particular speaker, and then slowly be forgotten (or rather, regress to the base line map). But in more weighty situations, a policy will become associated with a group of partners, or a type of situation; and conversely, a group is formed through the association with a policy. Membership in the group (being in a situation / doing a certain activity) then entails following this policy, this practice; and deviating from the practice rules out membership, or leads to irritation when done by a member of the group. (I.e., practices are a form of situated normativity, (Rietveld, 2008).) This is what a practice is, then: a policy that holds in certain situations, for certain groups of agents.

Returning to the more concrete subject matter of this paper: how does this account for the phenomenon of partner-specificity of lexical entrainment? It can explain the effect that "breaking the pact" has, the irritation that the partner exhibits at this, if we assume that breaking the pact simply is not behaving in the way that the chosen policy for this situation (which includes this partner) predicts. If we assume as a further element that the mechanism for copying and adapting maps is sensitive to aspects of the situation, then we might have a handle on the fact that beliefs about the partner can influence alignment with it. The results of (Branigan et al., forthcoming) might then be explained by differences in the assumptions of group membership: the computer believed to be more advanced is assumed to conform stronger to normal practices, whereas the simpler one does not trigger strong presuppositions. Working out these rather general ideas will have to remain for future work, however.

We close this section by noting that the general ideas sketched here are far from being new. The godfather of this line of thinking of course is Wittgenstein (1953) with the notion of *language game*. The way we have restricted policies to situation types (or rather, have stipulated that this should be done) could be seen as one way of spelling out this notion. The particular line on Wittgensteinian thought I'm following here is further represented by (Levinson, 1992) and (Bourdieu, 1990) (from whom the label *practice* is taken). More recently, the idea of conversations establishing micro-languages explored by Larsson (2008) is very closely related; our understanding is that the ideas sketched here should be complementary to this approach. The precise connections to this and other related work, however, need to be worked out more clearly in future work.

5 Conclusions

We have provided a formal model of a game which has often been used to study the phenomenon of lexical entrainment (Garrod and Anderson, 1987).

We have used this model to make precise some implications of the lexical pact model (Brennan and Clark, 1996), and the interactive alignment model (Pickering and Garrod, 2004), and our own practice-based proposal, which we have also tested in simulation, showing that it can model the process of aligning on language use in this very simple game.

What we have provided here is quite clearly only a first sketch of such a practice-based account of lexical entrainment (and, much more so, a first sketch of a more general theory of the role of practices in conversation). It is our hope, however, that our formalisation of the problem as outcome of a kind of a signalling game already is a useful contribution, and helps to better understand what the debate actually is about—and that our sketch of a practice-based account has at least succeeded in positioning it as an alternative that would be promising to work out further.

Acknowledgments This work was funded by DFG Grant SCHL845/3-1, Emmy Noether Programme.

References

Raffaele Argiento, Robin Pemantle, Brian Skyrms, and Stanislav Volkov. 2009. Learning to signal: analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*, 119(2):373–390.

Pierre Bourdieu. 1990. *The Logic of Practice*. Stanford University Press, Stanford, CA, USA. *transl.* R. Nice; orig. *Le sens practique*, Les Éditions de Minuit, 1980.

Holly P. Branigan, Martin J. Pickering, J. Pearson, and J.F. McLean. *forthcoming*. Linguistic alignment between humans and computers. *Journal of Pragmatics*.

Susan E. Brennan and Herbert H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493.

Hendrik Buschmeier, Kerstin Bergmann, and Stefan Kopp. 2009. An alignment-capable microplanner for natural language generation. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 82–89, Athens, Greece.

Herbert H. Clark and Catherine R. Marshall. 1978. Reference diaries. In D. L. Waltz, editor, *Theoretical Issues in Natural Language Processing (Vol. 2)*, pages 57–63. ACM, New York, USA.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

Melissa C. Duff, Julie Hengst, Daniel Tranel, and Neal J. Cohen. 2006. Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, 9(1):140–146.

Simon Garrod and Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218.

Simon Garrod and Martin J. Pickering. 2007. Automaticity in language production in monologue and dialogue. In A.S. Meyer, L.R. Wheeldon, and A. Krott, editors, *Automaticity and control in language processing*, pages 1–21. Psychology Press, Hove, UK.

Jonathan Ginzburg and Zoran Macura. 2006. Lexical acquisition with and without metacommunication. In *The Emergence of Communication and Language*, pages 287–301. Springer, Heidelberg, Germany.

Ellen A. Isaacs and Herbert H. Clark. 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116(1):26–37.

Robert M. Krauss and Sidney Weinheimer. 1964. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1:266–278.

Staffan Larsson. 2008. Formalizing the dynamics of semantic systems in dialogue. In Robin Cooper and Ruth Kempson, editors, *Language in Flux – Dialogue Coordination, Language Variation, Change and Evolution*. College Publications, London, UK.

Stephen C. Levinson. 1992. Activity types and language. In P. Drew and J. Heritage, editors, *Talk at work: Interaction in Institutional Settings*, pages 66–100. Cambridge University Press, Cambridge, UK.

Danielle Matthews, Elena Lieven, and Michael Tomasello. 2007. How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, 78:1744–1759.

Charles Metzing and Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49:201–213.

Sherry Ortner. 1984. Theory in anthropology since the sixties. *Comparative Studies in Society and History*, 26(1):126–166.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

Matthew Purver, Jonathan Ginzburg, and Patrick Healey. 2001. On the means for clarification in dialogue. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg, Denmark.

Matthew Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College, Unversity of London, London, UK, August.

David Reitter. 2008. *Context Effects in Language Production: Models of Syntactic Priming in Dialogue Corpora*. Ph.D. thesis, School of Informatics, Edinburgh, UK.

Erik Rietveld. 2008. Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind*, 117(468):973–1001.

David Schlangen. 2004. Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th Workshop of the ACL SIG on Discourse and Dialogue*, Boston, USA, April.

Yoav Shoham and Kevin Leyton-Brown. 2009. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, UK.

Ludwig Wittgenstein. 1953. *Tractatus Logicus Philosophicus und Philosophische Untersuchungen*, volume 1 of *Werkausgabe*. Suhrkamp, Frankfurt am Main. this edition 1984.

Modelling Sub-Utterance Phenomena in Spoken Dialogue Systems

Okko Buß **David Schlangen**
Department of Linguistics
University of Potsdam, Germany
{okko|das}@ling.uni-potsdam.de

Abstract

Dialogue systems that process user contributions only in chunks bounded by silence miss opportunities for producing helpful signals, such as backchannel utterances concurrent with the ongoing utterance—to the detriment of interaction quality. We survey in some detail what we call 'sub-utterance' phenomena, which such an approach misses, and then discuss two approaches to overcoming this limitation. The first is to add a 'reactive layer' to an otherwise unchanged, utterance-based dialogue manager. The other approach, less often taken, is to make the dialogue manager itself capable of producing such reactions. To explore the viability of such an approach, we sketch a dialogue management strategy capable of working at a sub-utterance level.

1 Introduction

It is generally agreed that human language processing in dialogue proceeds continuously (i.e., not at certain points in bulk, but at all times during the contribution of the other participant) and incrementally (with new bits of information building on previous ones, forming larger units); see e.g. (Marslen-Wilson and Tyler, 1981). This is a fact that manifests itself in subtle phenomena like eye movement patterns, studied in psycholinguistics (Tanenhaus et al., 1995), but also in 'surface-phenomena' like backchannel utterances ("uhu"), studied in Conversation Analysis for example (e.g., Schegloff (1982)).

Dialogue systems, but also dialogue theories, are, in contrast, more utterance-oriented. They typically discretise dialogue into utterance-chunks, either for technical reasons (simpler segmentation boundary detection) or for theoretical ones (propositions as smallest unit).[1]

In this paper, we focus on what we label 'sub-utterance phenomena', and ask how we can equip dialogue systems with the capability to produce and understand those. We discuss two possible approaches, one where additional machinery takes care of such phenomena, and another where the dialogue manager is adapted more fundamentally. As the latter is an approach that is less represented in the literature than the former, we sketch an approach to dialogue management that promises to enable this strategy.

The remainder of the paper is structured as follows. We first survey in some detail the phenomena of interest, looking at what may be appropriate reactions to user contributions that exhibit them, and what may be internal events that might require a system to produce them. We then discuss the two approaches to modelling such phenomena and sketch our attempts at the second type of approach.

2 Sub-Utterance Phenomena

So far we have used the term "sub-utterance phenomenon" informally. To make more precise what we mean by it, we need to explain in a bit more detail the way that most current dialogue process user contributions: In such systems, the speech recogniser delivers output only once it has detected silence of a certain duration (often something between 750 and 1500ms, (Ferrer et al., 2002)). Hence, the unit for processing for later modules is a continuous stretch of user speech ending in silence. This segmentation is performed without input from later modules (like parsers of dialogue managers), and those later modules only see complete utterance units. "Sub-utterance phe-

[1] With the notable exceptions of, on the theoretical side, PTT (Poesio and Traum, 1997; Poesio and Rieser, 2010) and Dynamic Syntax (Cann et al., 2005), and on the implementational side (DeVault and Stone, 2003; Aist et al., 2007; Skantze and Schlangen, 2009).

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

nomena", in our use of the term, then are all phenomena that make reference to units smaller than such silence-bounded speech chunks.

Table 1 gives an illustration of the types of phenomena we are interested in here. We divide our analysis of the phenomena into what possible system reactions are to user-produced sub-utterance phenomena, and what internal system events could be that require a system to produce them.

2.1 System Reactions to User-Produced Sub-Utterance Phenomena

Hesitations The first phenomenon we look at, hesitations, or more specifically, unfilled pauses, poses a direct problem to the approach to contribution segmentation sketched above. As this approach uses silence to endpoint the user utterance, there is a danger of wrongly endpointing during a hesitation, and being left with an incomplete utterance (and confusion when the user then resumes talking). Hence, as a 'minimal' type of reaction to a hesitation we would ideally like the dialogue system to not confuse it with an utterance end, and to simply wait for the speaker to continue. A more sophisticated system could offer signals of support, such as backchannel utterances, or even cooperative replies such as suggesting words the speaker may be looking for, or even completing the utterance for her. (See for example (Clark, 1996) for a discussion of such strategies.)

A more subtle effect of hesitations, and disfluencies in general, has been discussed in the psycholinguistics literature: under certain conditions, dialogue participants seem to draw conclusions from the fact that a speaker is disfluent. When producing descriptions of objects, disfluencies can be taken as indication that an object with a non-obvious name is being described (Brennan and Schober, 2001; Bailey and Ferreira, 2007; Arnold et al., 2007); a system that can attend to sub-utterance phenomena could make use of such implications (Schlangen et al., 2009).

Backchannel Utterances Backchannel utterances are "short messages such as *yes* and *uh-huh*", which are given "without relinquishing the turn" (Yngve, 1970, p. 568). Ward and Tsukahara (2000, p.1182) give a more formal definition:

Back-channel feedback:

> (D1) responds directly to the content of an utterance of the other,
> (D2) is optional, and
> (D3) does not require acknowledgement by the other.

From these descriptions, we can directly derive what reactions of a dialogue system to user backchannel utterances should be. First, they should not be taken as an attempt by the user to take the floor (D3). This requires fast recognition of the user utterance as a backchannel (and not an interruption, discussed below). This would enable the system to simply ignore such utterances. But such a strategy would disregard observation (D1), namely that these utterances nevertheless are *reactions* to the content of the system's own ongoing utterance. It seems more appropriate, then, to take into account the role `BC` plays for *grounding*, the process of reaching a common understanding of the ongoing dialogue (Clark and Schaefer, 1989; Clark, 1996). At the very least, `BC`s signal 'continued attention', and it may be useful to represent this fact in the system (perhaps in order to draw inferences from the *absence* of such signals). If this effect is to be modelled, then timing information becomes important, in order to determine to which parts of the utterance the `BC` may be reacting.

Interruptions Concurrent speech from the user that is not classified as a `BC` should be treated as an interruption. Again, there are various degrees of sophistication possible in reactions to interruptions. A sensible default behaviour perhaps is to simply stop talking. Ideally, a system would also be informed where exactly, after which parts of the own utterance, the interruption occurred. However, an interruption need not necessarily lead to a turn-change: in certain situations, a system may be interested in trying to hold the turn and to continue talking.

Turn-Taking One of the immediately striking features of human–human dialogue is that transitions between speakers are often seamless, with very little gap or overlap ((Jaffé and Feldstein, 1970; Sacks et al., 1974; Beattie and Barnard, 1979)). Such seamless transitions can obviously not be achieved with the segmentation model sketched above, which relies on gaps to determine whether a speaker wants to release the turn. Other cues within the utterance must hence be responsible for the other being able to determine when to take the turn. (This is why we include this under "sub-utterance" phenomena here, given our

Phenomenon	Example
Hesitation (HES)	A: From Boston *uhm* on Monday.
Backchannel (BC)	A: From Boston on Monday. B: Mhm
Interruption (INT)	A: From Boston on- B: Sorry, Boston airport is closed!
Turn-Taking (TT)	A: From Boston. B: Erm, hang on, I'll check.
Relevant Non-Linguistic Act (RNLA)	A: From Boston on Monday Sys: [Boston lights up on map]

Table 1: Examples of Sub-Utterance Phenomena in Dialogue

endpointing-based definition of utterance.) There is a rich literature on what exactly the nature of these signals might be, syntax, semantics / pragmatics or prosody (see, *inter alia*, (Ferrer et al., 2002; Ford and Thompson, 1996; Caspers, 2003; Koiso et al., 1998; Sato et al., 2002; Ferrer et al., 2003; Schlangen, 2006)). Assuming that our system could detect such signals, what would be the appropriate reaction? Looking at human–human dialogue, it seems that even in cases where a contentful reply isn't immediately ready, it is a good strategy to acknowledge the obligation to take the turn by producing non-committal "hedges" such as *erm* (Norrick, 2009), as in the TT example in Table 1.

Relevant Non-Linguistic Actions So far, we have restricted the discussion to *linguistic* reactions. In situations where other modalities are available (that is, in face-to-face settings), it can also be appropriate to react non-linguistically to ongoing utterances. An example for such a reaction is shown in Table 1. We will discuss more examples below in Section 4. (One could also subsume under this heading non-*verbal* actions expression functions listed above, such as head-nods as backchannel signals. We however restrict this category here to actions that are non-linguistic in a wider sense.)

With regards to their immediate discourse effect, RNLA are related to (one aspect of) backchannels: they indicate some degree of understanding of the ongoing utterance. In fact, they give much deeper evidence of understanding than BCs, in that they *display* what this understanding is (Clark and Schaefer, 1989). A system capable of registering a user's RNLA should then use them to check whether the displayed understanding is congruent with the intended effect of the utterance-so-far. If it is, that part of the utterance can be taken as understood; if not, corrective measures can be taken, such as providing more information or directly correcting the user's understanding.

2.2 System Conditions Triggering Production of Sub-Utterance Phenomena

We now turn to a discussion of the conditions under which a system might want to produce such phenomena itself.

Hesitations Hesitations in human speech are normally seen as reflecting planning problems (Levelt, 1989; Clark and Fox Tree, 2002), for example due to problems with lexical access. Given current language generation architectures (Theune, 2003) and how they differ to human language generation (for example, with lexical access as an error-free database look-up, no incremental formulation, etc), there doesn't seem to be a natural reason for dialogue systems to produce hesitations. It is an interesting, but to our knowledge unresearched question whether *simulating* such problems could have interactional benefits—one could speculate that inferences from the fact that production is disfluent (as mentioned above; a disfluent description might be of a hard-to-name object) could be usefully triggered in this way.

Backchannel Utterances The situation is different for backchannel utterances. Systems working in more conversational settings may well profit from being able to produce backchannel utterances. Mirroring what was said above about the interpretation of BCs, their production should ideally reflect a desire to signal the grounding status (as 'acoustically perceived') of material concurrently to the continuation of the utterance. Alternatively, one could tie the production of BCs closer to the user's utterance, assuming that there

are indeed 'backchannel-inviting cues' (Ward and Tsukahara, 2000; Gravano and Hirschberg, 2009); we discuss this approach below.

Interruptions A system that is continuously monitoring the user's input may want to interrupt for the same reasons that a human might do so: to be able to immediately address some parts of the utterance (for example challenging its truth), or to make a choice in a long list of alternatives. Another reason for interrupting the user can be that some other event occurs that requires notification of the user with high priority; this could happen in applications where the system controls and monitors some real-world objects (Boye et al., 2000; Lemon et al., 2002). As such interruptions are thematically unrelated to the user utterance, they can be performed without continuous understanding of the user utterance.

Turn-Taking Ideally, a dialogue system would plan its utterances in such a way that turn-endings can be projected easily by the user. As discussed above, there is a variety of candidate cues that may be appropriate here; one that may be in reach is the variation of prosodic structure (e.g., using rising pitch to indicate non-finality; see e.g. (Caspers, 2003)). Moreover, devices for preparing for longer turns could be used ("Let me list the possible options.").

Relevant Non-Linguistic Actions Together with BCs, RNLAs form the class of production behaviours that seem most promising for systems in the near-term. If a modality other than speech is available, it may be advantageous to use it for displaying understanding. We will discuss examples below.

3 Two Approaches

Having surveyed the phenomena, we now turn to describing two possible directions for changes to the current dialogue system architecture model, which relies on full-utterance-based processing. Our question will be to what extent these changes bring the phenomena into the reach of the dialogue systems—with the underlying assumption that making systems capable of handling these phenomena will make them more natural (Ward et al., 2005; Edlund et al., 2008). Before we do so, we need to say just a little bit more about one component of dialogue systems, the dialogue manager. We define this as that component (or collection of components) which a) computes appropriate updates to the context given the new incoming material (which in traditional systems is the endpointed utterance in one block), and b) computes the system reaction to this update. We can then characterise the two directions according to what they modify. In the *reactive approach*, the dialogue manager still works on full utterances only, but does not remain the only place that computes reactions; rather, a further component is added that works on sub-utterance information and controls some system actions. In the *incremental DM approach*, the dialogue manager remains the only module responsible for computing system behaviour, but the basis on which it does so is changed, by allowing updates triggered by units smaller than the utterance.

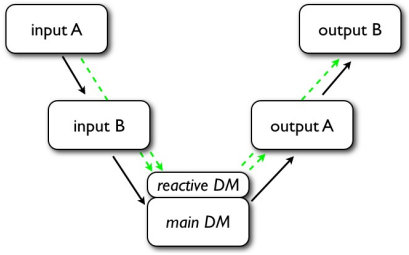


Figure 1: Schematic view of the architectural choices

Figure 1 gives a schematic view of the two possibilities: in a system with a reactive layer, continuous information flows (along the dashed lines) from "input-side" modules (the boxes on the left; let's say speech recognition and natural language understanding module) to a separate reactive DM module. This module can decide independently on the need for 'reactive behaviour', e.g. BCs. Meanwhile, the usual utterance-sized information flows along the normal channel (the solid lines), from the input modules into the dialogue manager, which computes system reactions concerning the "official business" of the dialogue as in normal systems. In a fully incremental system, on the other hand, continuous information can travel along the normal pathway, and the main DM itself is reactive enough to decide on reactions like BCs.

3.1 Keeping the Dialogue Manager, Adding a Reactive Layer

The strategy of adding a 'reactive layer' to an otherwise largely unchanged dialogue system architecture is the more prominent in the literature;

despite differences in detail, (Thórisson, 2002; Lemon et al., 2003; Raux and Eskenazi, 2007) can all be categorised in this way. It is clearly an attractive strategy, as it allows one to keep tried-and-tested traditional dialogue management paradigms; we will explore here to what extent it can cover the behaviours listed in the previous section.

Hesitations We have explained above that hesitations pose a big problem for systems that rely on silence thresholding for determining the end of a user contribution. A system with a reactive layer that has continuous access to more information than just voice activity can improve on this. Raux and Eskenazi (2008) describe such a system, where continuous information from a voice activity detector is combined with continuous information from a language understanding component that works on hypotheses of what was said so far. Using a simple proxy for detecting semantic completeness ("are expected slots already filled?"), their system can classify silences and use optimised thresholds, overall improving the latency of the replies. Note that the architectural changes only concern the additional layer; once the endpointing decision is made, all further updates and computations of system reactions are made by the unchanged dialogue manager.

Backchannel Utterances We've described above as one possible reaction to a user backchannel utterance to simply ignore them. For this, a system with a reactive layer would need the capability to quickly classify incoming audio from the user during a system utterance as backchannel or genuine interruption. A backchannel utterance could then simply be withheld from the rest of the system. (Note that this strategy entails that none of the discourse effects of BCs described above can be modelled.) We are not aware of any implemented system that makes use of such a strategy.

On the production side, a reactive layer could decide to produce a backchannel utterance in response to so called "backchannel-inviting cues" (Ward and Tsukahara, 2000; Gravano and Hirschberg, 2009), which are prosodic and lexical features of the utterance. (Their presence could presumably be detected on the basis of continuous information comparable to what was described in the previous section.) (Beskow et al., 2009) have shown that it is indeed possible, at least for short whiles, to plausibly accompany user speech with BCs produced as reaction to such cues. However, there is a danger in decoupling BCs from actual grounding state. The reactive layer and the main dialogue manager can get out of sync, as illustrated in (1) (constructed; system is B), a situation where the reactive layer produced BCs and hence signalled at least some form of understanding, which however isn't backed up by the main dialogue manager, which requests clarification. This shows that some form of synchronisation is needed in such an architecture, if behaviours produced by the reactive layer commit the system publicly to a certain discourse state.

(1)
A: Take the green block
B: uhu
A: and place i:t in the
B: yeah?
A: middle of the board.
B: OK.
B: I'm sorry, what did you say?

Interruptions, Turn-Taking From the perspective of a system with a reactive layer, these two phenomena are flip-sides of being able to deal with BCs and HES, respectively: if those phenomena are classified correctly, appropriate reactions to these phenomena can be made as well. For INT, that could be to stop talking (but again presumably losing information about what did get said), for TT this would be to start talking, or, if nothing is prepared, to produce a turn-initial non-committal signal like "erm".

Relevant Non-Linguistic Actions This seems to be an area where just adding a reactive layer cannot help. There must be a way to compute the relevance of such an action to what was already said, and for this, it seems, proper context updates have to be performed on such partial inputs. This is what the kind of architecture we turn to next promises to be able to do.

3.2 Incrementalising the Dialogue Manager

With some reflection, it should be clear that a system with an incremental dialogue manager (perhaps fed with more than just output of what is typically the previous module, NLU) can do at least that what was described above for the reactive-layer approach, as it is a proper superset. As such an approach has, to our knowledge, not been described in the literature, the more interesting ques-

tion is whether something like this is actually practically possible. (There is important prior work: (Allen et al., 2001) describe a general architecture for dialogue systems that somewhat falls under this heading; however, the focus there is more on architecture and no general DM strategy is described. Similarly, (Skantze and Schlangen, 2009) describe an implemented system that to some extent realises incremental DM, but again, the DM strategy is not the focus of that paper.)

We hence will not go through the list of phenomena again but instead devote the remaining space to sketching what a plausible incremental dialogue management approach could look like. Before we turn to this, we should note that a precondition of such an approach is that the modules feeding into the dialogue manager also work incrementally; recent work suggests that this precondition can be met (incremental ASR (Baumann et al., 2009); incremental NLU (Atterer and Schlangen, 2009; Schlangen et al., 2009; Atterer et al., 2009); generation (Kilger and Finkler, 1995; Otsuka and Purver, 2003)).

4 Sketch of an Incremental Dialogue Manager

For concreteness, we set ourselves the task here to model BC and RNLA behaviour with an incremental dialogue manager. What is needed to achieve this? First, a context representation that can be updated with partial information and that tracks grounding state (and how it is influenced by performing BCs and also RNLAs), and second, rules that compute when to perform these behaviours.

Perhaps surprisingly, it turns out that not very many deep conceptual changes are needed to get this from extant dialogue modelling paradigms. (2) shows the structure of the plan of an information state update-based system to provide ticket price information (from (Larsson, 2002, p.52)).

(2)
ISSUE: $?x.\textbf{price}(x)$
PLAN: $\langle$
findout($?x.\textbf{means\_of\_transport}(x)$),
findout($?x.\textbf{dest\_city}(x)$),
findout($?x.\textbf{depart\_city}(x)$),
findout($?x.\textbf{depart\_month}(x)$),
findout($?x.\textbf{depart\_day}(x)$),
findout($?x.\textbf{depart\_class}(x)$),
consultDB($?x.\textbf{price}(x)$)
$\rangle$

The items in this plan are questions that the system must get answered in order to handle the issue of providing price information. Raising them in the form of a sequence of questions, however, is what leads to the typical slightly rigid structure characteristic of dialogues with enquiry-based systems ("how do you want to travel?", "Where do you want to go?", "When?", etc. etc.). The assignment of each bit of necessary information to a separate question, and hence a separate user reply, appears somewhat unnatural; and this is indeed is reflected by the often made observation that human users tend to react to such restricted questions with what in the field of dialogue system design is called *overanswering*, that is by providing more information than the question taken on its own asks for (McTear, 2004).

If we remove the direct connection between the questions in the plan and expected utterances, and allow the user to address more than one of those question within a single utterance, and without them having been raised explicitly, we have made a first step towards an incremental dialogue manager. We then also note that the utterance bits answering individual questions (e.g. "*I want to fly...*", "*...from Amsterdam...*", "*...on Monday*") seem like good candidates for chunks that can be acknowledged by BCs.

Figure 2 shows an example of the information state format used in a system we are currently building.[2] The domain of the system is constructing a puzzle, where the user controls the computer to select and move around pieces, getting immediate visual feedback. The structure shown is our equivalent of a Question-Under-Discussion stack (Ginzburg, 1996), and is to be read as follows. In angle brackets, all information is grouped together that the system needs to collect to perform one domain action. Here, we have the actions take and delete; the curly brackets indicate that they are alternatives. I.e., the system needs to collect information about which action to perform, and about which tile to perform it on. (Here, both actions have the same parameter, but that is just a coincidence). After the first semicolon, we specify what an appropriate RNLA is when the information chunk specified in this line has been provided. E.g., once we know that the action to perform is

[2] We should note that we are in the early stages of the realisation of the system. While first experiments indicate that the concepts sketched here should work, the devil will, as always, be in the details of the implementation.

```
{< a ( 1 action=A=take; 2 prepare(A)   ; 3 U),
     ( 4 tile=T       ; 5 highlight(T) ; 6 U),
     ( 7              ; 8 execute(A,T) ; 9 U) >

 < b (10 action=A=del ;11 prepare(A)   ;12 U),
     (13 tile=T       ;14 highlight(T) ;15 U),
     (16              ;17 execute(A,T) ;18 U) >}
```

Figure 2: Example Information State

take, we can prepare for this action. The last column in each row records the resolution/grounding state: has this question / bit of information been resolved / provided? Has the provided value been grounded with the user? Etc. The last line finally records what to do when all bits of information have been collected.

The idea now is that users can provide this information (initially) unprompted and within one (or more) utterance(s), and that it is "struck out" immediately once provided. Additionally, both BC and RNLA feedback can be given during the utterance. In the appendix, we give two worked examples that illustrate some nice consequences of this setup: replies to RNLAs mid-utterance (e.g., after highlighting a piece, a "right" and then a continuation of the utterance), and delivery in installments with trial intonation (Clark, 1996) can be modelled.

5 Conclusion

We have surveyed what we call "sub-utterance phenomena", that is, phenomena that require processing in a dialogue system of units smaller than full utterances. We have discussed two possible approaches to such processing, namely either providing a parallel structure that takes care of some reactive behaviours, or else making the system as a whole more responsive and able to process small units of user input. We should stress that we do not necessarily favour one approach over the other. If the emphasis is on keeping a legacy dialogue model, then adding a reactive layer is a good way to increase reactivity. If however the emphasis is on full semantic modelling, we think that an incremental dialogue manager may have some advantages—despite being a paradigm that clearly needs more work to be fully understood.

Acknowledgements
This work was funded by DFG grant SCHL845/3-1, Emmy Noether Programme.

References

Gregory Aist, James Allen, Ellen Campana, Carlos Gomez Gallo, Scott Stoness, Mary Swift, and Michael K. Tanenhaus. 2007. Incremental understanding in human-computer dialogue and experimental evidence for advantages over nonincremental methods. In *Proceedings of Decalog (Semdial 2007)*, Trento, Italy.

James Allen, George Ferguson, and Amanda Stent. 2001. An architecture for more realistic conversational systems. In *Proceedings of the conference on intelligent user interfaces*, Santa Fe, USA, June.

Jennifer E. Arnold, Carla L. Hudson Kam, and Michael K. Tanenhaus. 2007. If you say *Thee uh* you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5):914–930.

Michaela Atterer and David Schlangen. 2009. RUBISC – a robust unification-based incremental semantic chunker. In *Proceedings of SRSL 2009*, Athens, Greece, March.

Michaela Atterer, Timo Baumann, and David Schlangen. 2009. No sooner said than done? testing incrementality of semantic interpretations of spontaneous speech. In *Proceedings of Interspeech 2009*, Brighton, UK, September.

Karl G D Bailey and Fernanda Ferreira. 2007. The processing of filled pause disfluencies in the visual world. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. l. Hill, editors, *Eye Movements: A Window on Mind and Brain*, pages 485–500. Elsevier.

Timo Baumann, Michaela Atterer, and David Schlangen. 2009. Assessing and improving the performance of speech recognition for incremental systems. In *Proceedings of NAACL-HLT 2009*, Boulder, Colorado, USA, May.

G. W. Beattie and P. J. Barnard. 1979. The temporal structure of natural telephone conversations. *Linguistics*, 17:213–229.

Jonas Beskow, Rolf Carlson, Jens Edlund, Björn Granström, Mattias Heldner, Anna Hjalmarsson, and Gabriel Skantze. 2009. Multimodal interaction control. In Alexander Waibel and Rainer Stiefelhagen, editors, *Computers in the Human Interaction Loop*, pages 143–158. Springer, Berlin, Germany.

Johan Boye, Beth Ann Hockey, and Manny Rayner. 2000. Asynchronous dialogue management: Two case-studies. In *Proceedings of the 4th Workshop on Semantics and Pragmatics of Dialogue (Götalog2000)*, pages 51–55, Gothenburg, Sweden.

Susan E. Brennan and Michael F. Schober. 2001. How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44:274–296.

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Amsterdam, The Netherlands.

Johanneke Caspers. 2003. Local speech melody as a limiting factor in the turn-taking system in dutch. *Journal of Phonetics*, 31:251–276.

Herbert H. Clark and Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84:73–111.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

David DeVault and Matthew Stone. 2003. Domain inference in incremental interpretation. In *Proceedings of ICOS*

4: Workshop on Inference in Computational Semantics, Nancy, France, September. INRIA Lorraine.

Jens Edlund, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson. 2008. Towards human-like spoken dialogue systems. *Speech Communication*, 50:630–645.

Luciana Ferrer, Elizabeth Shriberg, and Andreas Stolcke. 2002. Is the speaker done yet? Faster and more accurate end-of-utterance detection using prosody. In *Proceedings of ICSLP2002*, Denver, USA, September.

Luciana Ferrer, Elizabeth Shriberg, and Andreas Stolcke. 2003. A prosody-based approach to end-of-utterance detection that does not require speech recognition. In *Proceedings of ICASSP-2003*, Hong Kong, China.

Cecilia E. Ford and Sandra A. Thompson. 1996. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E.A. Schegloff, and S.A. Thompson, editors, *Interaction and Grammar*, pages 134–184. CUP, Cambridge, UK.

Jonathan Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.

Agustín Gravano and Julia Hirschberg. 2009. Backchannel-inviting cues in task-oriented dialogue. In *Proceedings of Interspeech 2009*, pages 1019–1022, Brighton, UK, September.

Joseph Jaffé and Stanley Feldstein. 1970. *Rhythms of Dialogue*. Academic Press, New York, NY, USA.

Anne Kilger and Wolfgang Finkler. 1995. Incremental generation for real-time applications. Technical Report RR-95-11, DFKI, Saarbrücken, Germany.

Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, and Yasuharu Den. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map-task dialogs. *Language and Speech*, 41(3–4):295–321.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University, Göteborg, Sweden.

Oliver Lemon, Alexander Gruenstein, Alexis Battle, and Stanley Peters. 2002. Multi-tasking and collaborative activities in dialogue systems. In *Proceedings of SIGDIAL 2002*, Philadelphia, USA, July.

Oliver Lemon, Lawrence Cavedon, and Barbara Kelly. 2003. Managing dialogue interaction: A multi-layered approach. In Alexander Rudnicky, editor, *Proceedings of SIGdial 2003*, pages 168–177, Sapporo, Japan, July.

Willem J.M. Levelt. 1989. *Speaking*. MIT Press, Cambridge, USA.

W. D. Marslen-Wilson and L. K. Tyler. 1981. Central processes in speech understanding. *Philosophical Transactions of the Royal Society London*, B295:317–332.

Michael F. McTear. 2004. *Spoken Dialogue Technology*. Springer Verlag, London, Berlin.

Neal R. Norrick. 2009. Interjections as pragmatic markers. *Journal of Pragmatics*, 41(5):866–891.

Masayuki Otsuka and Matthew Purver. 2003. Incremental generation by incremental parsing. In *Proceedings of the student conference "Computational Linguistics in the UK"*, Edinburgh, UK, January.

Massimo Poesio and Hannes Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1(1):1–89.

Massimo Poesio and David Traum. 1997. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347.

Antoine Raux and Maxine Eskenazi. 2007. A multi-layer architecture for semi-synchronous event-driven dialogue management. In *Proceedings of ASRU 2007*, Kyoto, Japan.

Antoine Raux and Maxine Eskenazi. 2008. Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. In *Proceedings of the 9th SIGdial Workshop in Discourse and Dialogue*, Columbus, Ohio, USA.

H. Sacks, E. A. Schegloff, and G. A. Jefferson. 1974. A simplest systematic for the organization of turn-taking in conversation. *Language*, 50:735–996.

Ryo Sato, Ryuichiro Higashinaka, Masafumi Tamoto, Mikio Nakano, and Kiyoaki Aikawa. 2002. Learning decision trees to determine turn-taking by spoken dialogue systems. In *Proceedings of ICSLP-2002*, pages 861–864, Denver, USA, September.

Emanuel A. Schegloff. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen, editor, *Analyzing Discourse: Text and Talk*, pages 71–93. Georgetown University Press, Washington, D.C., USA.

David Schlangen, Timo Baumann, and Michaela Atterer. 2009. Incremental reference resolution: The task, metrics for evaluation, and a bayesian filtering model that is sensitive to disfluencies. In *Proceedings of SIGdial 2009*, London, UK, September.

David Schlangen. 2006. From reaction to prediction: Experiments with computational models of turn-taking. In *Proceedings of Interspeech 2006, Panel on Prosody of Dialogue Acts and Turn-Taking*, Pittsburgh, USA, September.

Gabriel Skantze and David Schlangen. 2009. Incremental dialogue processing in a micro-domain. In *Proceedings of EACL 2009*, pages 745–753, Athens, Greece, March.

Michael K. Tanenhaus, Michael J. Spivey-Knowlton, Kathllen M. Eberhard, and Julie C. Sedivy. 1995. Intergration of visual and linguistic information in spoken language comprehension. *Science*, 268.

M. Theune. 2003. Natural language generation for dialogue: system survey. Technical Report TR-CTIT-03-22, Centre for Telematics and Information Technology, University of Twente, Enschede.

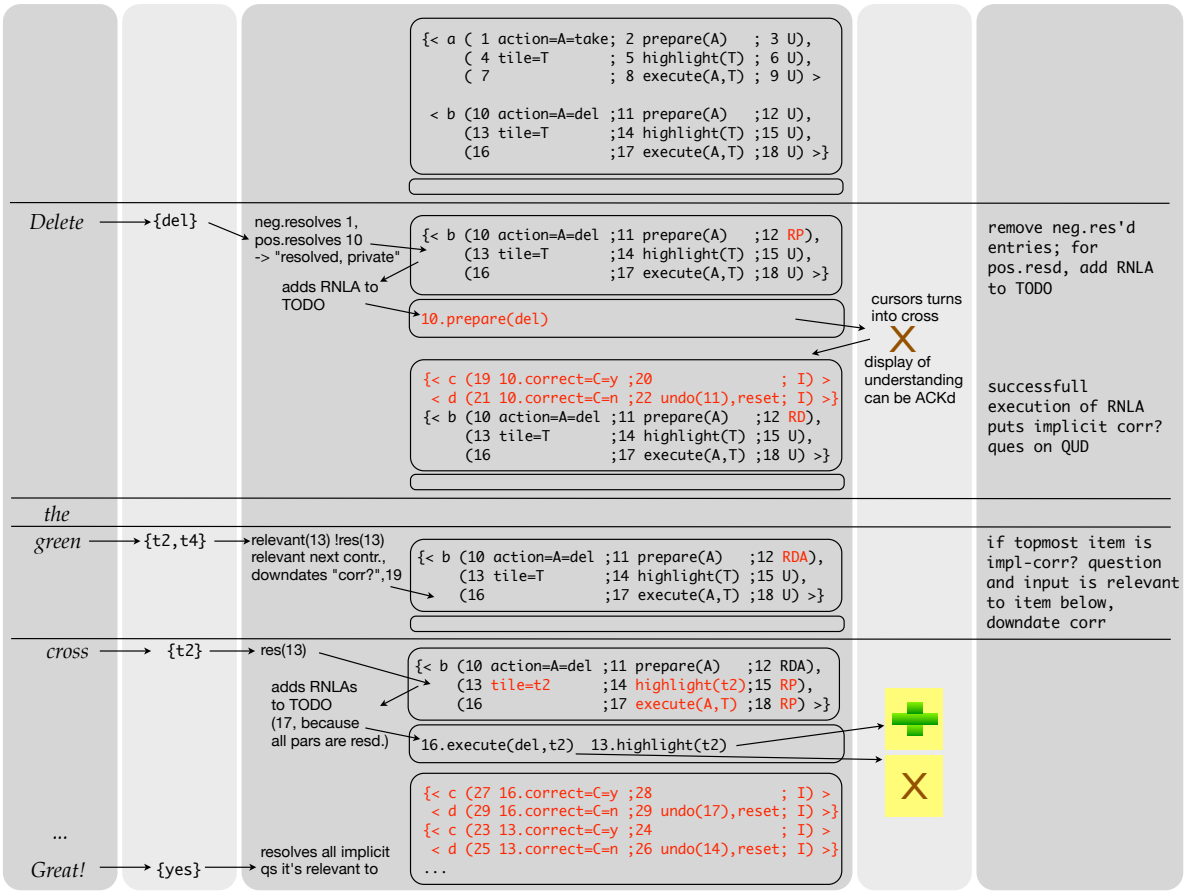
Kristinn R. Thórisson. 2002. Natural turn-taking needs no manual: computational theory and model, from perception to action. In Björn Granström, David House, and Inger Karlsson, editors, *Multimodality in Language and Speech Systems*, pages 173–207. Kluwer, Dordrecht, The Netherlands.

Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel repsonses in english and japanese. *Journal of Pragmatics*, 32:1177–1207.

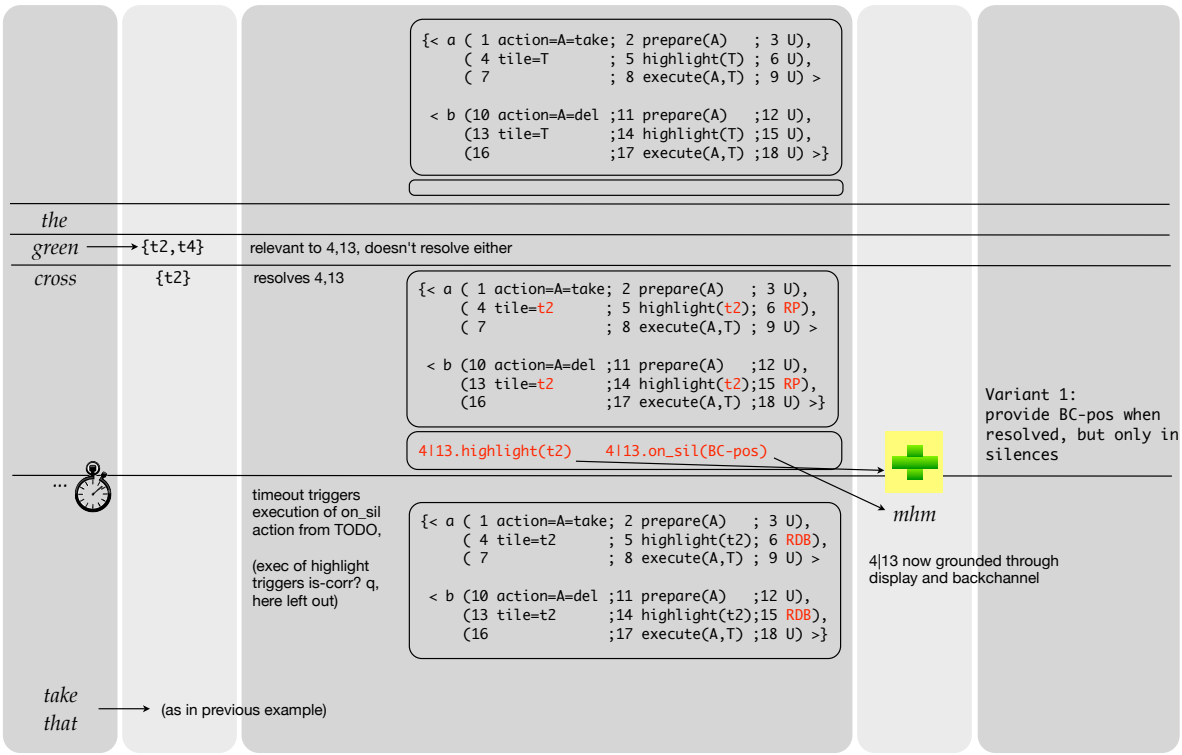
Nigel G. Ward, Anais G. Rivera, Karen Ward, and David G. Novick. 2005. Root causes of lost time and user stress in a simple dialog system. In *Proceedings of the 9th European Conference on Speech and Communication Technology (Interspeech2005)*, Lisbon, Portugal, September.

V. H. Yngve. 1970. On getting a worde in edgewise. In *Papers from the 6th Regional Meeting*, pages 567–578, Chicago, USA. Chicago Linguistics Society.

APPENDIX



Example 1. Columns are, from left to right: user utterance, semantics, updates and resulting information state (consisting of QUD and TO-DO field), system reaction, and comments. Utterance of *delete* eliminates other candidate (*take*) from QUD, triggers visible action (cursor turns into cross), which implicitly raises question "was this correct?". Question is answered by relevant continuation ("the green. . . "), and hence removed.



Example 2. Columns as above. "*the green cross*" is relevant to both action alternatives (*take* and *delete*). Is uttered with rising pitch (trial intonation), which leads to short timeout, at which a confirming BC is uttered.
Resolution / grounding states shown: U, unresolved; RP, resolved, but private (not grounded); I, implicitly raised; RD, resolved, understanding displayed; RDB, resolved, displayed and BC offered.

Figure 3: Worked Examples

Splitting the "*I*"s and Crossing the "*You*"s: Context, Speech Acts and Grammar

Matthew Purver
Department of Computer Science,
Queen Mary University of London
m.purver@qmul.ac.uk

Eleni Gregoromichelaki, Wilfried Meyer-Viol
Department of Philosophy, King's College London
eleni.gregor@kcl.ac.uk
wilfried.meyer-viol@kcl.ac.uk

Ronnie Cann
Linguistics and English Language, University of Edinburgh
r.cann@ed.ac.uk

Abstract

The occurrence of split utterances (SUs) in dialogue raises many puzzles for grammar formalisms, from formal to pragmatic and even philosophical issues. This paper presents an account of some of the formal details that grammars need to incorporate in order to accommodate them. Using *Dynamic Syntax* (DS), we illustrate how by incorporating essential interaction with the context into the grammar itself, we can deal with speaker change in SUs: not only its effects on indexicals *I* and *you*, but also the multiple illocutionary forces that can arise. We also introduce a *Split Turn Taking Puzzle* (STTP) showing that the current speaker and the agent of the resulting speech act are not necessarily the same.

1 Introduction

Split utterances (SUs) – utterances split across more than one speaker or dialogue contribution – are common in spontaneous conversation and provide an important source of data that can be used to test the adequacy of linguistic theories (Purver et al., 2009). Previous work has suggested that Dynamic Syntax (DS) (Kempson et al., 2001) is well placed to analyse these phenomena as it is strictly word-by-word incremental, allowing an account of speaker changes at any point (Purver et al., 2006) and interruptive phenomena such as mid-sentence clarification sequences (Gargett et al., 2009). However, other less incremental grammar formalisms have also been applied to particular kinds of SUs: Poesio and Rieser (2010) use LTAG[1] in an analysis of *collaborative completions*. But given that such accounts employ grammars which license *strings of words*, a direct account for SUs is prevented, as the reference and binding of indexical speaker/addressee pronouns changes as the speaker transition occurs:

(1) A: Did you give me back
B: your penknife? It's on the table.

(2) A: I heard a shout. Did you
B: Burn myself? No, luckily.

A grammar must rule out the sentence *did you burn myself?* as ungrammatical if spoken by one single speaker, but allow it as grammatical if the identity of the speaker changes as in (2) – this will be hard for a string-based account. In (1), *you* and *your* must be able to take different referents due to the speaker change between them. In contrast, as DS defines grammatical constraints in terms of the incremental construction of semantic content (rather than through licensing strings via an independent layer of syntax over strings), we show that such examples are not problematic given an independently motivated definition of the lexical entries for indexicals.

SUs can also perform diverse dialogue functions, with the speech acts associated with the individual speakers' contributions often being different – see (Purver et al., 2009). In (1)-(2), B's continuations seem to function as clarifications of A's intended queries. Others have pointed out that continuations can function as e.g. adjuncts (Fernández and Ginzburg, 2002) or clarification requests (Purver et al., 2003); and Poesio and Rieser (2010) show how a completion (in their terms) can get its function in the dialogue (in their case, to act as a collaborative contribution to a plan). A full account of SUs therefore requires some representation of such dialogue function information in the model of context that guides dialogue interpretation and production.

[1] Unlike DS, LTAG must be supplemented by a parsing and/or generation model (a set of defeasible inference rules for Poesio and Rieser) to derive the incrementality required.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

DS, however, currently incorporates no notion of illocutionary force or dialogue act type, as it is assumed that derivation of such information is not linguistically determined. In the case of SUs, it has been assumed that the *grammar* itself provides adequate means of continuing/taking over somebody else's utterance, and that this does not *necessarily* involve strategic reflection or fully-formed intentions as to what function the utterance should perform: this provides the possibility for speakers to 'blurt out' utterances without necessarily having any specific plans/intentions in mind, and for hearers to respond without reflection as to the speaker's plan. But, as pointed out in (Kempson et al., 2007; Gargett et al., 2009), this is not an in-principle objection to the specification of speech act information as part of the representation derived by the parse of an utterance, as DS provides mechanisms for allowing the inclusion of optional inferred information. We present here an extension to DS which allows it to include such information explicitly and draw the distinctions relevant for SUs.

We also show how this extension is motivated by the resolution of the *Split Turn Taking Puzzle* (STTP). This is a version of Ginzburg (1997)'s Turn Taking Puzzle applied to SUs, where it appears that distinct empirical results are obtained: given a SU split between two people, the possible interpretations of a subsequent "Why?" depend not on the most recent speaker, but on who can be taken as the agent of the speech act performed – which may be distinct from the notion of 'speaker' tracked by indexical pronouns like *I* and *my*.

2 Combining Dynamic Syntax with TTR

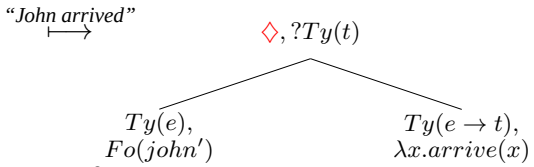
2.1 Dynamic Syntax (DS)

DS combines word-by-word incrementality with context-dependent, goal-directed parsing defined over partial trees. Importantly, these trees are semantic objects, rather than reflecting syntax or word order. Parsing in DS relies on the execution of licensed *actions*, as incorporated in lexical entries (as in (3) below); such actions resolve outstanding requirements (here, $?Ty(e)$) to decorate the tree with information about semantic type Ty and content (formula) Fo:

(3)

john:	
IF	$?Ty(e)$
THEN	put($Ty(e)$)
	put($Fo(john')$)
ELSE	abort

(4)



Application of lexical actions is interspersed with the execution of computational rules which provide the predictive element in the parse and provide the compositional combinatorics. For example, eventual type deduction and function application is achieved by means of the rule of *Elimination* (5). This derives the value of a mother node's semantic type Ty and content Fo from that of its daughters, in (4) providing the values $Ty(t), Fo(arrived(john))$ at the top node:

(5)

Elimination:	
IF	$?Ty(T_1)$,
	$\downarrow_0 (Ty(T_2), Fo(\alpha))$
	$\downarrow_1 (Ty(T_2 \rightarrow T_1), Fo(\beta))$
THEN	put($Ty(T_1)$)
	put($Fo(\beta(\alpha))$)
ELSE	abort

Grammaticality is then defined in terms of a resulting complete (requirement-free) tree. Generation is defined in terms of parsing, and therefore also functions with partial trees, uses the same action definitions, and has the same context-dependence, incrementality and predictivity.

DS is thus well-placed to account for SUs: equal incrementality in parsing and generation, and the use of the same partial tree representations, allows the successful processing of "interruptive" SUs with speaker changes at any point. As *goal trees* (planned messages driving generation) may also be partial, utterances may be produced before a total propositional message has been constructed, and completions may be analysed without necessarily involving "guessing". The parser-turned-producer has just to access a word that seems to them an appropriate completion, without *necessarily* considering whether it matches the previous speaker's intention.

DS doesn't incorporate a notion of dialogue act type (in contrast to e.g. Ginzburg et al. (2003)) as it is assumed that the linguistically provided information is highly underspecified, namely just an indication of sentence mood as declarative, interrogative, imperative.[2] However, as the DS formalism

[2]Such specifications are currently encoded as features translatable into use-neutral procedural instructions, unless there are "grammaticised" associations between moods and speech acts, an empirical issue to be decided on a language-by-language basis.

is designed to interact with context incrementally at any point, the possibility of deriving speech act information from context exists; although the interface to enable this must be specified. For this reason, we now turn to TTR, a transparent representation format allowing the specification and interaction of multiple types of information.

2.2 Type Theory with Records (TTR)

TTR has already been used in dialogue modelling (Cooper and Ginzburg, 2002; Ginzburg, forthcoming). Tokens (*records*) and types (*record types*) are treated uniformly as structured representations – sequences of $label : type$ pair *fields* – with the result that their interaction can be modelled in a single system, as required when dealing with meta-communicative uses of language such as 'repair'-constructions or grounding.

Here, the attraction of TTR is that it allows the stratification of multiple types of information, using distinct field labels. The device of *dependent types* allows linking of information between fields, as types can depend on types occurring earlier in the record (higher up in the graphical representation). This allows us to separate contextual information (e.g. information about conversational events, including speaker, addressee, time, location etc.) from the semantic content directly derived from the linguistic string, but allow interaction between the two; this is what we need for phenomena like resolution of ellipsis or assigning values to indexicals and anaphoric elements.

2.2.1 Using TTR in DS

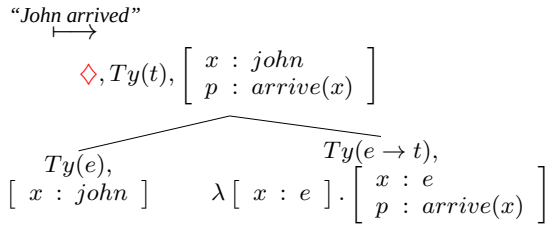
TTR has not, however, been defined in an incremental manner.[3] Here, then, we use TTR representations within the DS vocabulary of trees and actions, replacing the unstructured content of the $Fo()$ labels with TTR record types, and interpreting $Ty()$ simple type labels (and requirements) as referring to *final* TTR field type. Compare the modified lexical entry and eventual tree representation below with the ones displayed in (3)-(4):

(6)

john:	
IF	$?Ty(e)$
THEN	put($Ty(e)$)
	put($\left[\; x \; : \; john \; \right]$)
ELSE	abort

[3]Work is underway to introduce incrementality in the TTR model via the subtyping relation (White (in prep); Meyer-Viol (in prep)). Here we pursue a more conservative strategy.

(7)



Function application and type deduction will now apply under a suitably modified rule of the DS *Elimination* process; see (9) below.

3 Utterance Events

An account of SUs must explain how indexical pronouns can assume distinct values around a change of speaker (and addressee). We therefore require some record of the utterance event/situation which includes information about speaker/addressee identity. Note that the availability of utterance events to the semantics is independently motivated by e.g. event reference via anaphora ("what do you mean by *that*?") (see also Poesio and Rieser 2010).

We assume that utterance events should at minimum record participant information and who is uttering which particular word(-string). We therefore introduce a partition within the TTR representation of content, with utterance event information held in a *context* (or $ctxt$) field, and linguistically derived semantic content in a *content* (or $cont$) field. The $ctxt$ field is itself structured, containing the required information about utterance event, speaker and addressee; we assume this is available directly from the real-time context of utterance:[4]

(8)
$$\left[\begin{array}{ll} a & : participantA \\ b & : participantB \\ u & : utt - event \\ s_s & : spkr(u, a) \\ s_a & : addr(u, b) \end{array} \right]$$

In a fuller treatment, this utterance context information should also include further information such as time of utterance, world etc, but we omit these here for simplicity.

The DS *Elimination* process must now perform beta-reduction (as before) for the $cont$ field, and TTR extension (i.e. concatenation (Cooper, 1998), shown here as $\oplus$) for the $ctxt$ field, as shown in (9), (10). Parsing a two-word utterance *John arrived* spoken by one speaker, A, will therefore now result in a representation as in Fig 1.

[4]This is a simplification, of course: determination of addressee is not trivial – see (Goffman, 1981) amongst others.

Figure 1: Tree structure derived from *John arrived* spoken by a single speaker $participantA$

$$\Diamond, Ty(t), \left[\begin{array}{ll} ctxt: & \left[\begin{array}{ll} a & : participantA \\ u_0 & : utt-event \\ s_{s0} & : spkr(u_0, a) \\ u_1 & : utt-event \\ s_{s1} & : spkr(u_1, a) \end{array}\right] \\ cont: & \left[\begin{array}{ll} x & : john \\ p & : arrive(x) \end{array}\right] \end{array}\right]$$

$$\left[\begin{array}{ll} Ty(e), & \\ ctxt: & \left[\begin{array}{ll} u_0 & : utt-event \\ s_{s0} & : spkr(u_0, a) \end{array}\right] \\ cont: & \left[\begin{array}{l} x : john \end{array}\right] \end{array}\right] \quad \left[\begin{array}{ll} Ty(e \to t), & \\ ctxt: & \left[\begin{array}{ll} u_1 & : utt-event \\ s_{s1} & : spkr(u_1, a) \end{array}\right] \\ cont: & \lambda[x].\left[\begin{array}{l} p : arrive(x) \end{array}\right] \end{array}\right]$$

(9) *Elimination:*

IF $?Ty(T_1)$,
$\downarrow_0 (Ty(T_2), \left[\begin{array}{ll} ctxt: & c_1 \\ cont: & \alpha \end{array}\right])$
$\downarrow_1 (Ty(T_2 \to T_1), \left[\begin{array}{ll} ctxt: & c_2 \\ cont: & \beta \end{array}\right])$
THEN $\texttt{put}(Ty(T_1), \left[\begin{array}{ll} ctxt: & c_1 \oplus c_2 \\ cont: & \beta(\alpha) \end{array}\right])$
ELSE $\texttt{abort}$

(10)

$$\Diamond, Ty(t), \left[\begin{array}{ll} ctxt: & c_1 \oplus c_2 \\ cont: & \left[\begin{array}{ll} x & : e \\ p & : f(x) \end{array}\right] \end{array}\right]$$

$$\left[\begin{array}{ll} Ty(e), & \\ ctxt: & c_2 \\ cont: & \left[\begin{array}{l} x : e \end{array}\right] \end{array}\right] \quad \left[\begin{array}{ll} Ty(e \to t), & \\ ctxt: & c_1 \\ cont: & \lambda\left[\begin{array}{l} x : e \end{array}\right].\left[\begin{array}{ll} x & : e \\ p & : f(x) \end{array}\right] \end{array}\right]$$

3.1 Indexical Pronouns

Importantly, the definitions of TTR mean that semantic $cont$ information can depend on values in the earlier $ctxt$ context field (although not vice versa). Given this, an explanation of the reference of *I* and *you* becomes expressible. First-person pronouns are defined to take their semantic value from the value of the speaker information in $ctxt$; second-person pronouns from the addressee ($\mathbf{x}$ and $\mathbf{u}$ are rule-level variables binding terms on the nodes where the rules apply).

(11) *I*:
IF $?Ty(e), \left[\ ctxt\ :\ \left[\ s_s : spkr(\mathbf{u}, \mathbf{x})\ \right]\ \right]$
THEN $\texttt{put}(Ty(e))$, $\texttt{put}(Fo(\mathbf{x}))$
ELSE $\texttt{abort}$

(12) *You*:[5]
IF $?Ty(e), \left[\ ctxt\ :\ \left[\ s_a : addr(\mathbf{u}, \mathbf{x})\ \right]\ \right]$
THEN $\texttt{put}(Ty(e))$, $\texttt{put}(Fo(\mathbf{x}))$
ELSE $\texttt{abort}$

As grammatical constraints in DS are phrased in terms of semantic features (rather than syntactic features), the grammaticality of examples like (2) now becomes almost trivial. While a syntactic account would have trouble explaining how *myself* can be co-referential with its antecedent *you*, there is no such problem here: as *you* uttered by A and *myself* uttered by B annotate the trees with co-referential semantic variables. The lexical entries for reflexives such as *myself* must check for a suitably co-referential subject elsewhere in the tree (via the co-argument constraint $\uparrow_0\uparrow_{1*}\downarrow_0 Fo(\mathbf{x})$), and here, this will be available:

(13) *myself*:
IF $?Ty(e), \left[\ ctxt\ :\ \left[\ s_s : spkr(\mathbf{u}, \mathbf{x})\ \right]\ \right]$,
$\uparrow_0\uparrow_{1*}\downarrow_0 Fo(\mathbf{x})$
THEN $\texttt{put}(Ty(e))$, $\texttt{put}(Fo(\mathbf{x}))$
ELSE $\texttt{abort}$

4 Speech acts

Purver et al. (2009) show that SUs are often not straightforward in speech act terms: sometimes they continue/complete the original speech act; sometimes they perform a new one, clarifying/confirming a suggested completion; sometimes they are ambiguous and/or multifunctional. In order to express these important differences, we need the ability to represent and reason about speech act information (see e.g. (Ginzburg et al., 2003; Asher and Lascarides, 2003)).

Importantly, we would like any inferences about speech acts to be *optional*. A parser should enable these inferences when the appropriate function of the turn is at issue (e.g. in cases of 'repair'), but they should not have to be derived for intelligibility or the determination of grammaticality. They should also be derivable retrospectively: as a result of an interlocutor's feedback, one can assign a particular force (even to one's own contribution) that had not occurred to them beforehand.

Any computational rules that introduce such inferences must therefore be available in the grammar but optional (except where the association of a specific construction with a particular interpre-

[5] A more complex set of actions may be required to account for the fact that *you* may be singular or plural in reference, may include the hearer or not and may be generic.

tation has been grammaticised); and the resulting representations should be kept distinct from those derived directly from the parsing of linguistic input. DS already provides a mechanism which suits these requirements: the use of LINKed trees (trees which share some semantic variable), as in the analysis of non-sentential fragments (Gargett et al., 2009) and relative clauses (Kempson et al., 2001). This device of LINKed trees expresses the cognitive reality of distinguished local domains as evinced by standard syntactic tests, e.g. island-constraints and binding restrictions (see e.g. (Gregoromichelaki, 2006)). As TTR currently does not provide the means for such syntax-semantics interface restrictions we retain the notion of LINKed trees here.

As speech act information can be highly underspecified and context-dependent, we do not wish to assume here either a fixed range of speech acts or a fixed set of inferences from linguistic form to speech act type. We therefore take the rules introducing such information to be of the form sketched in (14). When applied, this rule will introduce a new LINKed tree and provide a Fo value $\mathbf{A}(\mathbf{V}, \mathbf{U}, \mathbf{F}(\mathbf{p}))$ where $\mathbf{A}$ is a metavariable ranging over speech act specifications, $\mathbf{V}$ the agent responsible for the speech act, $\mathbf{U}$ an utterance event (or sequence of events), and $\mathbf{F}$ some function over the semantic content of the utterance ($\mathbf{p}$ and $\mathbf{x}$ are rule-level variables binding terms on the nodes where the rules apply):[6]

(14)

IF	$Ty(\mathbf{x}), Fo(\mathbf{p})$
THEN	make(L), go(L)
	put($\mathbf{A}(\mathbf{U}, \mathbf{V}, \mathbf{F(p)})$)
ELSE	abort

In order to distinguish content that is derived directly on the basis of linguistically provided information and content derived on the basis of such inferences we introduce a partition in the TTR representation: we take the *cont* field to indicate the (linguistically-derived) truth conditional content and introduce an *inf* field for the speech act content derived by means of such rules (this roughly corresponds to the *explicature*/*high level explicature* distinction in Relevance Theory). So, for illustration, a suitable (optional) rule for assertions might perhaps apply to $Ty(t)$ trees with proposition p and speaker a, allowing one to infer the extra content $assert(a, p)$:

(15)

$$Tn(0), Ty(t), \left[\begin{array}{ll} ctxt: & \left[\begin{array}{ll} a & : participantA \\ u_0 & : utt-event \\ s_{s0} & : spkr(u_0, a) \end{array}\right] \\ cont: & \left[\begin{array}{ll} x & : john \\ p & : arrive(x) \end{array}\right] \end{array}\right]$$

$$Ty(e), \left[\begin{array}{l} x : john \end{array}\right] \qquad Ty(e \to t), \lambda[x].\left[\begin{array}{l} p : arrive(x) \end{array}\right]$$

$$\langle L\rangle Tn(0), \left[\begin{array}{ll} ctxt: & \left[\begin{array}{ll} a & : participantA \\ u_0 & : utt-event \\ s_{s0} & : spkr(u_0, a) \end{array}\right] \\ cont: & \left[\begin{array}{ll} x & : john \\ p & : arrive(x) \end{array}\right] \\ inf: & \left[\begin{array}{l} p' : assert(u_0, a, p) \end{array}\right] \end{array}\right]$$

4.1 SUs and speech acts

Given this, we can outline an account of SUs in which the same linguistic input can be construed as performing different possible speech acts (perhaps simultaneously). Consider the simple (and constructed) example in (16):

(16) A: John ...
B: arrived?

There are (at least) two possible readings of the resulting collaboratively produced contribution: one in which B is (co-)querying whether John arrived; and one in which B is clarifying A's original speech act, i.e., B is asking whether A was asking that John arrived. The tree resulting from parsing (or producing) this SU will be similar to the one in Fig 1 above, except that, due to the speaker change, the second utterance event u_1 is shown as spoken by B (see the unboxed part of Fig 2).

Applications of computational rules as in (14) above allow us to infer the speech act information corresponding to the two possible readings, deriving LINKed sub-trees which indicate speech acts performed by whichever participant is taken as the agent. One possible rule would derive the simple "co-querying" reading (based on the interrogative intonation and the identity of the final speaker B) adding the speech act proposition that B is asking whether John arrived – see the upper box in Fig 2. An alternative rule would derive the "clarificational" reading shown in the lower box. Of course, other inferences may also be possible.[7]

[6] The nature of **F** will depend on speech act type; for an assertion, it may simply be the identity operator; for irony, negation (see e.g. Asher and Lascarides (2003) for suggestions on how speech act type may relate to semantics).

[7] If such inferences become grammaticised, i.e. a particular construction is associated with a particular act (e.g. *clarification*), only one rule may be available. This is an empirical issue which we set aside here, but see (Ginzburg, forthcoming).

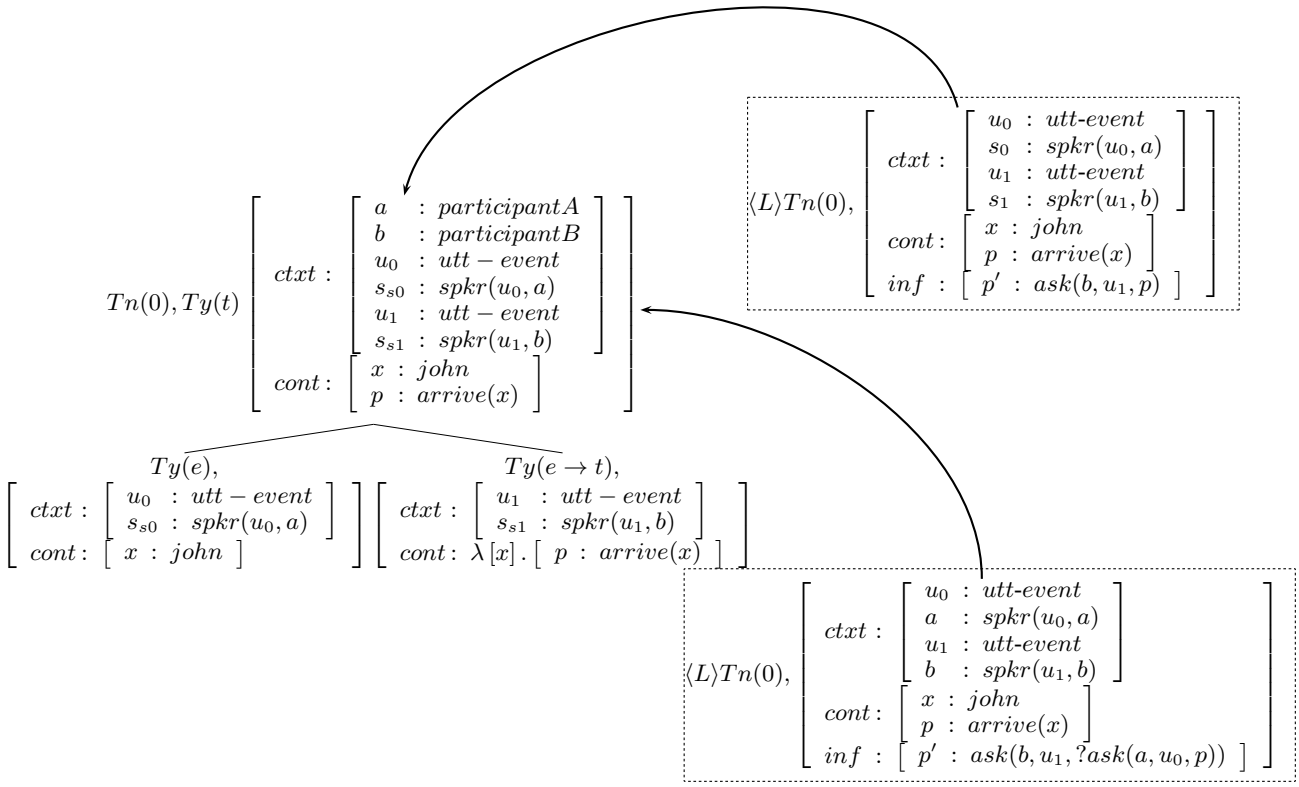
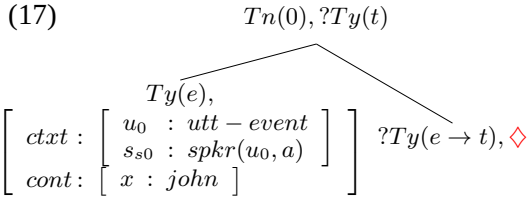


Figure 2: SU-derived tree

Note that Fig 1 and Fig 2 display representations of the final state that a parser might be in after B's contribution; from an incremental processing point of view, we are also interested in the state at the transition point (the change in speaker). Without considering any speech act inference, the tree at this transition point will be as follows:

(17)



This tree is partial (i.e. incomplete, having as yet unsatisfied requirements), but in itself is enough for B to begin generating – provided that they have some suitable message in mind (encoded as a *goal tree* in DS) which is subsumed by this partial tree. There is no requirement for B (or indeed A) to complete this tree, or perform any inference about speech acts, in order to begin generation (or, in A's case, parsing). In cases where B's continuation matches what the original speaker A could have intended to convey, the appearance would be one of "guessing", even though B has not performed any kind of inference regarding A's speech act. In fact, as (18) shows, completions of another speaker's utterance by no means need to be what the original speaker actually had in mind:

(18) Daughter: Oh here dad, a good way to get those corners out
Dad: is to stick yer finger inside.
Daughter: well, that's one way.
[from (Lerner, 1991)]

Such continuations can be completely the opposite of what the original speaker might have intended as in what we will call "hostile continuations" or "devious suggestions" – which are nevertheless collaboratively constructed from a syntactic point of view:

(19) (A and B arguing:)
A: In fact what this shows is
B: that you are an idiot

(20) (A mother, B son)
A: This afternoon first you'll do your homework, then wash the dishes and then
B: you'll give me £10?

Given a suitable model of the domain at hand, B, sometimes, will presumably be able to determine the content of A's intended speech act and represent it as such, i.e., as a speech act emanating from A, in their goal tree (see e.g. Poesio and Rieser (2010)). We take this not to be an essential process for the production of SUs, although it could be necessary in cases where B's next move is specifically intended as a confirmation request for such a representation.

4.1.1 The Split Turn-Taking Puzzle

Ginzburg (1997) describes a Turn-Taking Puzzle (TTP), which, he argues, shows that options for ellipsis resolution are distinct for speaker and hearer. This is illustrated by means of *why*-fragments:

(21) A: Which members of our team own a parakeet?
B: Why? (= *'Why are you asking which members of our team own a parakeet?'*)

(22) A: Which members of our team own a parakeet? Why?
(a) = *'Why own a parakeet?'*
(b)# *'Why am I asking this?'*

(23) A: Which members of our team own a parakeet? Why am I asking this question?

According to Ginzburg, the reading in which *why* queries the intended speech act (the why_{meta} reading) is available when asked by B (21) but unavailable when asked by the original speaker A (22). However, this is not simply due to coherence or plausibility, as it is available in (23) when expressed by non-elliptical means. Its unavailability must therefore be related to the way context is structured differentially for speaker and hearer.

Our explanation of this puzzle takes the why_{meta} interpretation as querying the intention/plan[8] behind the original speaker's speech act.[9] Since ellipsis resolution requires the potential for immediate accessibility of a salient representation, the infelicity of (22b) shows that the speaker's own intention behind their speech act is, in general, not salient enough for them to question it through *why*-ellipsis[10] (in Ginzburg's formulation such a fact does not belong in the TOPICAL FACTS field; however, this fact obtaining is not impossible, as (23) shows). Under this explanation, the TTP then reveals which agent takes responsibility for performing the relevant speech act, and hence can be queried about their intentions behind this act. In terms of (Goffman, 1981)'s distinctions among "speaker"-roles, the relevant agent is the 'Principal'. This can be evident in cases of SUs in multi-party dialogue. Now the utterer of a completion (the *final* "speaker" in the general sense discussed so far, and as indexed by pronouns like *my*) can felicitously ask elliptical why_{meta} questions of the *original* speaker (we will call this phenomenon the STTP, or Split Turn-Taking Puzzle):

(24) A to C: Have you finished sharpening ...
B to C/A: my loppers? B to A: Why?
(a) = *'Why are you asking C whether she has finished sharpening my loppers?'*
A to B: Because I want her to sharpen my secateurs too.

We can explain B's *why*-fragment interpretation in (24a) if we assume that although B's fragment *my loppers?* completes A's question, B does not necessarily assume responsibility for the performance of the speech act. That is, A must be taken as the agent of the querying speech act even though there is a sequence of utterance events which A and B have performed severally.[11] The availability of the why_{meta} reading then follows, even though apparently in contrast to (22b).

In some cases, then, even though the turn is collaboratively constructed, the original speaker maintains the authority or responsibility for the turn even though it was completed by somebody else. In other cases, see e.g. the hostile completions (19) and devious suggestions (20), this is not the case: the eventual content derived has to be taken as solely attributable to the second speaker. Notice however that in all cases (except those of direct quotation), the content of indexicals like *my* and *you* tracks directly the actual speaker/addressee, irrespective of who is taking responsibility for the content (or speech act performance). Even in helping out somebody to finish their sentence such indexicals will track the actual utterer/listener:

(25) Child (playing with toy garden tools): Give me my ...
Mum: your secateurs. Here they are, in fact these are loppers.

(26) A: Next cut it with your ...
B: my loppers. No, this we cut with the secateurs.

[8] Note that this approach does not necessitate that speech act and therefore intention information is available PRIOR to the processing of the *why*-question: instead, seeking to interpret such questions can be the trigger for optional (speech-act inducing) rules to apply. Hence, this approach is perfectly compatible with the general view on intentions as post-facto constructs (see e.g. Suchman (2007)) and the fact that conversational participants negotiate the content of speech acts with such assignments able to emerge retrospectively.

[9] As (Ginzburg, forthcoming) notes, recognition of this intention is *not* necessary for grounding.

[10] However, it is not impossible:

(i) A: Piss off. Why? Probably because I hate your guts.

[11] In fact, the specification of the why-fragment as why_{meta} can be taken to trigger the inference that A is solely responsible for the query as B dissociates himself from it.

This provides evidence for the dissociation of speech act performance and performance of the utterance event: these are two distinct actions whose agent might coincide but not necessarily so (these two roles roughly correspond to Goffman (1981)'s 'Author/Principal' and 'Animator'). Most accounts conflate the two: Lascarides and Asher (2009) argue that each time a speaker makes a conversational move they undertake a *public commitment*. However, SU examples such as (1)-(2) and (25)-(26) show that the person undertaking the public commitment (the 'Principal') does not necessarily coincide with actual utterer (the 'Animator'). We therefore conclude that the notion of 'commitment' should be correlated with something else, namely, who is performing (the agent of) the associated speech act (which could be the two speakers jointly but not necessarily and not only for SUs). Speech act inference rules as outlined in (14) must therefore maintain the flexibility to assign the inferred speech act to any of the speakers involved, and not only the final one.

5 Conclusions

The STTP and the multifunctionality of SU fragments motivates our claim that information manipulated during a parse has to be distinguished at three levels: semantic content which is directly derived on the basis of the linguistic input, context specifications arising from the utterance situation (utterance events) and optional speech act information. Formulation of this information in a DS-TTR combined formalism allows the interactions required for appropriate processing of SUs.

Acknowledgments

We gratefully acknowledge valuable support and advice from Ruth Kempson and Greg Mills. Thanks also to Pat Healey, Graham White, Arash Eshghi and Chris Howes. This work was supported by DynDial (ESRC-RES-062-23-0962).

References

N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

R. Cooper and J. Ginzburg. 2002. Using dependent record types in clarification ellipsis. In *Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*.

R. Cooper. 1998. Information states, attitudes and dependent record types. In *Proceedings of ITALLC-98*.

R. Fernández and J. Ginzburg. 2002. Non-sentential utterances: A corpus-based study. *Traitement Automatique des Langues*, 43(2).

A. Gargett, E. Gregoromichelaki, R. Kempson, M. Purver, and Y. Sato. 2009. Grammar resources for modelling dialogue dynamically. *Cognitive Neurodynamics*, 3(4):347–363.

J. Ginzburg, I. Sag, and M. Purver. 2003. Integrating conversational move types in the grammar of conversation. In *Perspectives on Dialogue in the New Millennium*, volume 114 of *Pragmatics and Beyond New Series*, pages 25–42. John Benjamins.

J. Ginzburg. 1997. On some semantic consequences of turn taking. In *Proceedings of the 11th Amsterdam Colloquium on Formal Semantics and Logic*.

J. Ginzburg. forthcoming. *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

E. Goffman. 1981. *Forms of talk*. University of Pennsylvania Press.

E. Gregoromichelaki. 2006. *Conditionals: A Dynamic Syntax Account*. Ph.D. thesis, King's College London.

R. Kempson, W. Meyer-Viol, and D. Gabbay. 2001. *Dynamic Syntax: The Flow of Language Understanding*. Blackwell.

R. Kempson, A. Gargett, and E. Gregoromichelaki. 2007. Clarification requests: An incremental account. In *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue (DECALOG)*.

A. Lascarides and N. Asher. 2009. Agreement, disputes and commitments in dialogue. *Journal of Semantics*, 26(2):109.

G. Lerner. 1991. On the syntax of sentences-in-progress. *Language in Society*, pages 441–458.

M. Poesio and H. Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1:1–89.

M. Purver, J. Ginzburg, and P. Healey. 2003. On the means for clarification in dialogue. In *Current and New Directions in Discourse & Dialogue*, pages 235–255. Kluwer Academic Publishers.

M. Purver, R. Cann, and R. Kempson. 2006. Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation*, 4(2-3):289–326.

M. Purver, C. Howes, E. Gregoromichelaki, and P. G. T. Healey. 2009. Split utterances in dialogue: a corpus study. In *Proceedings of SIGDIAL*.

L. Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Coherence and Rationality in Grounding

Matthew Stone
Computer Science and Cognitive Science
Rutgers University
Piscataway NJ 08854-8019 USA
Matthew.Stone@Rutgers.edu

Alex Lascarides
School of Informatics
University of Edinburgh
Edinburgh EH8 9AB Scotland UK
alex@inf.ed.ac.uk

Abstract

This paper analyses dialogues where understanding and agreement are problematic. We argue that pragmatic theories can account for such dialogues only by models that combine linguistic principles of discourse coherence and cognitive models of practical rationality.

1 Introduction

Interlocutors in conversation have only indirect evidence as to whether others understand and agree with them. Take the joke about the old folks in the bus shelter:

(1) a. *A*: Windy, en'it?
b. *B*: No it's not, it's Thursday.
c. *C*: So am I. Let's go and 'ave a drink!

Evidently *B* mishears *windy* as *Wednesday* and *C* mishears *Thursday* as *thirsty*. Even when somebody says they understand and agree, it's no guarantee that they do.

Our judgements about (1) depend on *principles of discourse coherence*. *B* formulates (1b) as a denial of (1a), so *B* must think that (1b) is semantically incompatible with (1a). Knowing this, *A* can infer information about *B*'s interpretation of (1a). The implicit discourse relation connecting the two halves of (1c) similarly shows that *C* thinks the salient property *B* and *C* share gives them a reason to go have a drink.

But there's more going on. After all, (1) is *not* a coherent discourse. Our judgements about (1) also rely on our knowledge of the kinds of mistakes that people can make in conversation—hearing one word for another, for example—and our presumption that people choose their utterances reasonably to fit the conversation as they understand it. Such inferences represent *cognitive modelling*.

The problem of managing understanding and agreement in conversation is known as grounding (Clark, 1996). In the formal and computational literature, previous approaches to grounding have focused either on discourse coherence or on cognitive modelling, but failed to consider the interactions between the two. In this paper, we outline how the two sets of considerations can be reconciled. We regiment utterance content so that it encapsulates the way dialogue moves are coherently or rhetorically connected to prior utterances. And we apply probabilistic reasoning to assess speakers' rationality in choosing to commit to specific contents. We analyse naturally-occurring examples involving implicit grounding, ambiguous grounding moves, misunderstandings and repair to illustrate the need for both kinds of reasoning.

Our work contributes to three different spheres of investigation. Firstly, it helps to explain how it might be possible for interlocutors to draw precise conclusions about others' mental states, despite the complexity and indirectness of the linguistic evidence. Secondly, it contributes to the Gricean programme of analysing conversation as cooperative activity, by showing how an important and independently-characterised set of conversational inferences might actually be calculated. (Of course, as in (1), these inferences need not always involve Gricean *implicature*.) And finally, we particularly hope that our work will inform the design of more robust and powerful conversational systems, by correlating the architecture, reasoning and knowledge that is realised in these systems with the grounding those systems can do.

2 Examples and Perspective

Following Clark (1996), we view grounding fundamentally as a skill rather than as an epistemic state. Interlocutors achieve grounding when they can detect misunderstanding, clarify utterances, negotiate meaning and coordinate their responses

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

in pursuit of successful joint activity. They may or may not thereby achieve *common ground* in the philosophical sense (Stalnaker, 1978). Our view is that the skill of grounding reflects the ability to entertain multiple hypotheses about the organisation of dialogue and to rank these hypotheses quantitatively to make strategic choices. Coherence and rationality are both essential to these calculations.

Dialogue (1) illustrates how principles of discourse coherence contribute to inferences about the nature of an implicit misunderstanding. Lascarides and Asher (2009) use dialogue (2) from Sacks et al. (1974, p.717) to illustrate how principles of coherence can also contribute to inferences about implicit agreement and understanding:

(2) a. *Mark (to Karen and Sharon):*
Karen 'n' I're having a fight,
b. after she went out with Keith and not me.
c. *Karen (to Mark and Sharon):*
Wul Mark, you never asked me out.

Intuitively, Mark and Karen agree that they had a fight, caused by Karen going out with Keith and not Mark. Thus *implicatures can be agreed upon*—that (2b) explains (2a) goes beyond compositional semantics. Furthermore, *agreement can be implicated*—Karen does not repeat (2a), (2b) or utter *OK* to indicate agreement.

As in (1), the basis for recognising Karen's implicit acceptance stems from coherence, which compels us (and Mark) to recognise the rhetorical connection between her contribution and Mark's. Here, the fact that Karen commits to (2c) *explaining why* (2b) is true should be sufficient to recognise that Karen accepts Mark's utterance (2b). Karen's implicit endorsement of (2b) also seems sufficient to conclude that she (implicitly) accepts its *illocutionary effects* as well—(2b) explaining (2a). The fact that Karen chooses to accept these contributions, rather than to ask about them, for example, offers very good evidence that she thinks she understands their content.

Incrementally, as discourse unfolds, interlocutors have only partial information about these contributions. As described by Clark (1996), grounding requires interlocutors to manage uncertainty at four levels: (1) the signals that they exchange with one another; (2) the words that are used; (3) the meanings that those words convey; and (4) what commitments interlocutors make to these meanings. The joke in (1) trades on the difficulty of grounding at Levels 1 and 2. Given the endemic semantic ambiguity, vagueness, and other forms of underspecification associated with utterances, interlocutors frequently also face transient uncertainties about their partners' contributions at Levels 3 and 4.

Interlocutors' choices in conversation reflect the specific ambiguities they encounter and the likelihood they assign to them. For example, when interlocutors see their uncertainty about a prior public commitment, or piece of logical form, as problematic, they can seek clarification, as the sales assistant *B* does in (3b)—a simplified version of a dialogue from the British National Corpus (Burnard, 1995) that is annotated with clarification acts (Purver et al., 2003) (we thank Matthew Purver for pointing us to this example):

(3) a. *A*: I would like one of the small reducers.
b. *B*: One going from big to small
or from small to big?
c. *A*: Big to small.
d. *B*: Big to small, ok.

(3b) is an example where specific clarification is sought on the intended meaning of *small reducers*. In (4), from DeVault and Stone (2007), *B* seeks specific clarification on the *illocutionary* content of *A*'s utterance (4b) rather than its locutionary content: was it an *Acceptance* of (4a) or something else, perhaps merely an *Acknowledgement*?

(4) a. *B*: Add the light blue empty circle please.
b. *A*: okay
c. *B*: Okay, so you've added it?
d. *A*: i have added it.

In both cases, *A*'s response to *B*'s clarification request is designed to help *B* resolve the specific ambiguity that *B* has called attention to.

Of course, as Clark (1996) underscores, not all uncertainty is problematic. If the issue is sufficiently unimportant or a misunderstanding is sufficiently unlikely, interlocutors can choose to tolerate the uncertainty and proceed anyway. This is crucial in systems where modules like speech recognition never offer certainty (Paek and Horvitz, 2000). But it could also be what *B* does in (1) or Karen does in (2), for instance.

So overall, grounding moves and anti-grounding moves can be implicit (see (2) for grounding and (1) for anti-grounding) or explicit (see (3cd) and (4ab) for grounding and (3ab)

and (4bc) for anti-grounding). Moreover, a misunderstanding or lack of grounding can be mutually recognised (see (3) and (4)) or not (see (1)). Even when a grounding move is explicit, there can still be uncertainty about both the level of grounding that the agent has reached—-e.g., *B* is uncertain whether *A*'s explicit endorsement in (4b) marks grounding at Level 3 or grounding at Level 4. There can also be uncertainty about the semantic scope of the endorsement—e.g., an utterance like *I agree* doesn't make explicit whether the acceptance is of all the clauses in the prior turn or only the last clause (see Lascarides and Asher (2009) for discussion).

3 Challenges

Our work draws on previous grounding models based on *discourse coherence* and those based on *probabilistic inferences about strategy*. Both of these traditions provide insights into the data of Sections 1 and 2, but neither tells a complete story.

Coherence approaches start from the insight that the relationships between utterances in discourse give evidence about mutual understanding. An early illustration of this type of reasoning is the work of McRoy and Hirst (1995), who recognise and repair misunderstandings in dialogue by identifying utterances that are best explained by assuming that the speaker's public commitments about the coherent organisation of the discourse are in conflict with those of the addressee.

More recent work in the coherence tradition tends to adopt the influential approach of Traum (1994), who posits specific categories of communicative action in dialogue, called *grounding acts*. The prototypical grounding acts model works by modeling assertions as introducing content with a status of *pending*. Subsequent acknowledgement acts may transfer that content out of what's *pending* and into what's *grounded*. Important work in this tradition includes both theoretical analyses (Poesio and Traum, 1998; Ginzburg, 2010) and system-building efforts (Matheson et al., 2000; Traum and Larsson, 2003; Purver, 2004).

Lascarides and Asher (2009) simplify and extend this idea. They analyse dialogue in terms of a single set of relational speech acts, formalised so as to represent what information each act commits its agent to, implicitly or explicitly. The account predicts facts about implicit grounding, illustrated in dialogue (2), without the need to describe the dialogue in terms of a separate layer of inferred grounding acts. We build on their account here.

Such models are good at characterising agreement but not as good at characterising uncertainty or misunderstanding. For example, in cases where interlocutors proceed despite uncertainty, neither a *pending* status nor a *grounded* status seems appropriate. On the one hand, interlocutors accept that there may be errors; on the other, they act as though the likely interpretation was correct. Such models are also limited by their *symbolically-defined* dynamics. Misunderstandings like those in (1) surface in the dialogue as inconsistencies that can potentially be corrected in a vast number of alternative ways—some of which are intuitively likely, others of which are not. Symbolic models need rules to specify which hypotheses are worth exploring—an open problem—while probabilistic models naturally assign each one a posterior probability based on all the available information.

Probabilistic approaches to grounding were inaugurated by Paek and Horvitz (2000), who describe the decision-theoretic choices of a spoken language interface directly in terms of Clark's model of contributing to conversation. Paek and Horvitz characterise their system's information state in terms of the probabilistic evidence it has about the real-world goals that users are trying to achieve with the system. This evidence includes the system's prior expectations about user behaviour, as well as the system's interpretations of user utterances. Paek and Horvitz show that this representation is expressive enough for the system to assess conflicting evidence about user intent, to ask targeted clarification questions, and to adopt an appropriate grounding criterion in trading off whether to seek more information or to act in pursuit of users' likely domain goals.

A range of related research has exploited probabilistic models in dialogue systems (Walker, 2000; Roy et al., 2000; Singh et al., 2002; Bohus and Rudnicky, 2006; DeVault and Stone, 2007; Williams and Young, 2007; Henderson et al., 2008). However, this research continues to focus primarily on inference about user goals, while largely sidestepping the knowledge and inference required to relate utterances to discourse context, as illustrated in dialogue (2). Moreover, because this work is generally carried out in the setting of spoken dialogue systems, researchers usually formalise whether the system understands the user, but draw no inferences about whether the user un-

derstands the system.

The present paper aims to reconcile these two perspectives in a common theoretical framework.

4 Public Commitments

We adopt from Lascarides and Asher (2009) a representation of the logical form (LF) of coherent dialogue.[1] This LF records the content to which each speaker is publicly committed through their contributions to the dialogue. Commitments are relational. Each utterance typically commits its speaker not only to new content, but also to a specific implied connection to prior discourse (maybe an utterance by another speaker), and perhaps indirectly to earlier content as well. The inventory of these *rhetorical relations* maps out the coherent ways dialogue can evolve—examples include *Explanation*, *Narration*, *Answer*, *Acknowledgement* and many others (Lascarides and Asher, 2009). Pragmatic rules for reconstructing implied relations provide a defeasible mechanism for resolving ambiguity and calculating implicatures.

More formally, the LF of a dialogue in Dialogue SDRT (DSDRT) is the LF of each of its turns, where each turn maps each dialogue agent to a Segmented Discourse Representation Structure (SDRS) specifying all his current public commitments. An SDRS is a set of labels (think of labels as naming dialogue segments) and a mapping from those labels to a representation of their content. Because content includes rhetorical relations $R(a,b)$ over labels a and b, this creates a hierarchical structure of dialogue segments. SDRSs are well-formed only if its set of labels has a unique root label—in other words, an SDRS represents just one extended dialogue segment consisting of rhetorically connected sub-segments.

Abstracting for now away from uncertainty, Lascarides and Asher (2009) suggest that by the end of dialogue (2) Mark and Karen are respectively committed to the contents of dialogue segments π_{1M} and π_{2K}, as shown in (2′) (contents of the 'minimal' segments a, b and c are omitted for reasons of space; we label the public commitments of speaker s in turn t with segment π_{ts}):

(2′) Mark: $\pi_{1M} : \mathit{Explanation}(a,b)$
Karen: $\pi_{2K} : \mathit{Explanation}(a,b) \wedge \mathit{Explanation}(b,c)$

[1] LF is a public construct like a game board, not a subjective construct related to mental state. Though controversial, this view is defensible—and it makes probabilistic modeling a lot easier (DeVault and Stone, 2006).

Karen's and Mark's public commitments share labels a and b. This reflects the reality that an agent's dialogue move relates in a coherent way to prior contributions. Assuming that agreement (or grounding at Level 4) is shared public commitment, LF (2′) entails that Mark and Karen agree that (2b) caused (2a). Lascarides and Asher (2009) infer that $\mathit{Explanation}(a,b)$ is a part of Karen's commitment, given her commitment to $\mathit{Explanation}(b,c)$, via default principles that predict or constrain the semantic scope of implicit and explicit endorsements and challenges. The relevant default principle here is that an implicit endorsement of a prior utterance normally involves acceptance of its illocutionary effects as well.

5 Strategy and Uncertainty

The assumption that interlocutors are pursuing reasonable strategies for pushing the conversation forward, given their information state, often allows observers to draw powerful inferences about what that information state is. For example, suppose Mark understands Karen's move correctly in (2), and thus assigns a high probability to the representation of Karen's commitments that we have ascribed in (2′). Mark can reason that since Karen has accepted the meaning that he intended to convey, then she must have understood it. Thus, implicit agreement—and even disagreement, as in (1)—should make it possible to draw conclusions about (implicit) grounding at lower levels.

In other cases, the representation of an agent A's public commitments may feature a segment a uttered by a prior agent B, and yet by the dynamic interpretation of A's SDRS A is not committed to a's content or its negation. In such cases, observers may not be able to tell whether A has identified the content associated with the earlier utterance a. In other words, the LF reveals a *lack of grounding*. For instance, A's public commitments may include a relation $CR(a,b)$ (*CR* for Clarification Request), whose semantics entails that b is associated with a question K_b all of whose possible answers help to resolve the meaning associated with utterance a. Normally, A would make such a move only when A was uncertain about that content. Seeing a *CR* thus allows interlocutors to infer a lack of grounding at Level 3. Some clarificatory utterances, such as *echo questions* or *fragment reprises* have additional constraints on their use, which reveals even more about what an interlocutor did or did not

recognise at Level 1 or Level 2. See Purver (2004) and Ginzburg (2010) for more formal details about clarification requests and their semantics.

In our view, these inferences are ultimately about what it's rational for a speaker to do. People tend to avoid agreeing with something they know they don't understand, or asking about something they know they do. Doing so doesn't move the conversation forward. In other words, these inferences rest on principles of *cognitive modelling* which are different from, and complementary to, the principles of interpretation which characterise the possible logical form of discourse.

In Figure 1, we schematise our approach to these inferences qualitatively in a dynamic Bayesian network (DBN). The model describes the discourse context as a public scoreboard that evolves, step by step, as a consequence of the moves interlocutors make to update it. M_t is the move made at time t. We think of it as a relational speech act; that is, as a bit of logical form with an intended rhetorical connection to the discourse context. In other words, M_t completely resolves anaphoric reference, discourse attachment, and the propositions expressed. The move for (2c), for example, would include the rhetorical connections $Explanation(a,b) \wedge Explanation(b,c)$ and the content of segment c. X_t is the discourse context at time t. We assume that it is a DSDRS, as illustrated in (2′). Finally, E_t is the observable utterance associated with the update at time t. Depending on the modality of conversation, this might be typed text, acoustic form, or the observable correlates of a multimodal communicative act.

The relevant dynamics involve two ingredients. A model of discourse coherence and discourse update, expressed as $X_{t+1} = u(X_t, M_t)$, describes how moves update the current context to yield a new context. This is the familiar update of dynamic semantics—when X_t and M_t are compatible, X_{t+1} is a new context that takes the information from both into account; otherwise, in cases of incoherence, presupposition failure and the like, X_{t+1} is a defective context that specifies the attempts made and the fact that they failed. A model of language, expressed as $P(E_t|X_t)$, describes the relationship between utterance form and meaning; uncertainties here reflect the variance in the way an utterance may be performed and observed.

The cognitive model surfaces in Figure 1 through models of discourse interpretation and discourse planning. The DBN casts the conversation as involving alternating contributions from two interlocutors A and B. We use the variables A_t and B_t to represent the subjective information state of these agents at time t. The models of discourse interpretation yield updates in the interlocutors' mental states as a function of their observations of an utterance produced by their partner. They are formalised as relationships $P(A_{t+1}|A_t, E_t)$ when t is even and $P(B_{t+1}|B_t, E_t)$ when t is odd. The models of discourse planning, meanwhile, describe the moves interlocutors make as a function of their current information state, and who takes the turn to speak. We have $P(M_{t+1}|A_t)$ when t is even and $P(M_{t+1}|B_t)$ when t is odd. Discourse coherence takes on new force in these planning models. Rational agents strive to make coherent moves, and thereby to commit to certain propositions that match their beliefs and interests.

The network as a whole is analogous to a Hidden Markov Model, with the observable state given by a sequence of utterances E_1 through E_n, and the hidden state at each time given by the joint distribution over X_t, A_t and B_t. A probabilistic *observer* of a conversation reasons in this network by observing the utterance sequence and reasoning about the hidden variables. That's the position we're in when we read an example like (1). The posterior distribution over the hidden state would normally permit specific conclusions about X_t. If the model predicts that A follows this aspect of the dialogue state, the model derives a match between A_t and X_t. If the model predicts A doesn't follow, the model would associate A_t with a value or values that don't match X_t. The model can make the same predictions even if it cannot pin down X_t. This situation would be realised by a broader posterior distribution over X_t and by correlations in the joint distribution over A_t and X_t.

The model would be used differently to implement a *participant* in a conversation. A participant in a conversation doesn't need to draw inferences about their own mental state; they actually implement particular interpretation and planning procedures. These procedures, however, would have a rational basis in the probabilities of the model, if the agent takes not only E_t but also the values for their own moves M_t as observed, and uses the model only to draw inferences about their partner.

This point bears on the nature of models such as $P(A_{t+1}|A_t, E_t)$, and $P(M_t|A_t)$. They can implicitly encode arbitrarily complex reasoning. Thus, Figure 1 is best thought of as shorthand for a *net-*

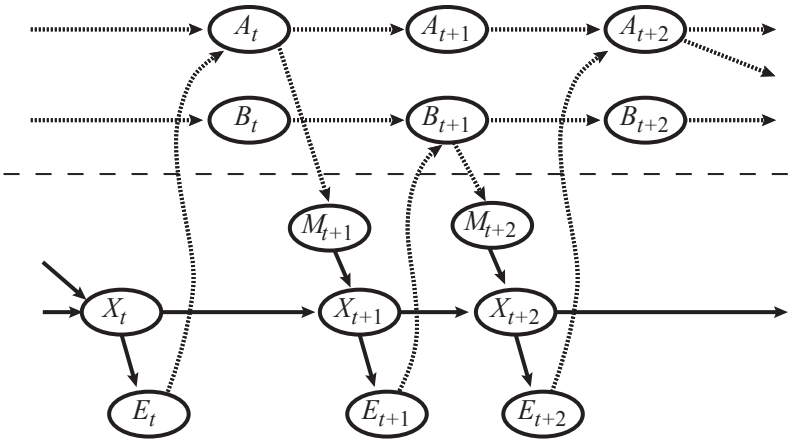


Figure 1: Fragments of the DBN indicating the probabilistic relationships relating one interlocutor A's mental state, the other interlocutor B's mental state, interlocutors' alternating discourse moves M, the evolving discourse context X and the observable correlates of discourse update E (including utterances). Solid dependencies indicate linguistic models; dotted ones, cognitive models.

work of influence diagrams (Gal, 2006). For example, suppose A tracks the hidden dynamics of the conversational record by Bayesian inference. Then A's information state at each time t includes expectations about the current discourse context X_t given the evidence A has accumulated so far in the discourse O_{At} (a combination of observed utterances and planned moves)—this serves as a prior distribution $P_A(X_t|O_{At})$ that's part of A's information state. A also has discourse expectations $P_A(X_{t+1}|X_t)$ and a linguistic model $P_A(E_{t+1}|X_{t+1})$. To describe the discourse interpretation of our Bayesian agent we use a standard DBN definition of filtering, as in (5).

(5) $P_A(X_{t+1}|O_{At+1}) \propto$
$\sum_{X_t} P_A(E_{t+1}|X_{t+1})P_A(X_{t+1}|X_t)P_A(X_t|O_{At})$.

This posterior distribution describes A's state at time $t+1$. This more specific model lets us flesh out how A acts to achieve coherence in planning M_{t+2}. For example, if $P_A(X_{t+1}|O_{At+1})$ assigns a high value to DSDRS K, then $P(M_{t+2}|A_{t+1})$ will be low when $u(K,M_{t+2})$ is incoherent.

Similarly, A may have a substantive model of B's planning and interpretation. Then A's information state will involve a distribution $P_A(B_t)$ over A's model of B, and A will have expectations of the form $P_A(B_{t+1}|E_{t+1},B_t)$ and $P_A(M_{t+1}|B_{t+1})$ describing B's interpretation and planning. These models now underwrite A's discourse expectations, allowing $P_A(X_{t+1}|X_t)$ to be described in terms of $P_A(M_{t+1}|B_t)$. It could be that A has a very simple model of B. Maybe A assumes B understands perfectly, or guesses interpretations at random. However, as familiar game-theoretic considerations remind us, A might instead model B as another Bayesian reasoner. That model may even describe B via a nested model of A! A useful assumption is that agents are uncertain about the exact degree of sophistication of their partner, but assume it is low (Camerer et al., 2004).

6 Worked Examples

We use the examples of Sections 1 and 2 to illustrate the dimensions of variation which our model affords. To make the discussion concrete, we will consider the reasoning of one interlocutor, typically A, using a probabilistic model of the form illustrated in Figure 1. Thus the whole of Figure 1 is understood to encode A's knowledge, with suitable variables (e.g., A_t, E_t and M_{t+1}) observed and the joint distribution over the other variables inferred. We are interested in cases where A speaks, and then A retrospectively assesses B's interpretation of what A has said in light of B's response. We will not assume that A maintains a detailed model of B's planning process. However, we assume that A tracks B's probabilistic representation of the discourse context, and moreover that A and B apply a common, public model of discourse update $u(X_t,M_{t+1})$ and of linguistic expression $P(E_t|X_t)$. For simplicity in treating the examples, we also assume that there are no pending ambiguities in the initial context, so that effectively X_0 and B_0 are observed (by both interlocutors). Obviously, this assumption does not hold in general in the model.

A's inference involves three mathematical constructs. The first is A's assessment of B's interpretation of an initial move M_1 made by A. A is uncertain about the probability B assigns to particular interpretations of the discourse up to time 1, given B's available evidence. This means the model has a continuous random variable z_i for each candidate DSDRS representation K_i for the discourse; z_i gives the probability that B thinks the interpretation of the discourse up to time 1 is K_i. If we take A and B to entertain N interpretations, $\vec{z}$ is a vector in N-dimensional space, subject to the constraint that coordinates sum to one (a point on the $N-1$ simplex). B's state at time 1 thus includes a vector $\vec{z}$ and the model includes a prior probability density over this vector, conditional on available evidence: $p(\vec{z}|X_1, B_0)$ which we represent mathematically as $pr(\vec{z})$. Assuming Bayesian inference by B, it is derived from B_0 via (5) by marginalising in expectation over E_1.

The second key construct describes A's expectations about what B will do next. This is realised in the model's value for the likelihood $P(M_2|B_1)$. Concretely, for each epistemic state $\vec{z}$ for B, the model assigns a likelihood $l(U_j|\vec{z})$ that B chooses move U_j in $\vec{z}$. For each epistemic state $\vec{z}$, the function $l(U|\vec{z})$ defines a point in the $D-1$ simplex, if there are D possible next moves, representing the model's expectation about B's behaviour there.

The final key construct is the model's retrospective assessment of what B's mental state must have been, given the move observed at E_2. That is the posterior $p(\vec{z}|X_1, B_0, E_2)$. We abbreviate this as $po(\vec{z})$; it is another density over the $N-1$ simplex. The model derives this by Bayesian inference:

$$(6)\ \ po(\vec{z}) \propto pr(\vec{z}) \sum_{U_j} l(U_j|\vec{z}) P(E_2|u(X_1, U_j))$$

The equation shows how an interlocutor gets retrospective insight into their partner's mental state by combining evidence from the observed utterance E_2 and discourse coherence, with inference about why the interlocutor might have planned such an utterance, $l(U_j|\vec{z})$, and expectations about what their mental state would have been, $pr(\vec{z})$.

Let's look at (1b). We track B's interpretation via a DSDRS K_1 saying it's windy and another K_2 saying it's Wednesday. We expect understanding, so our prior $pr(\vec{z})$ naturally favors $\vec{z}$ where K_1 has high probability. Now, given the utterance, we can assign high probability to an observed value U_2 for the variable M_2 with the form $Correction(a,b)$ where a is the immediately prior discourse segment and b says it's Thursday. Our posterior distribution $po(\vec{z})$ factors in our prior estimate of B's state, this evidence, and our model $l(U_2|\vec{z})$ of B's choice. Now $u(K_1, U_2)$ is incoherent and $u(K_2, U_2)$ is coherent, so $l(U_2|\vec{z})$ is going to be very low if $\vec{z}$ assigns much probability to K_1. That's how the model recognises the misunderstanding.

Now let's look at (2b). The model of language should predict that any value U_j for M_2 that features $Explanation(b,c)$ as a part will be more likely than an alternative that doesn't. This entails at least a partial commitment to the prior utterance. If we assume that agents tend not to commit in uncertain states—giving a low probability to $l(U_j|\vec{z})$ for such U_j when $\vec{z}$ has high entropy—then (6) sharpens our information about $\vec{z}$. Following Lascarides and Asher (2009), we assume a further constraint on dialogue policy: if you only partially endorse the prior discourse, you tend to say so. So the model predicts further probabilistic disambiguation: among those U_j that feature $Explanation(b,c)$, those that also feature $Explanation(a,b)$ will get a higher posterior probability than those that do not.

Next is (3b). Here the model of language should predict that any likely value U_j for M_2 will feature $CR(a,b)$. Rationality dictates for such U_j that $l(U_j|\vec{z})$ will be high only when $\vec{z}$ has high entropy, and this is reflected in our updated posterior over $\vec{z}$. Indeed, since the linguistic form of the clarification elicits particular information about the prior context, we can use similar reasoning to recognise particular points of likely uncertainty in $\vec{z}$. A similar analysis applies to (4).

7 Discussion and Conclusion

We have proposed a programmatic Bayesian model of dialogue that interfaces linguistic knowledge, principles of discourse coherence and principles of practical rationality. New synergies among these principles, we have argued, can lead naturally to more sophisticated capabilities for recognising and negotiating problematic interactions.

Of course, we must still specify a model in detail. We hope to streamline this open-ended effort by capturing important correlations in dialogue, as found in alignment phenomena for example, through simple generative mechanisms proposed in Pickering and Garrod (2004). We also face difficult computational challenges in fitting our models

to available data and drawing conclusions quickly and accurately from them. We would also like to determine whether existing Bayesian approaches to unsupervised learning, such as Goldwater and Griffiths (2007), can apply to our model. At any rate, until we can demonstrate our ideas through systematic implementation, training and evaluation, our account must remain preliminary.

Acknowledgements

Thanks to Mark Johnson, Matthew Purver, David Schlangen, Chung-chieh Shan and Mike Wunder for helpful discussion. Supported in part by NSF CCF 0541185 and HSD 0624191.

References

D. Bohus and A. Rudnicky. 2006. A K hypothesis + other belief updating model. In *Proceedings of the AAAI Workshop on Statistical and Emprical Approaches for Spoken Dialogue Systems*.

L. Burnard, 1995. *Users Guide for the British National Corpus*. British National Corpus Consortium, Oxford University Computing Service.

CF. Camerer, T. Ho, and J. Chong. 2004. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119:861–898.

H. Clark. 1996. *Using Language*. Cambridge University Press.

D. DeVault and M. Stone. 2006. Scorekeeping in an uncertain language game. In *Proceedings of SEMDIAL (BRANDIAL)*, pages 139–146.

D. DeVault and M. Stone. 2007. Managing ambiguities across utterances in dialogue. In *Proceedings of SEMDIAL (DECALOG)*.

Y. Gal. 2006. *Reasoning about Rationality and Beliefs*. Ph.D. thesis, Harvard.

J. Ginzburg. 2010. *The Interactive Stance: Meaning for Conversation*. CSLI Publications.

S. Goldwater and T. Griffiths. 2007. A Fully Bayesian Approach to Unsupervised POS Tagging. *Proceedings of ACL*, pages 744-751.

J. Henderson, P. Merlo, G. Musillo, and I. Titov. 2008. A latent variable model of syncrhonous parsing for syntactic and semantic dependencies. In *Proceedings of CoNLL*, pages 178–182.

A. Lascarides and N. Asher. 2009. Agreement, disputes and commitment in dialogue. *Journal of Semantics*, 26(2):109–158.

C. Matheson, M. Poesio, and D. Traum. 2000. Modelling grounding and discourse obligations using update rules. In *Proceedings of NAACL*, pages 2–9.

S. McRoy and G. Hirst. 1995. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):435–478.

T. Paek and E. Horvitz. 2000. Conversation as action under uncertainty. In *Proceedings of UAI*, pages 455–464.

MJ. Pickering and S. Garrod. 2004. Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225.

M. Poesio and D. Traum. 1998. Towards an axiomatisation of dialogue acts. In *Proceedings of SEMDIAL (TWENDIAL)*, pages 207–222.

M. Purver, J. Ginzburg, and P. Healey. 2003. On the means for clarification in dialogue. In R. Smith and J. van Kuppevelt, editors, *Current and New Directions in Discourse and Dialogue*, pages 235–255. Kluwer Academic Publishers.

M. Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College, London.

N. Roy, J. Peneau, and S. Thrun. 2000. Spoken dialogue management for robots. In *Proceedings of ACL*, pages 93–100, Hong Kong.

H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turntaking in conversation. *Language*, 50(4):696–735.

SP. Singh, DJ. Litman, MJ. Kearns, and MA. Walker. 2002. Optimizing dialogue management with reinforcement learning: Experiments with the NJFun system. *Journal of Artificial Intelligence Research*, 16:105–133.

R. Stalnaker. 1978. Assertion. In P. Cole, editor, *Syntax and Semantics*, pages 315–322. Academic Press.

D. Traum and S. Larsson. 2003. The information state approach to dialogue management. In R. Smith and J. van Kuppevelt, editors, *Current and New Directions in Discourse and Dialogue*, pages 325–353. Kluwer Academic Publishers.

D. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, University of Rochester.

MA. Walker. 2000. An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artifical Intelligence Resesarch*, 12:387–416.

JD Williams and SJ Young. 2007. Partially Observable Markov Decision Processes for Spoken Dialog Systems. *Computer Speech and Language*, 21(2):231–422.

Dialogues with conflict resolution: goals and effects

Katarzyna Budzynska
Institute of Philosophy,
Cardinal Stefan Wyszynski University
Warsaw, Poland
k.budzynska@uksw.edu.pl

Kamila Dębowska
Institute of English Philology,
Adam Mickiewicz University in Poznan
Poznan, Poland
kamila@ifa.amu.edu.pl

Abstract

The aim of the paper is to propose the semi-formal model of dialogues with conflict resolution. We focus on the specification for goals and effects of this type of dialogue. Our proposal is based upon the popular and influential model by D. Walton. We show that this model, even though referring directly to conflict resolution, does not allow to express its important properties. The paper proposes the model's modification and extension, which enables describing various characteristics related to the goals of conflict resolution. Moreover, we combine formal and linguistic concepts to define different kinds of effects achieved in this type of dialogues.

1 Introduction

The paper proposes the semi-formal model of dialogues with conflict resolution (CR). The model is to serve as a heuristic tool for the study of dialogues aiming at conflict resolution. The properties of the heuristic tool are rooted in both interpersonal dialogues and multi-agent systems (MAS).

The paper is organized into Section 2 and Section 3. In Section 2, drawing on the influential model established by D. Walton, we discuss the possible goals of dialogues with conflict resolution. Considerations in Section 2 serve also as underpinnings of our claim that Walton's model should be extended to enhance its applicability. In Section 3, relying on the heuristic power of the model we have proposed in Section 2, we focus on the degrees of reaching the initial goal of CR.

Although Walton's typology of dialogues takes into account the dialogue with conflict resolution, it omits its significant properties and aspects. More specifically, it does not allow for the distinction of the two types of strategies for achieving conflict resolution: egoistic and collaborative. While the first strategy enables an agent to strive for his/her individual goal, the second one emphasizes the superiority of the collaborative CR. The importance of the individual standpoints for agents is significantly diminished in the second case.

Our contribution to the existing models of dialogues is the proposal of clearly defined properties of CR. The first part of the model specifies the goals of dialogues with conflict resolution taking into account the pre-planned goals of system and agents. The second part of the model determines the types of the effects which can be achieved through a dialogue with the pre-planned conflict resolution. The effects are understood as degrees of achieving conflict resolution. The model uses both formal (e.g. dialogue systems) and linguistic (e.g. topical relevance) concepts.

2 Goals of dialogues with conflict resolution

In this section, we explore what goals the dialogues with conflict resolution have. We limit our considerations to the initial goal of a dialogue, i.e. the goal related to the intention of initiating the dialogue. We start our consideration with the model proposed by Walton (Section 2.1) and then propose its extension which allows to enhance expressivity and applicability of this model (Section 2.2).

2.1 Limitations of Walton's model

Walton's model was originally presented in (Walton, 1989) and further developed in (Walton, 1995; Walton and Krabbe, 1995). Walton's work is broadly studied in linguistics and selectively discussed in social psychology, while Walton & Krabbe's work is widely applied in the formal dialogue systems and MAS.

In this model, the persuasion dialogue (also called critical discussion originally introduced by

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

(van Eemeren and Grootendorst, 1984)) is the only type of dialogue that is meant to be related to conflict resolution. There are other types of dialogues that have something in common with conflict, however, their goal is not to resolve it. Negotiation aims for conflict settlement, and eristics aims for reaching a (provisional) accommodation. The other types of dialogues do not refer to conflict at all (Walton and Krabbe, 1995, 80-81).

Walton specifies the dialogues by means of three properties: initial situation, main goal and participant's aims. The initial situation of the persuasion is a conflict of opinion. The main goal is to resolve this conflict by verbal means. The aim of each participant is to persuade the other party to take over its point of view (see Table 1).

initial state	**main goal**	**agent's goal**
conflicting points of view	resolution of conflict	persuade the other(s)

Table 1. The properties of persuasion dialogue from (Walton and Krabbe, 1995).

This model relates the persuasion dialogue to the issue of conflict resolution. However, it has some serious limitations if we want to analyze this issue in a more detailed manner. Below, we discuss three of those limitations.

Limitation 1: conflict's object. The first criticism refers to the types of conflicts identified in Walton's model. It distinguishes only two types of conflict - with respect to opinion (the conflict specific for persuasion) and interest (the conflict specific for negotiation). Since the goal of negotiation is not to resolve the conflict, the only dialogues with conflict resolution considered in this model are dialogues that aim for agreeing on opinions (i.e. resolving the conflict of opinion).

On the other hand, in the literature related to applications of dialogues allowing for conflict resolution, many other objects of conflict are distinguished and studied. Besides conflict between opinions (see (Prakken, 2006) for an overview), conflicts concern e.g. attitudes (Pasquier et al., 2006), actions (Bench-Capon et al., 2005), behavior (Sierra et al., 1997), intentions (Dignum et al., 2001), plans (Tang and Parsons, 2005), preferences (Sycara, 1990) or permissions for gaining access to information (Perrussel et al., 2007). The model restricted to conflict of opinion has, thus, a strong limitation in expressivity and applicability.

Limitation 2: the meaning of "main goal". The next limitation refers to the ambiguity of the notion "main goal". It is not clear if the main goal means: (1) that resolution of conflict is the basic, but still *individual* aim of an agent, while his secondary aim is to persuade the other agent, or (2) it means the goal of the *system* of all participants (i.e. the joint goal of agents as a group).

Throughout the paper, we use the word "system of agents" to denote the group of individuals (human or artificial). That is, a system may be a group of people (e.g. a council of doctors or a council of war), as well as a multi-agent system.

It seems that in Walton's model meaning (1) is assumed:[1]

> We must distinguish between the primary or main goal of a type of dialogue and the aims of the participants (...). Thus, the primary goal of negotiation could be characterized as "making a deal." By entering into negotiations the parties implicitly subscribe to this overall purpose. But, besides, each party pursues, within the dialogue, the particular aim of getting the best out of it for oneself (Walton and Krabbe, 1995, 67).

A negative consequence of this interpretation is discussed in Section 2.2 (see Proposition 1).

Limitation 3: the scope of persuasion's goal. The last criticism refers to the relation between the main goal and participants' aims. Assume the first meaning of the main goal described above (i.e., the main goal means a primary, individual goal). Two interpretations of the relation between the main goal and the participant's goal are possible: (1) the narrow one: the persuasion dialogue has to fulfill *both* of those goals, and (2) the broad one: the persuasion has to fulfill *at least one* of them.

Both of the interpretations generate some problems. If interpretation (1) is assumed, then collective methods of conflict resolution are inexpressible in Walton's model (i.e when agents aim to resolve a conflict and they do not care about their individual victories). From the point of view of its important applications such as MAS, it is a strong limitation. The multi-agents systems have some tasks to perform (e.g. to control the temperature in a building). A conflict among agents may be an

[1] Consider the goals of persuasion in analogy to negotiation.

obstacle in accomplishing those tasks: "Finding ways for agents to reach agreements in multiagent systems is an area of active research" (Parsons and Sklar, 2005, 297). Typically, in MAS persuasion represented by Walton's model is used as a tool for conflict resolution. As a result, the collective method for conflict resolution is entirely excluded from the studies. Since in the narrow interpretation an agent has to fulfill both of the goals, then he is restricted to adopt an egoistic strategy. This makes impossible to express the cooperative methods of conflict resolution which from the viewpoint of MAS should be equally (or maybe even more) desirable.

On the other hand, if interpretation (2) is assumed, then both of these methods of conflict resolution (egoistic and collective) are describable in Walton's model, however, they are undistinguishable. If the word "persuasion dialogue" denotes both of these strategies, then some additional subclasses of persuasion should be introduced to refer to a particular type of strategy.

In (Walton and Krabbe, 1995, 66), the authors indicate that their aim is not to propose the exhaustive typology of dialogues. However, the scope of applicability of this typology (e.g. in MAS) narrows down the functionality of the models which originate from the distinction. Therefore, despite its pioneering and important advances in the formal dialectics, Walton's model needs to be modified and extended such that limitations 1-3 could be avoided.

2.2 Extension of Walton's model

In this section, we propose the extension of Walton's (1989) model. Let $Agt = \{1, \ldots, n\}$ be a set of names of *agents*. Our model is built upon and uses the standard notions from the formal systems of persuasion dialogue (see e.g. (Prakken, 2006)), in particular the model proposed in (Prakken, 2005).[2] Let L_t be a *topic language* (a logical language including e.g. $p, \neg p$), and L_c be a *communication language* (a set of locutions including e.g. $claim\ p$, $why\ p$). Each agent maintains a list of utterances, called the *commitment store*. Intuitively, commitments are what an agent publicly declares as his beliefs, attitudes, intentions, plans, etc. A set of an agent i's commitments at a stage d of a dialogue is denoted by $C_d(i)$ (for $i \in Agt$). For example, if i makes a move $claim\ p$ at a dialogue stage d, then p is placed in his commitment store, i.e. $p \in C_d(i)$.

We introduce some simplifications for the clarity of presentation. First, we limit our considerations to a system of agents S which consists of two players in a dialogue, i.e. $S = \{i, j\} \subseteq Agt$. We use a symbol $\bar{i}$ to denote an agent i's adversary. It means that when $i \in prop(t)$ (i is a proponent for t), then $\bar{i} \in opp(t)$ ($\bar{i}$ is an opponent against t). Moreover, we assume that a conflict refers to one object (one belief, one attitude, etc.). This assumption corresponds to a single type of dialogue (van Eemeren and Grootendorst, 1984, 80).

Conflict. First we specify the notion of conflict. The conflict in relation to t is denoted by $\{t, \bar{t}\}$ ($\bar{t}$ denotes opposing standpoint to t). A topic t may refer to different objects, e.g. opinions, attitudes, actions, intentions, preferences, and so on. Moreover, for simplicity we do not consider "neutral" commitments, i.e. we assume that there is not such t that $t \notin C_d(i)$ and $\bar{t} \notin C_d(i)$. Consequently, $\{t, \bar{t}\}$ means that one of agents is committed to t, while his opponent is committed to $\bar{t}$, i.e. $t \in C_d(i)$ and $\bar{t} \in C_d(\bar{i})$.

Let $S = \{i, j\} \subseteq Agt$ be a system of agents, and $t \in L_t$ be a topic of disagreement between agents. We say that after a dialogue d a conflict $\{t, \bar{t}\}$ is resolved, when either both agents are committed to t, or both are committed to $\bar{t}$:

Definition 1 After execution of dialogue d, *conflict* $\{t, \bar{t}\}$ between agents in S *is resolved*, when exactly one of the following conditions holds: (1) $t \in C_d(i)$ and $t \in C_d(\bar{i})$, or (2) $\bar{t} \in C_d(i)$ and $\bar{t} \in C_d(\bar{i})$.

Observe that in this definition the disjunction is exclusive, i.e. it holds only if exactly one of the elements that it connects is true (either first one is true and second - false, or the opposite way).

Conflict resolution for system of agents. Now we can specify the class of dialogues in which the conflict resolution is the goal of the whole *system of agents*. We use the notion "conflict resolution" instead of "dialogue with conflict resolution", if there is no danger for confusion.

Definition 2 Dialogue d is *conflict resolution* for S in relation to t (denoted by $d \in \mathsf{CR}_{S,t}$), when

[2] In the paper, a detailed formal specification for a dialogue system is not needed; for the full details the reader is referred to the references.

the goal of S is to resolve conflict $\{t, \bar{t}\}$ between agents in S after execution of d.

We assume that a system's goal is a joint purpose of all agents in the system. For example, in MAS this goal may be a result of a need to accomplish a task by a system (e.g. to control the temperature in a building). Since a conflict is an obstacle for a system's joint performance, its goal will be to resolve this conflict. Observe that humans may be in some sense unaware of their joint purpose. Imagine that Bob and Ann have a walk in the mountains. They approach the crossroad. Ann thinks that they should go left, and Bob claims they should go right. They may be extremely competitive and willing to persuade the other party to take the path each of them have chosen, however, still their joint goal as a group is to move further and eventually come back home safely.

Conflict resolution for individuals. From the point of view of the system's members, the very same goal (of conflict resolution) may be accomplished in different manners. An agent may adopt one of the following individual strategies of fulfilling the system's goal: persuasive (egoistic), collaborative or passive.

The agent has a persuasive goal if he is interested only in such an outcome of a dialogue in which his standpoint wins. That is, if in conflict $\{t, \bar{t}\}$ an agent i is committed to t, then i executes persuasion when his goal is to make $\bar{i}$ be committed to t.

Definition 3 Conflict resolution d is *persuasion for agent* $i \in S$ in relation to t (denoted by $d \in \mathsf{CR}_{S,t}^{Per_i}$), when the goal of i is that after execution of d, it holds: $t \in C_d(i)$ and $t \in C_d(\bar{i})$.

The second strategy that an agent may adopt to achieve a system's goal is collaborative. Intuitively, the agent has a collaborative goal if he is interested in such an outcome of a dialogue in which any party wins. That is, i is collaborative when his goal is to reach an agreement regardless of whether both agents will be committed to t or both will be committed to $\bar{t}$:

Definition 4 Conflict resolution d is *collaborative dialogue for agent* $i \in S$ in relation to t (denoted by $d \in \mathsf{CR}_{S,t}^{Col_i}$), when the goal of i is to resolve conflict $\{t, \bar{t}\}$ between agents in S after execution of d.

Such dialogues are specific for cooperative systems, while persuasion is typical for adversarial domains. When doctors disagree and discuss what method of treatment to apply in a particular case, then they may adopt collaborative strategy of conflict resolution. On the other hand, when an insurance agent and a client disagree and discuss the type of insurance the client should buy, then the agent (typically) adopts egoistic strategy.

The goals of persuasion dialogues such as informativeness and the increase in understanding through the performance of maieutic function (Walton, 1995, 102-103) are specific for collaborative rather than for persuasive $\mathsf{CR}_{S,t}$ (see (Felton et al., 2009) for experimental results). Still, they are not the intended goals of $\mathsf{CR}_{S,t}$ but constitute some "extra" value.

Observe that persuasive and collaborative goals are not two mutually-exclusive strategies, but rather the prototypical forms representing the extremes of an intention continuum. While in MAS agents can be designed to behave according to given definitions, then in natural contexts humans may adopt strategies located somewhere between those two extremes (a person may be more or less persuasive, or collaborative).

The last individual "strategy" in conflict resolution is passive. Intuitively, the agent is passive in d, if he has no goal. He only reacts to the moves executed by his adversary. In other words, such an agent is an opponent against t, but not a proponent for $\bar{t}$ (i.e., $i \in opp(t)$, but $i \notin prop(\bar{t})$). This type of goal can be described as follows: "the other participant has a role of raising critical questions that cast doubt on that thesis" (Walton, 1995, 100).

Definition 5 Conflict resolution d is *passive dialogue for agent* $i \in S$ in relation to t (denoted by $d \in \mathsf{CR}_{S,t}^{\emptyset_i}$), when i has no goal of executing d.

Note that in $d \in \mathsf{CR}_{S,t}$, at least one agent cannot be passive.

Subclasses of conflict resolution. In the formal language of dialogue systems, no symbol representing a goal is specified. To account for our considerations, we introduce a preliminary version of its specification. Let $G_{i,t}(d)$ be a set of i's goals in a dialogue $d \in \mathsf{CR}_{S,t}$. We can distinguish the following types of system's conflict resolution:

- if ($G_{i,t}(d) = \{Per\}$ or $G_{i,t}(d) = \{Col\}$)[3], and $G_{\bar{i},t}(d) = \emptyset$, then d is asymmetric $\mathsf{CR}_{S,t}$ (which corresponds to simple dialogue derived from (van Eemeren and Grootendorst, 1984, 80));
- if ($G_{i,t}(d) = \{Per\}$ and $G_{\bar{i},t}(d) = \{Per\}$), or ($G_{i,t}(d) = \{Col\}$ and $G_{\bar{i},t}(d) = \{Col\}$), then d is symmetric $\mathsf{CR}_{S,t}$ (which corresponds to compound dialogue derived from (van Eemeren and Grootendorst, 1984, 80));
- if $G_{i,t}(d) = \{Per\}$ and $G_{\bar{i},t}(d) = \{Col\}$, then d is mixed $\mathsf{CR}_{S,t}$.

Recall that we discussed a problem with interpretation of "main goal" in Walton's model (Section 2.1). We argued that the goals of persuasion (plausibly) mean both resolving a conflict and persuading the other agent (see Limitation 3). Moreover, those goals seemed to be individual aims of the dialogue's participant (see Limitation 2). If this is the case, then an agent has two individual goals, i.e. $\{Per\} \subseteq G_{i,t}(d)$ and $\{Col\} \subseteq G_{i,t}(d)$. Then, Walton's persuasion would be a multi-goal dialogue. However, it can be easily shown that the multi-goal d reduces to the single-goal d of persuasive conflict resolution (for a given agent):

Proposition 1 Let $i \in S \subseteq Agt$ and $t \in L_t$. If $d \in \mathsf{CR}_{S,t}$ and $\{Per\} \subseteq G_{i,t}(d)$ and $\{Col\} \subseteq G_{i,t}(d)$, then $G_{i,t}(d) = \{Per\}$.

To prove the proposition formally, we would need to introduce more precise specifications. For example, we would have to decide how we want to understand goals (e.g. as formulas (Budzynska et al., 2009), or as states (Tokarz, 1985; Tang and Parsons, 2005)). However, such a level of formalization is outside the scope of this paper. Instead, we will give the intuitions for such a proof. Say that an agent's goal in a dialogue is understood as a set of states which could be reached after the dialogue. Then, the goal Per is a set of states where the following condition is satisfied: $t \in C_d(i)$ and $t \in C_d(\bar{i})$. Further, the goal Col is a set of states where the exclusive disjunction of the two conditions is satisfied: ($t \in C_d(i)$ and $t \in C_d(\bar{i})$) or ($\bar{t} \in C_d(i)$ and $\bar{t} \in C_d(\bar{i})$). If $\{Per\} \subseteq G_{i,t}(d)$ and $\{Col\} \subseteq G_{i,t}(d)$, then a goal set of states $G_{i,t}(d)$ have to be an intersection of two of the sets described above. Since the condition for Per is the first element of the exclusive disjunction for Col, then only this element will be true. It means that $G_{i,t}(d)$ is reduced to the set $\{Per\}$.

Consequently, in Walton's model the main goal of conflict resolution is an (unintended) result of an individual persuasive goal adopted by an agent, rather than an additional (primary, main) goal of this agent. Moreover, a collaborative conflict resolution cannot be defined within Walton's model. The solution that we propose in the paper is to extend this model with the separation of system's goal from agents' goals, and the distinction between different individual strategies of reaching system's goal.

To conclude, the extension of Walton's model generates five subclasses of the dialogues with conflict resolution for a system of agents:

$$\mathsf{CR}_{S,t} = (\mathsf{CR}_{S,t}^{Per_i} \cap \mathsf{CR}_{S,t}^{\emptyset_{\bar{i}}}) \cup (\mathsf{CR}_{S,t}^{Per_i} \cap \mathsf{CR}_{S,t}^{Per_{\bar{i}}}) \cup (\mathsf{CR}_{S,t}^{Col_i} \cap \mathsf{CR}_{S,t}^{\emptyset_{\bar{i}}}) \cup (\mathsf{CR}_{S,t}^{Col_i} \cap \mathsf{CR}_{S,t}^{Col_{\bar{i}}}) \cup (\mathsf{CR}_{S,t}^{Per_i} \cap \mathsf{CR}_{S,t}^{Col_{\bar{i}}}).$$

The first subclass of conflict resolution is asymmetric persuasive $\mathsf{CR}_{S,t}$, the second one - symmetric persuasive $\mathsf{CR}_{S,t}$, the third - asymmetric collaborative $\mathsf{CR}_{S,t}$, the forth - symmetric collaborative $\mathsf{CR}_{S,t}$, and the last one - mixed $\mathsf{CR}_{S,t}$. Observe that Walton's model allows to express and explore only two first subclasses.

3 Effects of conflict resolution

This section determines the types of the effects which can be achieved through a dialogue with the pre-planned conflict resolution. Four degrees of accomplishing conflict resolution are distinguished and exemplified: fully unsuccessful dialogue (Section 3.1), partially successful dialogue (Section 3.2), fully successful dialogue (Section 3.1) and over-successful dialogue (Section 3.4).

3.1 Unsuccessful vs. successful dialogue

Recall that $C_d(i)$ means agent i's commitment store at the stage of dialogue d. Following our specification for the goals in dialogues with conflict resolution, unsuccessful and successful $\mathsf{CR}_{S,t}$ may be defined. Let $S = \{i, j\} \subseteq Agt$ and $t \in L_t$.

Definition 6 Conflict resolution $d \in \mathsf{CR}_{S,t}$ is (fully) *unsuccessful for system* S in relation to t, when conflict $\{t, \bar{t}\}$ between agents in S is not resolved after execution of d.

[3] For simplicity, we do not introduce the precise specification for goals. Intuitively, $G_{i,t}(d) = \{Per\}$ means that i has persuasive goal in d with respect to t.

Since in Definition 1 the disjunction was exclusive, $\mathsf{CR}_{S,t}$ is fully unsuccessful for S in two cases: (1) if $t \in C_d(i)$, and $\bar{t} \in C_d(\bar{i})$, or (2) if $\bar{t} \in C_d(i)$, and $t \in C_d(\bar{i})$.

Definition 7 Dialogue $\mathsf{CR}_{S,t}^{Per_i}$ is (fully) *unsuccessful persuasion for agent* i in relation to t, when after execution of d, it holds: $\bar{t} \in C_d(i)$ or $\bar{t} \in C_d(\bar{i})$.

That is, a persuasive conflict resolution is fully unsuccessful for i in three cases: (1) if $t \in C_d(i)$ and $\bar{t} \in C_d(\bar{i})$, or (2) if $\bar{t} \in C_d(i)$ and $t \in C_d(\bar{i})$, or (3) if $\bar{t} \in C_d(i)$ and $\bar{t} \in C_d(\bar{i})$. In the cases (1) and (2), an agent i is unsuccessful, since the goal of the system is not accomplished. Thus, even though in (2) i managed to make $\bar{i}$ be committed to i's initial standpoint t, he failed to resolve the conflict. In case (3), even though the conflict is resolved, i is unsuccessful, since he failed to make $\bar{i}$ be committed to t.

Definition 8 Dialogue $\mathsf{CR}_{S,t}^{Col_i}$ is (fully) *unsuccessful collaborative dialogue for agent* i in relation to t, when conflict $\{t, \bar{t}\}$ between agents in S is not resolved after execution of d.

A passive conflict resolution for an agent i cannot be successful or unsuccessful for i, since a set of i's goals is empty. In other words, there is no goal to achieve for this agent in a given dialogue.

The definitions of successful dialogues are analogical to Definitions 6-8:

Definition 9 Conflict resolution $d \in \mathsf{CR}_{S,t}$ is (fully) *successful for system* S in relation to t, when conflict $\{t, \bar{t}\}$ between agents in S is resolved after execution of d.

Definition 10 Persuasion $\mathsf{CR}_{S,t}^{Per_i}$ is (fully) *successful for agent* i in relation to t, when after execution of d, it holds: $t \in C_d(i)$ and $t \in C_d(\bar{i})$.

Interestingly, when $d \in \mathsf{CR}_{S,t}^{Per_i}$ accomplishes individual goal of i, then d accomplishes a system's goal. Observe that the relationship in the opposite direction does not hold. That is, if a system's goal is accomplished, the persuasive goal of i does not have to be fulfilled, since $\bar{i}$ could be successful.

Definition 11 Collaborative dialogue $\mathsf{CR}_{S,t}^{Col_i}$ is (fully) *successful for agent* i in relation to t, when conflict $\{t, \bar{t}\}$ between agents in S is resolved after execution of d.

Clearly, if i accomplishes a collaborative goal, then he will accomplish the goal of a system. In this case, the relationship in the opposite direction does hold.

3.2 Fully unsuccessful vs. partly successful conflict resolution

In formal approach, it is not possible to differentiate partially successful $\mathsf{CR}_{S,t}$ from the fully unsuccessful $\mathsf{CR}_{S,t}$. Therefore, a need arises to define partially successful dialogues in pragmatic terms, using both Cartesian and non-Cartesian approach. Non-Cartesian approach relies on ratio-empirical pragmatics which allows for gradualistic reasoning and non-discreteness (Walton, 1995, 158)(Kopytko, 2002). Cartesian approach, as (Kopytko, 2002, 523) indicates, refers to 'discreteness/categoriality of pragmatic phenomena'. To define partially successful dialogues we need not only discrete terminology (such as 'move', 'goal', 'agent'), but also non-discrete procedures and concepts (such as 'gradual reasoning', 'topical relevance'). Consider the following examples (where i_1 means a first move in a dialogue performed by agent i):

(d^1) Bob_1: Let's go to the cinema today.
Ann_2: No.

(d^{1a}) Bob_3: But Avatar is playing at the Odeon.
Ann_4: OK, but we'll go tomorrow. I have no time today.

(d^{1b}) Bob_3: So let's go to the theatre.
Ann_4: OK, let's go.

Dialogue d^{1a} and dialogue d^{1b} are possible continuations of dialogue d^1. They are sequential procedures in which partial conflict resolution can be described in terms of the non-Cartesian approach. At the last stage of d^{1a} and d^{1b}, Ann is not committed to statement expressed in move Bob_1. At this point, it is essential to distinguish between the goal of the system of conflict resolution and the topic (t) of the conversation. In each case, topic t relates to the positive attitude to going to the cinema today while the goal of the system is a conflict resolution on topic t. The effectiveness of the achievement of conflict resolution is different in each case. In d^1 Ann rejects topic t and thus conflict resolution is

not achieved. In d^{1a} and d^{1b}, however, the partial accomplishment of conflict resolution occurs due to the agreement with statement in move Bob_3. It is evident in d^{1a} and d^{1b} that the moves refer to a set of topics $T_{i,t}$ which specify the area of i's interest of conflict resolution with respect to t. The commitment of agent i to one of the possible topics from the set $T_{i,t}$ (e.g. positive attitude to going to the cinema tomorrow not today) in his last turn manifests topical relevance and partial accomplishment of conflict resolution.

In dialogue d^{1b}, Bob divides his conflict resolution into submoves. As the achievement of conflict resolution is gradually strived for, in $move_3$ Bob attempts to remain within the set of topic $T_{Bob,t}$. Although topic t is rejected by Ann in $move_2$, Bob in the statement in move Bob_3 still tries to get some benefit for himself and be relevant. Ann treats the positive attitude to going to the theatre (move Bob_3) as an allowable alternative within the set of topics $T_{Bob,t}$. Only if both move Bob_1 and move Bob_3 were acclaimed by Ann, the whole conflict in d^{1b} would be successfully resolved. The accomplishment of part of i's conflict resolution occurs since only the statement in $move_3$ is accepted. Treating conflict resolution as related to a set of topics $T_{i,t}$ points to non-discreteness of the process of achieving conflict resolution.

3.3 Over-successful dialogue

Similarly, in formal approach there is no way to express over-successful dialogue. We have to take into account that the topics in $T_{i,t}$ differ in the degree of acceptance and degree of importance for agent i. If conflict resolution is achieved due to topical relevance and agent i is not only completely satisfied with the last move of agent $\bar{i}$ but also attributes the high degree of importance to it, then we can talk about over-successful conflict resolution.

We propose to draw the distinction between radial topics and prototype topics. The distinction is motivated by the Lakoff's categorization of concepts into prototypical and radial ones (Lakoff, 1987). In the approach we advocate, prototypical and radial categories should not exclusively relate to single concepts, but also to topics (opinions, attitudes, actions, intentions, preferences, etc.). Prototype topics manifest essential, stereotypical and salient examples of topic t. Radial topics are indirectly concerned with the prototype topics. It means that depending on the context they can relate to the prototype topics or not. This can be expressed by the formula $T_{i,t} = TP_{i,t} \cup TR_{i,t}$ in which $T_{i,t}$ is the set of topics relevant to t, $TP_{i,t}$ are the prototype topics relevant t and $TR_{i,t}$ are the radial topics indirectly relevant to t. If the last move of Ann manifests both prototype relevance ($TP_{i,t}$) and radial relevance ($TR_{i,t}$) and the radial relevance has a high degree of importance for agent i, then we can talk about over-successful $\mathsf{CR}_{S,t}$. Consider dialogue d^2:

(d^2) Bob_1: Let's go to the cinema today.
Ann_2: Why?
Bob_3: Avatar is playing at the Odeon.
Ann_4: OK and I'll invite you for dinner afterwards.
Bob_5: OK.

In d^2, $move_4$ of Ann manifest both $TP_{i,t}$ (the positive attitude to going to the cinema to see Avatar today) and $TR_{i,t}$ (the positive attitude to going for dinner afterwards). Since $TR_{i,t}$ has a high degree of importance and acceptance for agent i we can observe over-successful $\mathsf{CR}_{S,t}$.

4 Conclusions

Dialogues with conflict resolution play an important role in different contexts. In MAS, the most important issue is the realization of system's tasks. A $\mathsf{CR}_{S,t}$ dialogue enables to resolve a disagreement on t, which enhances a cooperative accomplishment of S's tasks. The minor issue is how the resolution of a conflict is achieved - either in an egoistic or a collaborative way. The conflict resolution is also important in educational setting, since the collaborative goal of performing a dialogue supports teaching students of knowledge construction through argumentation in science classrooms (Felton et al., 2009).

From this point of view, constraining a model to the subclass of dialogues that aim to resolve a conflict in an egoistic manner is a serious limitation. In Proposition 1, we show that the main (individual) goal of resolving the conflict is reducible to the secondary individual goal of persuading the other party. As a result, in Walton's model there is no reason to consider the property of the main goal defined in such a way, since it does not provide any additional information to the dialogue's specification beyond that information which is provided by the property of the persuasive individ-

ual goal. Thus, it is not clear in what sense Walton talks about the "collective goal" (contrasting it with the "individual goals") or the cooperativeness as a property of persuasion dialogue (Walton, 1995, 101).

In the paper, we propose the extensions that allow to specify dialogues in which the superior goal is to resolve a conflict. In our model, this goal may be achieved by individuals in different manners. First, an agent may be persuasive, i.e. he may be interested only in such a resolution in which his standpoint wins. An agent may also choose the collaborative goal, i.e. he may aim at any outcome which brings a resolution of conflict (no matter if his or the opponent's standpoint wins). Finally, an agent may be passive and only react to the other party's dialogue moves. Depending on the type of the individual goal, different strategy will be adopted by an agent. It means that we may need to specify a distinct formal dialogue system for each of the five subclasses of conflict resolution. The model proposed in this paper does not suffer from the problem discussed in Proposition 1, since we differentiate the goal of the system of agents from the goals of its members, instead of distinguishing two types of individual goals as assumed in Walton's model.

We also specify the four different types of effects in $\mathsf{CR}_{S,t}$. The dialogue may be fully successful (or unsuccessful), when a given goal (of a system, persuasive or collaborative) is fulfilled (or not fulfilled, respectively). Moreover, we explore such an effect when a goal is not achieved, however, an agent still "benefits" to a certain degree, as well as the effect when an agent achieves more that he initially intended. To identify the partially successful and over-successful dialogues, we use the linguistic concepts of topical relevance and gradualistic reasoning.

References

Trevor Bench-Capon, Katie Atkinson, and Alison Chorley. 2005. Persuasion and Value in Legal Argument. *Journal of Logic and Computation*, (15):1075–1097.

Katarzyna Budzynska, Magdalena Kacprzak, and Pawel Rembelski. 2009. Logic for reasoning about components of persuasive actions. *Proc. of 18th ISMIS09*, LNAI(5722):201–210.

Frank Dignum, Barbara Dunin-Keplicz, and Rineke Verbrugge. 2001. Creating collective intention through dialogue. *Logic Journal of the IGPL*, (9):289–303.

Frans van Eemeren and Rob Grootendorst. 1984. *Speech acts in argumentative discussions*. Dordrecht: Foris Publications.

Mark Felton, Merce Garcia-Mila and Sandra Gilabert. 2009. Deliberation versus Dispute: The Impact of Argumentative Discourse Goals on Learning and Reasoning in the Science Classroom. *Informal Logic*, 29(4).

Roman Kopytko. 2002. *The mental aspects of pragmatic theory: An integrative view*. Poznan: Motivex.

George Lakoff. 1987. *Women, Fire, and Dangerous Things*. University of Chicago Press.

Simon Parsons and Elizabeth Sklar. 2005. How agents revise their beliefs after an argumentation-based dialogue. *Proc. of ArgMAS05*.

Philippe Pasquier, Iyad Rahwan, Frank Dignum, and Liz Sonenberg. 2006. Argumentation and Persuasion in the Cognitive Coherence Theory. *Proc. of COMMA06*.

Laurent Perrussel, Sylvie Doutre, Jean-Marc Thevenin, and Peter McBurney. 2007. A persuasion dialog for gaining access to information. *Proc. of ArgMAS07*.

Henry Prakken. 2005. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, (15):1009–1040.

Henry Prakken. 2006. Formal systems for persuasion dialogue. *The Knowledge Engineering Review*, (21):163–188.

Carles Sierra, Nicholas R. Jennings, Pablo Noriega, and Simon Parsons. 1997. A Framework for Argumentation-Based Negotiation. *Proc. of the 4th International Workshop on Intelligent Agents*, LNCS (1365):177–192.

Katia Sycara. 1990. Persuasive argumentation in negotiation. *Theory and Decision*, (28):203-242.

Yuqing Tang and Simon Parsons. 2005. Argumentation-based dialogues for deliberation. *Proc. of ArgMAS05*.

Marek Tokarz. 1985. Goals, results and efficiency of utterances. *Bulletin of the Section of Logic*, 4(14):150-155.

Douglas Walton. 1989. *Informal Logic*. Cambridge: Cambridge Univ. Press.

Douglas Walton. 1995. *A Pragmatic Theory of Fallacy*. The Univ. of Alabama Press.

Douglas Walton and Erik Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of N.Y. Press.

Erotetic Search Scenarios: Revealing Interrogator's Hidden Agenda

Mariusz Urbański
Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland
Mariusz.Urbanski@amu.edu.pl

Paweł Łupkowski
Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland
Pawel.Lupkowski@amu.edu.pl

Abstract

The way of formal modeling of a hidden agenda of an interrogator is described in terms of Inferential Erotetic Logic. Two examples are given: one is based on a simple detective story, the other is based on an analysis of a judge's strategy in the Turing Test.

1 Introduction: erotetic basis

Wiśniewski (2003) defines erotetic search scenarios (e-scenarios for short) within the framework of Inferential Erotetic Logic (IEL) as a possible technique for solving problems expressed by questions. He claims that:

> When a problem is expressed by a question which has a well-defined set of direct[1] [...] answers, one can [...] apply an e-scenario in order to find the solution to the problem. Viewed pragmatically, an e-scenario provides us with conditional instructions which tell us what questions should be asked and when they should be asked. Moreover, an e-scenario shows where to go if such-and-such a direct answer to a query appears to be acceptable and does so with respect to any direct answer to each query (Wiśniewski, 2003, p. 422).

Thus an e-scenario may be interpreted as a plan for an interrogation (the questioned being a human, a database, an Oracle etc.) that describes a "hidden agenda" of an interrogator.

Suppose that I am questioning a certain suspect in order to determine if this person is guilty or not (for definiteness let my problem be expressed by a question: "Who stole the tarts?"). In such a situation addressing the question directly may not be the most brilliant idea, unless the suspect is willing to plead guilty. If I am interested in something more than a declaration of a person in question I have to seek for a more or less indirect solution by gathering evidence and by making inferences on its basis. Description of this evidence and a plan for further inferences forms in this case my hidden agenda. It expounds *a*) initial information relevant to the case and *b*) inferential steps made on its basis. Inferential steps involved are possibly of two kinds: standard declarative ones and erotetic ones. Erotetic inferences are these in which questions play the role of conclusion and/or premises. Questions arise when there is a gap in available information (initial or derived), but from the investigator's point of view it is important to pose only such auxiliary questions that are both informative and cognitively useful, that is, answers to which are helpful in answering the initial question. This may be formally explicated in terms of erotetic implication, an erotetic counterpart to the entailment relation (Wiśniewski, 1995):[2]

Definition 1 *A question Q* implies *a question Q^** on the basis of *a set of d-wffs X (in symbols:*

[1] *Direct answers* are the answers which "are directly and precisely responsive to the question, giving neither more nor less information than what is called for" (Belnap, 1969, p. 124). For the sake of generality they may be called *principal possible answers* (Wiśniewski and Pogonowski, 2010).

[2] Our language is the language of First-order Logic enriched with question-forming operator ? and brackets $\{,\}$ (call this language $\mathbf{L}$). Well formed formulas of FoL (defined as usual; additionally, we allow for names of formulas to appear as arguments of predicate symbols) are *declarative well-formed formulas* of $\mathbf{L}$ (d-wffs for short). Expressions of the form $?\{A_1, \ldots, A_n\}$ are *questions* or *erotetic formulas* of $\mathbf{L}$ (e-formulas for short) provided that $A_1, \ldots, A_n$ are syntactically distinct d-wffs and that $n > 1$. The set $\mathbf{d}Q = \{A_1, \ldots, A_n\}$ is the set of all the direct answers to the question $Q = ?\{A_1, \ldots, A_n\}$. Thus an erotetic formula $?\{A, \neg A\}$ expresses a simple yes-no question: "Is it the case that A or is it the case that $\neg A$?"; this kind of questions we shall abbreviate by $?A$. Let also the symbol T stand for any logically valid formula. Intuitively, T stands for the lack of factual knowledge: a question of the form $?\{A_1, \ldots, A_n, \mathsf{T}\}$ reads "Is it the case that A_1 or ... or is it the case that A_n or no required information is available?" (Wiśniewski, 2007).

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

$Im(Q, X, Q^*)$) *iff*

1. *for each direct answer A to the question Q: $X \cup \{A\}$ entails the disjunction of all the direct answers to the question Q^*, and*

2. *for each direct answer B to the question Q^* there exists a non-empty proper subset Y of the set of direct answers to the question Q such that $X \cup \{B\}$ entails the disjunction of all the elements of Y.*

If $X = \emptyset$, then we say that Q implies Q^* and we write $Im(Q, Q^*)$.

The first condition requires that if the implying question is sound[3] and all the declarative premises are true, then the implied question is sound as well[4]. The second condition requires that each answer to the implied question is potentially useful, on the basis of declarative premises, for finding an answer to the implying question. To put it informally: each answer to the implied question Q^*, on the basis of X, narrows down the set of plausible answers to the implying question Q.

Consider a simple example. My initial question is;

(Q) Who stole the tarts?

Suppose that I manage to establish the following evidence:

(E_1) It is one of the courtiers of the Queen of Hearts attending the afternoon tea-party who stole the tarts.

Thus my initial question together with the evidence implies the question:

(Q^*) Which of the Queen of Hearts' courtiers attended the afternoon tea-party?

If moreover I know that:

(E_2) Queen of Hearts invites for a tea-party only these courtiers who made her laughing the previous day.

then Q^* and E_2 imply the question:

(Q^{**}) Which courtiers made the Queen of Hearts laughing the previous day?

[3] A question Q is *sound* iff it has a true direct answer (with respect to the underlying semantics).

[4] This property may be conceived as an analogue to the truth-preservation property of deductive schemes of inference.

Erotetic search scenarios may be defined as sets of so-called erotetic derivations (Wiśniewski, 2003) or, in a more straightforward way, as finite trees (Wiśniewski, 2010, p. 27–29):

Definition 2 *An* e-scenario for a question Q relative to a set of d-wffs X *is a finite tree Φ such that:*

1. *the nodes of Φ are (occurrences of) questions and d-wffs; they are called e-nodes and d-nodes, respectively;*

2. *Q is the root of Φ;*

3. *each leaf of Φ is a direct answer to Q;*

4. *$\mathbf{d}Q \cap X = \emptyset$;*

5. *each d-node of Φ:*
   (a) *is an element of X, or*
   (b) *is a direct answer to an e-node of Φ different from the root Q, or*
   (c) *is entailed by (a set of) d-nodes which precede the d-node in Φ;*

6. *for each e-node Q^* of Φ different from the root Q:*
   (a) *$\mathbf{d}Q^* \neq \mathbf{d}Q$ and*
   (b) *$Im(Q^{**}, Q^*)$ for some e-node Q^{**} of Φ which precedes Q^* in Φ, or*
   (c) *$Im(Q^{**}, \{A_1, ..., A_n\}, Q^*)$ for some e-node Q^{**} and some d-nodes $A_1, ..., A_n$ of Φ that precede Q^* in Φ;*

7. *each d-node has at most one immediate successor;*

8. *an immediate successor of an e-node different from the root Q is either a direct answer to the e-node, or exactly one e-node;*

9. *if the immediate successor of an e-node Q^* is not an e-node, then each direct answer to Q^* is an immediate successor of Q^*.*

Definition 3 *A* query *of an e-scenario Φ is an e-node Q^* of Φ different from the root of Φ and such that the immediate successors of Q^* are the direct answers to Q^*.*

We shall elaborate the idea of representing interrogator's hidden agenda via e-scenarios on two examples. The first one is a simple detective story based on a Smullyan's (1978) logical puzzle. The second one is an analysis of a judge's strategy in the Turing Test.

2 Vampires, zombies and humans

On a certain island the inhabitants have been bewitched by some kind of magic. Half of them turned into zombies, the other half turned into vampires. The zombies and the vampires of this island do not behave like the conventional ones (if any): the zombies move about and talk in as lively a fashion as do the humans, and the vampires even prefer drinking strong mocca over anything else. It's just that the zombies of this island always lie and the vampires of this island always tell the truth[5]. What is also important, both vampires and zombies never miss a reasonable opportunity to tell the truth or to lie, respectively. Thus they always do their best to answer questions addressed to them.

A native named Eugene has been suspected of an attempt to break in an ATM near the police station. The case has been assigned to Inspector Negombo (a vampire) of local police force. His first task was to establish if the accused is a vampire or a zombie. Inspector Negombo was clever enough to determine that Eugene is a vampire on the basis of the suspected's answer to a single question. What was Negombo's question?

There are many possibilities. The question could be *e.g.* "Is it the case that you a vampire or you are not a vampire?": the positive answer identifies the answerer as a vampire, the negative one identifies the person as a zombie. This solution may cause a usual astonishment of the Watson-like audience as well as rise the admiration of Negombo's methods[6]. However, its explanation would certainly cause as usual "It is pretty obvious now" reaction.

Let us reveal Inspector Negombo's hidden agenda. He knows the following fact:

1. Every native is either a vampire or a zombie.

He knows also the following rules:

2. Every native who is a vampire utters true sentences.

3. Every native who utters a true sentence is a vampire.

(These two rules could be expressed more succintly as an equivalence, but it is in our hidden agenda to leave them in an implicational form.) The only problem is to operationalise the rules in such a way that they will be applicable to questions. This may be done as follows:

$2'$. For every native x, if x gives back a true answer to a posed question, then x is a vampire.

$3'$. For every native x, if x is a vampire, then x gives back a true answer to a posed question.

An important premise in Negombo's reasoning is the following:

4. Eugene is a native of the island.

What remains is to find a suitable question. It would be useless to ask questions like: "Are you a vampire?". Every native of the island would answer this questions positively giving no clue who is lying and who is telling the truth. The point is to ask a question with such direct answers that both Negombo and the suspected will know their truth values – as in the case of questions about fairly simple logically true (or false) sentences. Let us express Negombo's hidden agenda in terms of a formalized language (Wiśniewski, 1995).

Let $V(x)$, $Z(x)$, $N(x)$ stand for expressions: "x is a vampire", "x is a zombie", "x is a native of the island" respectively, and let $U(x, A_i, ?\{A_1, \ldots, A_n\})$ stand for an expression "x gives back an answer A_i to the question $?\{A_1, \ldots, A_n\}$"[7] (provided that $i = 1, \ldots, n$). Finally, let the constant a represent Eugene. Negombo's agenda is depicted by the erotetic search scenario of example 1 (for brevity we assume that R stands for the formula $V(a) \vee \neg V(a)$).

Negombo's initial question, "Is Eugene a vampire, a zombie, or no information is available?" is expressed by the first e-formula: $?\{V(a), Z(a), \mathsf{T}\}$. The inspector makes use of four declarative premises. The first one, $\forall x(N(x) \rightarrow V(x) \perp Zx)$, expresses the fact 1, that every native is either a vampire or a zombie. The second and the third premises express the rules $2'$ and $3'$, respectively. The fourth premise states that Eugene is a native of the island.

[5] An important though tacit assumption is that both vampires and zombies of this island reason in accordance with classical logic.

[6] Although if a reader would like to argue that on this particular island this should be something like a standard procedure of interrogation, we shall agree.

[7] In a formula $U(x, A_i, ?\{A_1, \ldots, A_n\})$ the second argument of the predicate symbol U is a name of a d-wff and the third argument is a name of an e-formula. For the sake of brevity we omit the quotation marks, as no ambiguity arises in this context.

Example 1

$$?\{V(a), Z(a), \mathsf{T}\}$$
$$\forall x(N(x) \to V(x) \perp Z(x))$$
$$\forall x(U(x, R, ?R) \to V(x))$$
$$\forall x(V(x) \to U(x, R, ?R))$$
$$N(a)$$
$$?V(a)$$
$$?U(a, R, ?R)$$

$$U(a, R, ?R) \quad \neg U(a, R, ?R)$$
$$V(a) \qquad\quad \neg V(a)$$
$$\qquad\quad\; Z(a)$$

The initial question together with the first and the fourth declarative premise implies the question $?V(a)$[8]. Observe that this question, though implied, is not asked: it is not a query of the scenario[9]. The question $?V(a)$ together with the rules implies in turn the next question in the scenario, expressed by the erotetic formula $?U(a, R, ?R)$. This question is a query: it is answered in the scenario. The positive answer to it, expressed by the formula $U(a, R, ?R)$, together with the rule $2'$ entails the positive answer to $?V(a)$ which is also an answer to the initial question. On the other hand, the negative answer, expressed by the formula $\neg U(a, R, ?R)$, together with the rule $3'$ entails the negative answer to $?V(a)$ which, in turn, together with the first and the fourth premise, entails $Z(a)$, an answer to the initial question[10].

What this scenario shows is that the only question that needs to be actually addressed to Eugene is the question $?R$. What is more, this question and Eugene's answer would form all the interrogation conducted by Negombo. However, $?R$ is only mentioned and not used in the scenario (it is neither the initial question nor an implied question). The scenario reveals also what conclusions will Negombo derive on the basis of his declarative premises, questions and possible Eugene's answers[11].

One important property of e-scenarios is that they are "conditionally safe": if an initial question Q of a given e-scenario Φ has a true direct answer and if all the declarative premises of Φ are true, then at least one path of Φ leads to a true answer to Q; what is more, all d-wffs of this path are true and all e-formulas of this path have true direct answers as well. This is the essence of the Golden Path Theorem (Wiśniewski, 2003, p. 410–411).

It would be quite simple now to solve the problem: just ask Eugene if he really did try to break in the ATM. But this simple plan has been ruined by discovery that one premise on which Negombo's inferences were dependent is false: Eugene was not a native of the island. He came there as an immigrant from the nearby island, inhabited exclusively by humans, who are totally unpredictable as for the truth or falsity of what they tell[12]. Moreover, Eugene refused to give any sort of testimony. A short investigation among Eugene's friends (all confirmed being vampires by Negombo's test) led Negombo to establish the following rules:

5. If Eugene did run short of money, then he attempted to break in an ATM or borrowed some money from Eustace.

6. If Eugene didn't run short of money, then he went shopping or visited his favourite pub.

7. If Eugene attempted to break in an ATM or went shopping, then he has been seen in a local mall.

8. If Eugene borrowed some money from Eustace or visited his favourite bar, then he hasn't been seen in a local mall.

On this basis Inspector Negombo devised the plan for further interrogation of Elyssa, the only of Eugene's friends able to describe the course of events of that particular day (cf. example 2; $At(x)$ stands for "x attempted to break in an ATM", $Sh(x)$ stands for "x ran short of money", $Br(x, y)$ stands for "x borrowed money from y", $Sp(x)$ stands for "x went shopping", $Pb(x)$ stands for "x visited his favourite pub", $Ml(x)$ stands for "x

[8] In fact the question $?V(a)$ is implied by the question $?\{V(a), Z(a), \mathsf{T}\}$ on the basis of the empty set as well, but this holds for trivial reasons. Questions with T as a direct answer should be dealt with with some caution if one is not to fall into such triviality. One way to provide such caution is to employ *constructive* erotetic implication, instead of the standard one (Wiśniewski, 2007).

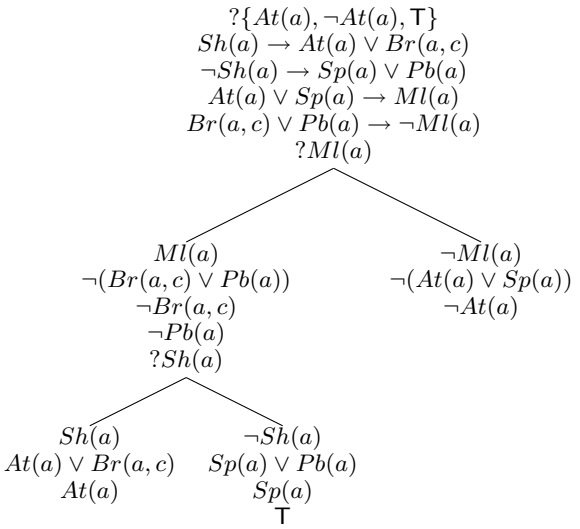
[9] However, this question is an important inferential step, because erotetic implication is not transitive.

[10] The relevant erotetic implications on which this and subsequent scenarios are based may be found in (Wiśniewski, 1995), (Wiśniewski, 2007), and (Łupkowski, 2010).

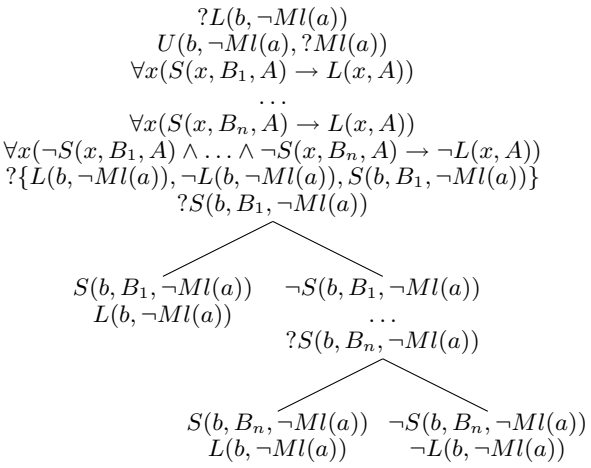
[11] Van Kuppevelt (1995) presents somewhat similar ideas of deriving questions (both explicit and implicit) from other questions and/or declaratives in the context of analysis of discourse structure. However, his informal account concerns wh-questions only.

[12] Although still *classically* unpredictable.

Example 2



Example 3



has been seen in a local mall" and c stands for Eustace).

"If he's guilty and if this is my lucky day, I'll send him to the court in two questions", Negombo said to himself. "If he's been in the mall but didn't run short of money then my information is insufficient and I will need new evidence. If he hasn't been in the mall, he's innocent. Well, we'll see". He ordered one of his lieutnants to conduct an interrogation according to this plan[13]. The lieutnant soon reported the outcome: Elyssa answered first query with "No". Eugene hasn't been in the mall. He's innocent!

However, it occured that Elyssa is a human, too. Unfortunately for Elyssa and Eugene, Negombo studied nonverbal behaviour of human liars (Ekman, 2001) and has been identified as a "Truth Wizard", that is, a person who can identify deception with exceptional accuracy of more than 80% (Harrington, 2009). Besides his natural talent he devised for himself a list of behaviours that help identifying lies with satisfactory precision. Negombo, highly suspicious as for the truth of what Elyssa testified, decided to try to kill two birds with one stone and possibly accuse her of false testimony. Negombo repeated Elyssa's interrogation, this time personally, having in mind an agenda represented by the e-scenario of example 3 ($L(x, A)$ stands for "x lies saying A", $S(x, B, A)$ stands for "x expresses the set of behaviours B while saying A" and b stands for Elyssa).

Again, this scenario shows that the only question that Negombo should actually pose to Elyssa is "Has Eugene been in the mall?" (the one represented by the e-formula $?Ml(a)$) although it is known what will the answer be. All the remaining questions play the role of milestones on Negombo's way of thinking in solving the initial problem. Notice that they are concerned not with the content of Elyssa statement but with the way she provided that statement.

Elyssa repeated her previous testimony that Eugene has not been in the mall. But saying this she expressed a set of behaviours characteristic for a liar (say that they were microexpressions of her lips indicating disbelief in what she has been saying[14]). On this basis Negombo determined that she is lying that Eugene has not been in the mall. To finish his investigation quickly he decided to employ ethically disputable means. He produced a fake witness (who testified that he has seen Eugene in the mall) and confronted Elyssa with him. Elyssa finally admitted that she was lying and that Eugene in fact has been in the mall on that particular day. Her statement has been recorded. The case has been sent to the court.

3 The Turing Test

The idea of erotetic reconstruction of an interrogator's hidden agenda comes from Łupkowski's (2010) analysis of a judge's strategies in the Turing Test (TT).

[13] The reader may notice that both queries of this scenario, that is $?Ml(a)$ and $?Sh(a)$, might demand plans for investigation in the form of e-scenarios on themselves. Such auxiliary e-scenarios can be incorporated into the main one by the embedding operation (Wiśniewski, 2003).

[14] An interested reader may have a look at the example of such microexpression on Paul Ekman's page: `http://www.paulekman.com`.

In the TT the judge (interrogator, **J**) poses questions and his/her aim is to establish on the basis of answers received if an interlocutor (answerer, **A**) is a human or not (in which case it is inferred that the interlocutor is a machine). It is reasonable to assume that the judge would not ask questions at random, but will use some kind of a strategy. This is due to the fact that it is the judge who is responsible for the way the test is performed – **J** asks questions and **A** answers them. It is **J** who also decides when the test will end. There are good reasons why it is beneficial for **J** to choose an e-scenario as a questioning plan in the TT:

- An e-scenario gives information when and what question should be asked (relative to the initial question and initial premises).
- What is more, it ensures that all subsequent questions asked would be relevant to the initial question (so we may say that no unnecessary information would be collected by **J**).
- E-scenarios also guarantee that each subsequent question asked is a step towards the answer to the initial question.
- The Golden Path Theorem guarantees that for a strategy expressed by an e-scenario there exist at least one such path that ends with the answer to the initial question which is true (relative to the initial premises).
- A strategy presented by an e-scenario is flexible in the sense that it can be modified and rearranged by the embedding procedure to fit the **J**'s needs for the current interrogation.

It would be useless for **J** to ask "Are you a human?", because everything he/she can get are just simply **A**'s doubtful declarations. A sound strategy for **J** is to devise a plan for an interrogation that is based on assumptions concerning criteria for being a human. The judge may formulate this criteria as sufficient and/or necessary conditions, representing his/her expectations as for the answers which would be given to his/her questions by a human.

One of the possible ways to put the judge's beliefs might be the following:

"If d is a human, then d fulfils a condition C_i."

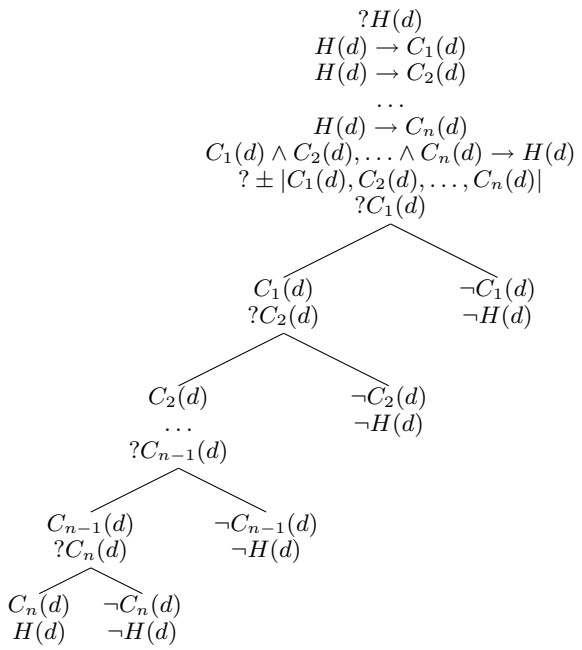
We may present the interrogator's beliefs as a set of formulas of the following form:

$H(d) \rightarrow C_1(d)$
$H(d) \rightarrow C_2(d)$
$\ldots$
$H(d) \rightarrow C_n(d)$
$C_1(d) \wedge C_2(d) \wedge \ldots \wedge C_n(d) \rightarrow H(d)$

where H (standing for "... is a human") is different from any of C_i $(1 \leq i \leq n)$.

The interrogator uses the necessary conditions of being a human in this case. Fulfilling all these rules together is – in the interrogator's opinion – a sufficient condition of being a human. The e-scenario which might be used as a strategy for the interrogator in this case is presented as example 4.[15]

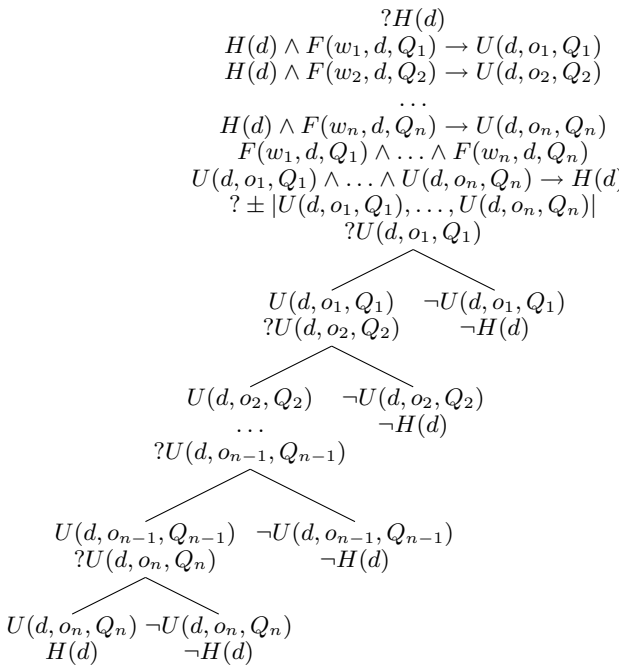
Example 4



The queries of this e-scenario should be treated as questions asked by **J** to himself/herself. Again, it seems fruitless to ask these questions directly to **A**. If **J** for example asks a question "Can you play chess?" and **A** answers "Yes", **J** do not obtain any interesting piece of information (at least if we consider the Turing Test's perspective). This is the reason why we will treat the queries of an e-scenario as only setting out the questions which should be asked to the answerer.

[15] In this scenario we make use of a *conjunctive question* represented by a formula $? \pm |A_1, \ldots, A_n|$ that should be read: "Is it the case that A_1 and ... and is it he case that A_n?". To grasp the idea, consider binary conjunctive question: "Is it the case that A_1 and is it the case that A_2?". This question, abbreviated by $? \pm |A_1, A_2|$, has four direct answers: $A_1 \wedge A_2, A_1 \wedge \neg A_2, \neg A_1 \wedge A_2, \neg A_1 \wedge \neg A_2$. For precise definition of a conjunctive question see (Urbański, 2001, p. 76).

Example 5



To clarify the intuition of setting out the questions to be asked by some queries of an e-scenario, we will present some operationalisation. To do this, we assume that the interrogator will accept premises of the following form (here formulated in a first person manner):

(*) if d is a human and I formulate the condition w_i (as a task's condition) and then I ask d the question Q_i, then d gives back an answer o_i to the question Q_i.

In this scheme o_i is an answer to question Q_i such that (in the interrogator's opinion) exactly that kind of an answer would be given by a human being, taking condition w_i set for the task into account. We will write the scheme in symbols as the following:

(**) $H(d) \wedge F(w_i, d, Q_i) \rightarrow U(d, o_i, Q_i)$,

where $F(w_i, d, Q_i)$ stands for "I formulate the condition w_i for the task and then I ask d a question Q_i", and U, H are understood as before.

Let us assume that the interrogator uses n such premises (where $n > 1$). Then, the strategy for the interrogator is expressed by the e-scenario of example 5.

Due to this kind of approach, we obtain an easy way of differentiating the questions which the interrogator asks to himself/herself from the ones asked to the answerer. The first group of questions are: $?U(a, o_1, Q_1), \ldots, ?U(a, o_n, Q_n)$ (they are used in the e-scenario as implied questions), while the second one are: $Q_1, \ldots, Q_n$ (they are only mentioned in the e-scenario as the third arguments of the predicate U).

When we take the Golden Path Theorem into account, we may say that the judge carrying out the presented e-scenario will end the interrogation with accurate identification of the answerer. Of course, we should still have in mind that this would be possible if all declarative premises of the e-scenario were true (which is a rather strong assumption).

At this stage, one may clearly see that the final result of the TT relies heavily on the knowledge and beliefs of the interrogator (this is one of the serious issues of the TT's setting). An e-scenario ensures only that the interrogator will get the answer to the initial question of the e-scenario. This e-scenario, however, does not guarantee the true answer, which is understood as an accurate identification of the answerer. The identification's accuracy depends on the set of premises on the basis of which the interrogator builds his/her e-scenario for the TT. This consequence might be seen as a weak point of the TT setting. However, when we take a closer look, it appears that the problem has its roots much deeper, in the unclear and fuzzy criteria of "being a human" (cf. for example considerations on how we assign thinking to other human beings presented by Moor (1976) and Stalker (1976)).

It is worth noticing, that for the real-life Turing Test, at least some elements of reasoning involving probabilities would be necessary. Some additional statistical rules might be used e.g. to set a proportion of satisfactory to unsatisfactory answers obtained by **J**. We may however imagine that a procedure of defining these rules might be something like the proposal made by French (1996), i.e. the so-called *Human Subcognitive Profile*. According to French, it is possible to establish (using empirical procedures) the profile of human answers to questions concerning low level cognitive structures.

4 Conclusions and future work

Erotetic search scenarios allow for the formal modeling of an interrogator's hidden agenda. What is more they offer a possibility of a differentiation of questions posed by an interrogator to himself/herself from the ones that actually should

be asked to an interlocutor. However, adequacy of resulting models is far from satisfactory. This is due to the fact that the semantics of underlying logic is rather simple and that assumed inferential relations (both declarative and erotetic ones) are relatively modest in character. Thus further development of this kind of formal models of interrogator's hidden agenda requires extension of the IEL framework to logics offering deeper insights into mechanisms underlying dialogue and argumentation.

References

Nuel D. Belnap. 1969. Åqvist corrections-accumulating question sequences. In: *Philosophical Logic*. J. W. Davis, P. J. Hockney, and K. Wilson (eds.), pp. 122–134. Reidel.

Paul Ekman. 2001. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W. W. Norton & Company, New York.

Robert French. 1996. The Inverted Turing Test: How a Mindless Program Could Pass It. *Psycholoquy*, 7(39).

Brooke Harrington (ed.). 2009. *Deception: From Ancient Empires to Internet Dating*. Stanford University Press.

Jan van Kuppevelt. 1995. Discourse structure, topicality and questioning. *Journal of Linguistics*, 31:109–147.

Paweł Łupkowski. 2010. *The Turing Test. Interrogator's Perspective* (*Test Turinga. Perspektywa sędziego*, in Polish). Adam Mickiewicz University Press, Poznań (in press).

James H. Moor. 1976. An Analysis of the Turing Test. *Philosophical Studies*, 30:249–257.

Raymond Smullyan. 1978. *What Is the Name of This Book*. Prentice-Hall, Englewood Cliffs, NJ.

Douglas F. Stalker. 1976. Why machines can't think: A reply to James Moor. *Philosophical Studies*, 34:317–320.

Mariusz Urbański. 2001. Synthetic tableaux and erotetic search scenarios: Extension and extraction. *Logique & Analyse*, 17–174–175:69–91.

Andrzej Wiśniewski. 1995. *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Kluwer Academic Publishers, Dordrecht–Boston–London.

Andrzej Wiśniewski. 2003. Erotetic Search Scenarios. *Synthese*, 134(3):389–427.

Andrzej Wiśniewski. 2007. Constructive Erotetic Implication. (manuscript).

Andrzej Wiśniewski. 2010. Erotetic Logics. Internal Question Processing and Problem-solving. UNILOG'03, Lisbon, `http://www.staff.amu.edu.pl/~p_lup/aw_pliki/Unilog`.

Andrzej Wiśniewski and Jerzy Pogonowski. 2010. Interrogatives, Recursion, and Incompleteness. *Journal of Logic and Computation*, 10:114, in press (doi:10.1093/logcom/exq014).

Cooperative answering and Inferential Erotetic Logic

Paweł Łupkowski

Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland
Pawel.Lupkowski@amu.edu.pl

Abstract

This paper addresses the issue of applicability of erotetic search scenarios, a tool developed within Inferential Erotetic Logic, in the area of cooperative answering for databases and information systems. Short descriptions of cooperative answering and Erotetic Search Scenarios are given. Some basic cooperative answering phenomena are modeled within the framework of Erotetic Search Scenarios.

1 Introduction

The issue of cooperative answering is important in the field of databases and information systems. Databases and information systems in general offer correct answers (as far as these systems contain valid data). The problem is to ensure that the answers will be also non-misleading and useful for a user.[1] To solve this problem certain specific techniques were developed. The most important are the following:

- consideration of specific information about a user's state of mind,
- evaluation of presuppositions of a query,
- detection and correction of misconceptions in a query (other than a false presupposition),
- formulation of intensional answers,
- generalization of queries and of responses.

A detailed description of the above techniques may be found in (Gaasterland et al., 1994) and (Godfrey, 1997). For their implementation in various database and information systems see e.g. (Godfrey et al., 1994), (Gal, 1988), (Benamara and Dizier, 2003b).

[1] H. P. Grice in (Grice, 1975) points out three features of what we may call a *cooperative answer*. It should be (i) correct, (ii) non-misleading, and (iii) useful answer to a query.

In this paper we will consider, among others, the following important aspect of cooperative answering: providing additional information useful for a user confronted with a failure. As Terry Gaasterland puts it:

> On asking a query, one assumes there are answers to the query too. If there are no answers, this deserves an explanation. (Gaasterland et al., 1994, p. 14)

We may consider a well known example here (cf. (Gal, 1988, p. 2)). Imagine that a student wants to evaluate a course before registering in it. He asks the secretary:

Q: *How many students failed course number CS400 last semester?*

The course CS400 was not given last semester, so the secretary would easily recognise the student's false assumption and correct it in her answer:

A_1: *Course number CS400 was not offered last semester.*

However, for most database interface systems the answer would be:

A_1^*: *None.*

Without additional explanations given, this response would be misleading for the student, and thus uncooperative from our perspective.

We will be addressing this issue, focussing our attention on the following cases: (a) the answer to a question is negative, (b) there is no answer available in a database, and (c) the asked question bares a misconception (i.e. it requests for information impossible to obtain from the database). We are going to make use of the Erotetic Search Scenarios framework, developed within A. Wiśniewski's Inferential Erotetic Logic

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

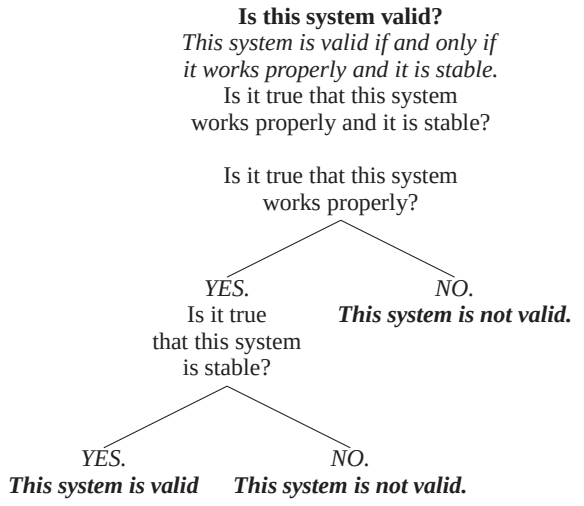
(cf. (Wiśniewski, 1995)). The reason for this choice is that Inferential Erotetic Logic provides the concept of validity of inferences which involve questions.[2]

For the reasons of space, only an informal characteristics of Erotetic Search Scenarios (thereafter referred to as *e-scenarios*) will be given here. The exact definition and many examples can be found in (Wiśniewski, 2001), (Wiśniewski, 2003) or (Wiśniewski, 2004).

E-scenarios are defined in terms of syntax and semantics. But:

> Viewed pragmatically, an e-scenario provides us with conditional instructions which tell us what questions should be asked and when they should be asked. Moreover, an e-scenario shows where to go if such-and-such a direct answer to a query appears to be acceptable and goes so with respect to any direct answer to each query. (Wiśniewski, 2003, p. 422)

For instance, let us imagine that we ask if a given system is valid and at the same time we construe the relevant concept of validity as follows: a system is valid just in case it works properly and is stable. How can one cope with this problem? A solution may be offered by an e-scenario. We can present this e-scenario as a downward tree with the main question as the root and direct answers to it as leaves. The relevant e-scenario for our exemplary problem is:



[2]For comparison of Inferential Erotetic Logic and J. Hintikka's Interrogative Model of Inquiry (with *interrogative game* as a central concept) see e.g. (Wiśniewski, 2004, p. 139–140).

The exemplary e-scenario, as well other e-scenarios, may be written in a formal language. Let us use here a language which we will call L_2; this language resembles a language characterised in (Wiśniewski, 2001, p. 20–21). The 'declarative' part of L_2 is a first-order language with identity and individual constants, but without function symbols. A *sentence* is a declarative well-formed formula (d-wff for short) with no occurrence of a free variable. The metalinguistic expression Ax refers to d-wffs of L_2 which have x as the only free variable. $A(x/c)$ designates the result of the substitution of an individual constant c for the variable x in Ax.

The vocabulary of the 'erotetic' part of L_2 consists of the signs: $?$, $\{$, $\}$, S, U, and the comma.

Questions of L_2 are expressions of the following forms:

(i) $?\{A_1, A_2, \ldots, A_n\}$
where $n > 1$ and $A_1, A_2, \ldots, A_n$ are syntactically distinct sentences of L_2,

(ii) $?\mathsf{S}(Ax)$
(iii) $?\mathsf{U}(Ax)$
where x is an individual variable and Ax is a d-wff of L_2 which has x as the only free variable.

Direct answers are defined as follows. For a question of the form (i), each of $A_1, A_2, \ldots, A_n$ is a direct answer to the question. For a question of the form (ii), a direct answer to it is a sentence of the form $A(x/c)$, where c is an individual constant. Direct answers to questions of the form (iii) fall under the schema:

$$A(x/c_1) \wedge \ldots \wedge A(x/c_n) \wedge \forall x(Ax \rightarrow x = c_1 \vee \ldots \vee x = c_n)$$

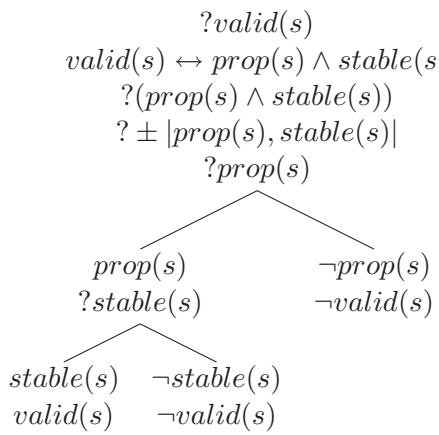
where $n \geq 1$ and $c_1, \ldots, c_n$ are distinct individual constants.

A question of the form (i) can be read, 'Is it the case that A_1, or is it the case that A_2, ..., or is it the case that A_n?'. A question of the form (ii) can be read, 'Which x is such that Ax?', whereas a question of the form (iii) can be read, 'What are all of the x's such that Ax?'.

For brevity, we will adopt a different notation for some types of questions. A question of the form $?\{A, \neg A\}$ ('Is it the case that A?') will be abbreviated as $?A$. The so-called (two-argument) *conjunctive questions*, namely $?\{A \wedge B, A \wedge \neg B, \neg A \wedge B, \neg A \wedge \neg B\}$ (to be read, 'Is

it the case that A and is it the case that B?'), will be abbreviated as $? \pm |A, B|$ — cf. (Wiśniewski, 2003, p. 399).

Here is the exemplary e-scenario written in the language L_2 ($valid$ stands for 'system is valid', $prop$ stands for 'system works properly', and $stable$ for 'system is stable'; the letter 's' is an individual constant, a name of the system):



As above, the e-scenario has a tree-like structure with the main question as the root and direct answers to it as leaves. Other questions are auxiliary. Either an auxiliary question has another question as the immediate successor (cf. e.g., '$?(prop(s) \wedge stable(s))$') or it has all the direct answers to it as the immediate successors (cf. e.g., '$?prop(s)$'). In the latter case the immediate successors represent the possible ways in which the relevant request for information can be satisfied, and the structure of the e-scenario shows what further information requests (if any) are to be satisfied in order to arrive at an answer to the main question. If an auxiliary question is a 'branching point' of an e-scenario, it is called a *query* of the e-scenario. Among auxiliary questions, only queries are to be asked; the remaining auxiliary questions serve as 'erotetic' premises only.

An e-scenario consists of *paths*, each of which leads from the main question through premises, auxiliary questions and answers to them, to a (direct) answer to the initial question. Viewed pragmatically, a path delivers some 'if/then' instructions. For instance, the instructions given by the leftmost path of our exemplary e-scenario are presented by Algorithm 1.

The key feature of e-scenarios is that auxiliary questions appear in them on the condition they are *erotetically implied* (in the sense of Inferential Erotetic Logic). Erotetic implication, Im, is a semantical relation between a question, Q, a (possibly empty) set of d-wffs, X, and a question, Q_1. Without going into details let us only say that Im ensures the following: (a) if Q has a true direct answer and X consists of truths, then Q_1 has a true direct answer as well ('transmission of soundness and truth into soundness'), and (b) each direct answer to Q_1, if true, and if all the X-es are true, narrows down the class at which a true direct answer to Q can be found ('open-minded cognitive usefulness'). For details see (Wiśniewski, 2003).

if *the main question is* $?valid(s)$ and *the initial premise is* $valid(s) \leftrightarrow prop(s) \wedge stable(s)$ **then**
  ask $?prop(s)$;

if *the answer is* $prop(s)$ **then**
  ask $?stable(s)$;

if *the answer is* $stable(s)$ **then**
  the answer to the main question is $valid(s)$;

Algorithm 1: Instructions given by the leftmost path of the exemplary e-scenario

Our exemplary e-scenario is based upon the following logical facts (A and B perform here the role of metalinguistic variables for sentences of L_2):

$\mathsf{Im}(?C, C \leftrightarrow A \wedge B, ?(A \wedge B))$
$\mathsf{Im}(?(A \wedge B), ? \pm |A, B|)$
$\mathsf{Im}(? \pm |A, B|, A)$
$\mathsf{Im}(? \pm |A, B|, B)$

2 Erotetic Search Scenarios and basic cooperative answering behaviours

In this section we will consider a very simple ('toy') example of a database. The database will be in a deductive database form. This exemplary database (thereafter we will refer to it as DB) will serve as a testing field for e-scenarios applicability in the domain of cooperative answering.

Each deductive database consists of:

- Extensional database ($\mathcal{EDB}$) — build out of facts,
- Intensional database ($\mathcal{IDB}$) — build out of rules,
- Integrity constraints ($\mathcal{IC}$).

For simplicity and notation coherence, we do not use the Datalog notation usually applied in similar contexts. Instead, we will be using the language L_2 described in the previous section.

In our case $\mathcal{EDB}$ consists of the following facts (where usr stands for 'is a user' and $live$ stands for 'lives in'):

$usr(a)$ $\quad$ $live(a,p)$
$usr(b)$ $\quad$ $live(b,zg)$
$usr(c)$ $\quad$ $live(c,p)$
$usr(d)$

The $\mathcal{IDB}$ contains the following rules (where loc_usr means 'is a local user'):

$$loc\_usr(x) \rightarrow usr(x)$$

$$loc\_usr(x) \rightarrow live(x,p)$$

$$usr(x) \wedge live(x,p) \rightarrow loc\_usr(x)$$

As for the $\mathcal{IC}$, there is only one, saying that it is not possible to live in zg and p at the same time.

$$\neg(\exists x(live(x,zg) \wedge live(x,p)))$$

Questions carried out against the DB may be polar questions asking if objects have some properties. For example, one can ask if object a_i is a 'user'; this is expressed by '$?usr(a_i)$'. A direct answer is either 'yes' or 'no' (expressed by $usr(a_i)$ and $\neg usr(a_i)$, respectively). But we may as well ask about an example of an object satisfying the condition 'is a user'. This question is expressed in L_2 by '$?\mathsf{S}(usr(x))$'. We can also ask of all the objects satisfying the condition 'is a user', by means of '$?\mathsf{U}(usr(x))$'. Then an answer is supposed to list all the objects in the database having the property of being a user (e.g. '$usr(a) \wedge usr(b)$').

E-scenarios are applied by a layer located between a user and the DB (let us call it a 'cooperative layer'). The layer proceeds a question asked by a user by carrying out the relevant auxiliary questions/queries against the DB in a way determined by an e-scenario. The received answers to queries are then transformed into an answer to the main question. The scheme of such a system is presented in Figure 1.

For the purposes of this paper we need to choose some e-scenarios that might be used to work with the exemplary DB. Most of them will fall under a certain general schema; in designing the schema we rely on the following logical facts:

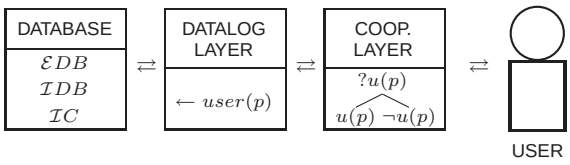
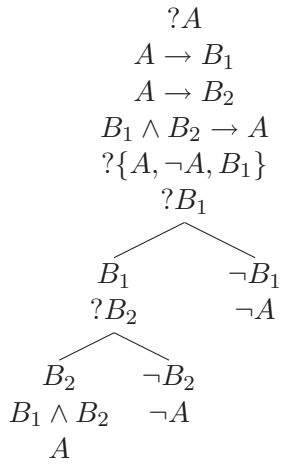


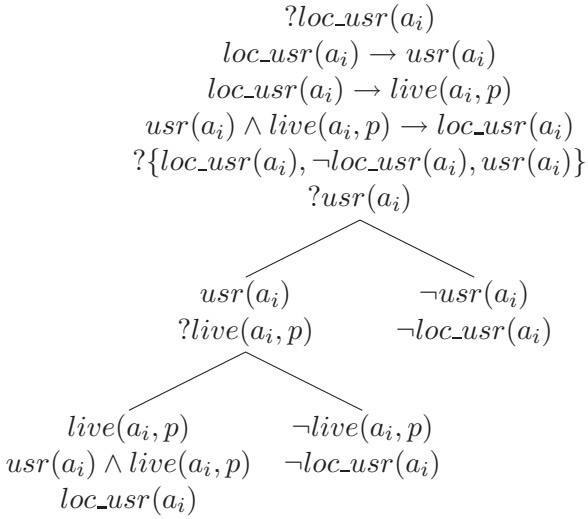
Figure 1: Scheme of the cooperative database system using e-scenarios

$\mathsf{Im}(?A, C \rightarrow A, ?\{A, \neg A, C\})$
$\mathsf{Im}(?\{A, \neg A, C\}, ?C)$
$\mathsf{Im}(?A, B_1 \wedge B_2 \rightarrow A, B_1, A \rightarrow B_2, ?B_2)$

Here is the schema:



We may adapt the above schema to our specific needs, obtaining e-scenarios for the concepts that occur in DB. For instance, the relevant e-scenarios for questions of the form 'Is a_i a local user?' fall under the schema (we will refer to it as ESS1).



Note that the $\mathcal{IDB}$ rules are used as premises of the e-scenario.

ESS1 should be regarded as a plan of questioning for the main question. It tells us which sub-

sequent questions should be executed against the DB and in which order/when it should be done.

Let let us see how ESS1 may be executed against the DB.

First, we consider an example of a question whose answer in view of the $\mathcal{IDB}$ is affirmative. This will help us to understand how ESS1 is executed against the DB. The question is, 'Is a is a local user?'. The question is executed against $\mathcal{IDB}$ as follows:

Q1: Is a a local user?

$$?loc\_usr(a)$$
$$loc\_usr(a) \rightarrow usr(a)$$
$$loc\_usr(a) \rightarrow live(a,p)$$
$$usr(a) \wedge live(a,p) \rightarrow loc\_usr(a)$$
$$?\{loc\_usr(a), \neg loc\_usr(a), usr(a)\}$$
$$?usr(a)$$
$$usr(a)$$
$$?live(a,p)$$
$$live(a,p)$$
$$usr(a) \wedge live(a,p)$$
$$loc\_usr(a)$$

answer: a is a local user ($loc\_usr(a)$),
since:
$loc\_usr(b) \rightarrow usr(b)$,
$loc\_usr \rightarrow live(b,p)$,
$usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$
and we know that:
a is a user **and** a lives in p

The diagram shows that only one path of ESS1 has been 'activated' or 'actualized' in order to get the answer (here the leftmost path). As long as a successful execution of an e-scenario is concerned, this is always the case. To indicate the unrealized questioning options we left the corresponding branches empty. For **Q1** the process of ESS1 execution against the DB boils down, in essence, to carrying out two subsequent queries, namely '$?usr(a)$', and (after receiving the affirmative answer '$usr(a)$'), the query '$?live(a,p)$'. The answers obtained entail the affirmative answer to the main question, which states that a is a local user, '$loc\_usr(a)$'. This answer is then provided (with additional explanations) to the user. Needles to say, the answer received can be regarded as cooperative.

Now we shall turn to other questions, chosen in such a manner that some more complex cooperative behaviors will be needed. First, let us consider the following question, 'Is b a local user?'.

Q2: Is b a local user?

$$?loc\_usr(b)$$
$$loc\_usr(b) \rightarrow usr(b)$$
$$loc\_usr(b) \rightarrow live(b,p)$$
$$usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$$
$$?\{loc\_usr(b), \neg loc\_usr(b), usr(b)\}$$
$$?usr(b)$$
$$usr(b)$$
$$?live(b,p)$$
$$\neg live(b,p)$$
$$\neg loc\_usr(b)$$

After ESS1 execution against the DB, one gets the negative answer to the main question. However, negative answers to polar questions are often less expected than affirmative ones; in some cases a negative answer can even be regarded as a failure. But we can easily supplement a negative answer received with an *explanation*. We do it by making use of the path just executed and the premises involved. Here is an example of an explanation:

answer: b is not a local user,
since: $loc\_usr(b) \rightarrow usr(b)$,
$loc\_usr \rightarrow live(b,p)$,
$usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$
and we know that: b does not live in p
$?live(b,p)$: $\neg live(b,p)$

The explanation contains information about the initial premises of the e-scenario (which reflect $\mathcal{IDB}$ part of the DB) and confront them with gained pieces of information. What is more it points out the query that failed, so a user knows exactly what information was not obtained from the DB.

As the example shows, e-scenarios allow to generate explanations of this kind in a quite easy way. The relevant procedure can be briefly described as follows.

We produce a list on the basis of the e-scenario's part just activated. We enumerate elements of the list consecutively; as a result we obtain an index, i.e. a sequence of indices.

1. $?loc\_usr(b)$
2. $loc\_usr(b) \rightarrow usr(b)$
3. $loc\_usr(b) \rightarrow live(b,p)$
4. $usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$

5. $?\{loc\_usr(b), \neg loc\_usr(b), usr(b)\}$
6. $?usr(b)$
7. $usr(b)$
8. $?live(b,p)$
9. $\neg live(b,p)$
10. $\neg loc\_usr(b)$

By means of the list we can identify the main question and the initial premises. The main question will be a formula of the form ?_ (i.e. formula beginning with the question mark ?') with index number 1. Then we identify a formula of the form ?_ that has the lowest index number greater than 1. Let the index number be k. All formulas with index numbers larger than 1 and lower than k are the initial premises; we write them down consecutively.

2. $loc\_usr(b) \rightarrow usr(b)$
3. $loc\_usr(b) \rightarrow live(b,p)$
4. $usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$

The task of finding the next remaining element of the explanation reduces to the issue of finding, on the list, a formula of the form ?_ with the largest index number. In this way the last 'active' query is identified. Then the query and the next element of the list (i.e. direct answer to this query) will be used in the explanation.

8. $?live(b,p)$
9. $\neg live(b,p)$

A formal description of the procedure is presented as Algorithm 2.

Let us now consider another example. By the way, the example shows how e-scenarios can decrease the number of queries executed agains the DB. The question is, 'Is e a local user?'

Q3: Is e a local user?

$?loc\_usr(e)$
$loc\_usr(e) \rightarrow usr(e)$
$loc\_usr(e) \rightarrow live(e,p)$
$usr(e) \wedge live(e,p) \rightarrow loc\_usr(e)$
$?\{loc\_usr(e), \neg loc\_usr(e), usr(e)\}$
$?usr(e)$

$\neg usr(e)$
$\neg loc\_usr(e)$

answer: e is not a local user ($\neg loc\_usr(e)$), **since:**

Data: E-scenario path as a list with index (we will denote element of the list as e and element of the index as i)
Result: Additional explanations for the answer to the initial question

1 $iq \leftarrow e$ with $i = 1$
2 $answ\_iq \leftarrow e$ with max i
3 $next\_q \leftarrow e$ of the form ?_ with min $i > 1$
4 i_{next_q} = i of $next\_q$
5 $premises \leftarrow e$ with $1 < i < i_{next\_q}$
6 $failing\_q \leftarrow e$ of the form ?_ with max i
7 $i_{failing\_q}$ = i of $failing\_q$
8 $answer\_failing\_q \leftarrow e$ with $i_{failing\_q} + 1$

Algorithm 2: Generation of additional explanations for the answer to the initial question

$loc\_usr(b) \rightarrow usr(b)$,
$loc\_usr \rightarrow live(b,p)$,
$usr(b) \wedge live(b,p) \rightarrow loc\_usr(b)$
and we know that: e is not a user
$?usr(e)$: $\neg usr(e)$

The next example illustrates how one can cope with a misconception of a question asked by a user. The question is: Does b live in zg and is a local user? We apply an e-scenario of a slightly different form than ESS1[3], i.e.

$?(live(a_i, zg) \wedge loc\_usr(a_i)$
$? \pm |live(a_i, zg), loc\_usr(a_i)|$
$?loc\_usr(a_i)$
$loc\_usr(a_i) \rightarrow usr(a_i)$
$loc\_usr(a_i) \rightarrow live(a_i, p)$
$usr(a_i) \wedge live(a_i, p) \rightarrow loc\_usr(a_i)$
$?\{loc\_usr(a_i), \neg loc\_usr(a_i), usr(a_i)\}$
$?usr(a_i)$

$usr(a_i)$ | $?live(a_i, p)$
$\neg usr(a_i)$ | $\neg loc\_usr(a_i)$ | $\neg(live(a_i, zg) \wedge loc\_usr(a_i))$

$live(a_i, p)$ | $usr(a_i) \wedge live(a_i, p)$ | $loc\_usr(a_i)$ | $?live(a_i, zg)$
$\neg live(a_i, p)$ | $\neg loc\_usr(a_i)$ | $\neg(live(a_i, zg) \wedge loc\_usr(a_i))$

$live(a_i, zg)$ | $\neg live(a_i, zg)$

[3] An additional logical fact is used here, namely: $\mathsf{Im}(?(A \wedge D), A, ?D)$.

Question **Q4** is executed against to DB as follows:

Q4: Does b live in zg and is a local user?

$$?(live(b, zg) \wedge loc\_usr(b)$$
$$? \pm |live(b, zg), loc\_usr(b)|$$
$$?loc\_usr(b)$$
$$loc\_usr(b) \rightarrow usr(b)$$
$$loc\_usr(b) \rightarrow live(b, p)$$
$$usr(b) \wedge live(b, p) \rightarrow loc\_usr(b)$$
$$?\{loc\_usr(b), \neg loc\_usr(b), usr(b)\}$$
$$?usr(b)$$

$$usr(b)$$
$$?live(b, p)$$

$$\neg live(b, p)$$
$$\neg loc\_usr(b)$$
$$\neg(live(b, zg) \wedge loc\_usr(b))$$

A simple negative answer to the initial question will not make a user aware of the misconception in the question. But when explanations are added, a user can learn about the database schema and understand better the obtained answer.[4]

answer: no ($\neg(live(b, zg) \wedge loc\_usr(b))$)
since: b is not a local user ($\neg loc\_usr(b)$)
this is due to:
$loc\_usr(b) \rightarrow usr(b)$,
$loc\_usr \rightarrow live(b, p)$,
$usr(b) \wedge live(b, p) \rightarrow loc\_usr(b)$
and: $\neg live(b, p)$
(*correction of user's misconception about the* DB *schema*)
but we know that: $usr(b)$

Yet another example shows one of the possible ways of dealing with information gaps in the database. We ask if d is a local user. During ESS1 execution against the DB a query fails since the requested information is not present in the database (this is indicated by the ⊠ symbol on the schema below). From this point, a further execution of ESS1 is no longer possible. However the executed part of the e-scenario enables us to generate a sensible explanation of this fact and to report the gained information relevant to the main question. Let us stress that the information gained in the process of ESS1 execution will always stay in connection with the main question. Consequently, we may simply report the obtained answers to queries to a user as a piece of information relevant to his/her question.

Q5: Is d a local user?

$$?loc\_usr(d)$$
$$loc\_usr(d) \rightarrow usr(d)$$
$$loc\_usr(d) \rightarrow live(d, p)$$
$$usr(d) \wedge live(d, p) \rightarrow loc\_usr(d)$$
$$?\{loc\_usr(d), \neg loc\_usr(d), usr(d)\}$$
$$?usr(d)$$

$$usr(d)$$
$$?live(d, p)$$

⊠ ⊠

answer: the answer is unknown, since:
$loc\_usr(b) \rightarrow usr(b)$,
$loc\_usr \rightarrow live(b, p)$,
$usr(b) \wedge live(b, p) \rightarrow loc\_usr(b)$
and: query $?live(d, p)$ failed
but we know that: d is a user

3 Summary and further works

I have presented here only some simple examples of cooperative answering behaviours that can modelled by means of the e-scenarios framework. But, in my opinion, Inferential Erotetic Logic provides many useful tools for investigating the area of cooperative answering. Erotetic Search Scenarios framework presented in this paper allows to join two focus points of cooperative answering research: question analysis and fundamental reasoning procedures — cf. (Benamara and Dizier, 2003b, p. 63). It also allows to develop techniques which are domain unspecific (in contrast to limited domains systems like the WEBCOOP developed by Benamara and Dizier (2003a), (2003b)).

Future works will concentrate mainly on incorporating techniques developed in (Gal, 1988) (based on integrity constraints processing) into presented framework. Also new algorithmic procedures working on e-scenarios that enable implementations of specific cooperative techniques will be developed. The area of automatic generation of premises for e-scenarios using Formal Concept Analysis will also be explored.

[4] At this stage of this research integrity constraints are not employed in the process of generating cooperative answers using e-scenarios. However concerning techniques and results presented in (Gal, 1988) it would be necessary to incorporate $\mathcal{IC}$ into e-scenarios' premises to obtain better cooperative behaviours (especially when users' misconceptions are considered).

References

Farah Benamara and Patrick Saint Dizier. 2003a. Dynamic Generation of Cooperative Natural Language Responses. Ninth European workshop on Natural Language Generation, Budapest, Hungary.

Farah Benamara and Patrick Saint Dizier. 2003b. WEBCOOP: a cooperative question-answering system on the web. *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, 63–66.

Terry Gaasterland, Parke Godfrey, and Jack Minker. 1994. An Overview of Cooperative Answering. In R. Demolombe and T. Imielinski, editors, *Nonstandard Queries and Nonstandard Answers*, 1–40.

Annie Gal. 1988. Cooperative Responses in Deductive Databases. PhD Thesis, University of Maryland, Department of Computer Science.

Parke Godfrey, Jack Minker, and Lev Novik. 1994. An Architecture for a Cooperative Database System. In W. Litwin and T. Risch, editors, *Proceedings of the First International Conference on Applications of Databases.* Lecture Notes in Computer Sicence, 819: 3–24.

Parke Godfrey. 1997. Minimization in Cooperative Response to Failing Database Queries. *International Journal of Cooperative Information Systems*, 6(2): 95–149.

Herbert P. Grice. 1975. Logic and Conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, 41–58, New York: Academic Press.

Mariusz Urbański. 2001. Synthetic Tableaux and Erotetic Search Scenarios: Extension and Extraction. *Logique & Analyse*, No. 173–174–175: 69–91.

Andrzej Wiśniewski. 1995. *The Posing of Questions: Logical Foundations of Erotetic Inferences,* Kluwer AP, Dordrecht, Boston, London.

Andrzej Wiśniewski. 2001. Questions and Inferences. *Logique & Analyse*, No. 173–175: 5–43.

Andrzej Wiśniewski. 2003. Erotetic Search Scenarios. *Synthese*, No. 134: 389–427.

Andrzej Wiśniewski. 2004. Erotetic Search Scenarios, Problem-Solving, and Deduction. *Logique & Analyse*, No. 185–188: 139–166.

Accommodating innovative meaning in dialogue

Staffan Larsson
University of Gothenburg
Sweden
sl@ling.gu.se

Abstract

Meaning accommodation occurs when a dialogue participant learns new meanings by observing utterances in context and adapting to the way linguistic expressions are used. This paper sketches a formal account of meaning accommodation, focusing on the detection of subjectively innovative uses of an expression. It also provides an example of adaptation of lexical meaning to accommodate innovation.

1 Introduction

Two agents do not need to share exactly the same linguistic resources (grammar, lexicon etc.) in order to be able to communicate, and an agent's linguistic resources can change during the course of a dialogue when she is confronted with a (for her) innovative use. For example, research on alignment shows that agents negotiate domain-specific microlanguages for the purposes of discussing the particular domain at hand (Clark and Wilkes-Gibbs, 1986; Pickering and Garrod, 2004; Brennan and Clark, 1996; Healey, 1997; Larsson, 2007). We use the term *semantic coordination* to refer to the process of interactively coordinating the meanings of linguistic expressions.

Several mechanisms are available for semantic coordination in dialogue. These include corrective feedback, where one DP implicitly corrects the way an expression is used by another DP, as well as explicit definitions and negotiations of meanings. However, it also possible to coordinate silently, by DPs observing the language use of others and adapting to it.

This paper aims towards a formal account of learning meanings from observation and accommodation in dialogue. The examples we present are from first language acquisition, where the child detects innovative (for her) uses and adapts her take on the meaning accordingly[1].

In the following, we will first introduce the basics of the TTR framework (section 2). We then review earlier work on learning meaning from corrective feedback, and discuss how various aspects of meaning can be represented using TTR (section 5). The following two sections show how we use TTR to represent multiple aspects of lexical meaning, including "perceptual type". Section 6 introduces the notion of meaning accommodation and briefly describes an experiment demonstrating that children can learn lexical meaning using meaning accommodation. Sections 7 and 8 explain how contexts are represented using TTR and provide an example of "normal" contextual interpretation. In section 9 we define a notion of innovation that we use to detect innovation in an example interaction where normal contextual interpretation breaks down, and specify the meaning update resulting from this interaction.

2 TTR

The received view in formal semantics (Kaplan, 1979) assumes that there are abstract and context-independent "literal" meanings (utterance-type meaning; Kaplan's "character") which can be regarded formally as functions from context to content; on each occasion of use, the context determines a specific content (utterance-token meaning). Abstract meanings are assumed to be static and are not affected by language use in specific contexts. Traditional formal semantics is thus ill-equipped to deal with semantic coordination, because of its static view of meaning.

[1] We regard semantic coordination in first language acquisition as a special case of semantic coordination in general, wherethere is a clear asymmetry between the agents involved with respect to expertise in the language being acquired when a child and an adult interact. However, we believe that the mechanisms for semantic coordination used in these situations are similar to those which are used when competent adult language users coordinate their language.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

We shall make use of type theory with records (TTR) as characterized in Cooper (2005) and elsewhere. The advantage of TTR is that it integrates logical techniques such as binding and the lambda-calculus into feature-structure like objects called record types. Thus we get more structure than in a traditional formal semantics and more logic than is available in traditional unification-based systems. The feature structure like properties are important for developing similarity metrics on meanings and for the straightforward definition of meaning modifications involving refinement and generalization. The logical aspects are important for relating our semantics to the model and proof theoretic tradition associated with compositional semantics.

We will now briefly introduce the TTR formalism. If $a_1 : T_1, a_2 : T_2(a_1), \ldots, a_n : T_n(a_1, a_2, \ldots, a_{n-1})$, the record to the left is of the record type to the right:

$$\begin{bmatrix} l_1 & = & a_1 \\ l_2 & = & a_2 \\ \ldots & & \\ l_n & = & a_n \\ \ldots & & \end{bmatrix}$$

is of type:

$$\begin{bmatrix} l_1 & : & T_1 \\ l_2 & : & T_2(l_1) \\ \ldots & & \\ l_n & : & T_n(l_1, l_2, \ldots, l_{n-1}) \end{bmatrix}$$

Types constructed with predicates may also be *dependent*. This is represented by the fact that arguments to the predicate may be represented by labels used on the left of the ':' elsewhere in the record type.

Some of our types will contain *manifest fields* (Coquand et al., 2004) like the ref-field in the following type:

$$\begin{bmatrix} \text{ref=obj123} & : & \text{Ind} \\ \text{c} & : & \text{glove(ref)} \end{bmatrix}$$

$\left[\text{ref=obj123:Ind}\right]$ is a convenient notation for $\left[\text{ref} : \text{Ind}_{\text{obj123}}\right]$ where $\text{Ind}_{\text{obj123}}$ is a *singleton type*. If $a : T$, then T_a is a singleton type and $b : T_a$ (i.e. b is of type T_a) iff $b = a$. Manifest fields allow us to progressively specify what values are required for the fields in a type.

An important notion in this kind of type theory is that of *subtype*. Formally, $T_1 \lhd T_2$ means that T_1 is a subtype of T_2. Two examples will suffice as explanation of this notion:

$$\begin{bmatrix} \text{ref} & : & \text{Ind} \\ \text{c} & : & \text{glove(ref)} \end{bmatrix} \lhd \begin{bmatrix} \text{ref} & : & \text{Ind} \end{bmatrix}$$

$$\begin{bmatrix} \text{ref=obj123} & : & \text{Ind} \end{bmatrix} \lhd \begin{bmatrix} \text{ref} & : & \text{Ind} \end{bmatrix}$$

Below, we will also have use for an operator for combining record types. The $\dot{\wedge}$ operator works as follows. Suppose that we have two record types C_1 and C_2:

$$C_1 = \begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{clothing}} : \text{clothing(x)} \end{bmatrix}$$

$$C_2 = \begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{physobj}} : \text{physobj(x)} \end{bmatrix}$$

$C_1 \wedge C_2$ is a type. In general if T_1 and T_2 are types then $T_1 \wedge T_2$ is a type and $a : T_1 \wedge T_2$ iff $a : T_1$ and $a : T_2$. A meet type $T_1 \wedge T_2$ of two record types can be simplified to a new record type by a process similar to unification in feature-based systems. We will represent the simplified type by putting a dot under the symbol $\wedge$. Thus if T_1 and T_2 are record types then there will be a type $T_1 \dot{\wedge} T_2$ equivalent to $T_1 \wedge T_2$ (in the sense that a will be of the first type if and only if it is of the second type).

$$C_1 \dot{\wedge} C_2 = \begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{physobj}} : \text{physobj(x)} \\ \text{c}_{\text{clothing}} : \text{clothing(x)} \end{bmatrix}$$

The operation $\dot{\wedge}$ corresponds to unification in feature-based systems and its definition (which we omit here) is similar to the graph unification algorithm.

3 Learning compositional and ontological meaning from corrective feedback and explicit definition

This section gives a brief overview of the work presented in (Cooper and Larsson, 2009) , which will be useful as a backdrop to the description of meaning accommodation. It will also introduce some important concepts which we will be making use of later.

Recent research on first language acquisition (Clark, 2007; Saxton, 2000) argues that the learning process crucially relies on negative input, including corrective feedback. We see corrective feedback as part of the process of negotiation of a language between two agents. Here is one of the

examples of corrective feedback that we discuss in connection with our argument for this position:

"Gloves" example (Clark, 2007):

- Naomi: mittens
- Father: **gloves**.
- Naomi: gloves.
- Father: when they have fingers in them they are called gloves and when the fingers are all put together they are called mittens.

We think that an important component in corrective feedback of this kind is syntactic alignment, that is, alignment of the correcting utterance with the utterance which is being corrected. This is a rather different sense of alignment than that associated with the negotiation of a common language, although the two senses are closely linked. By "syntactic alignment" here we mean something related to the kind of alignment that is used in parallel corpora. It provides a way of computing parallelism between the two utterances. Syntactic alignment may not be available in all cases but when it is, it seems to provide an efficient way of identifying what the target of the correction is.

Following Montague (1974) and Blackburn and Bos (2005) compositional semantics can be predicted from syntactic information such as category. For example, for common nouns we may use the formula

$$\text{commonNounSemantics}(N) = \lambda x N'(x)$$

or, using TTR,

$$\text{commonNounSemantics}(N) = \lambda r{:}\left[\text{x} : \mathit{Ind}\right] \left(\left[\text{c} : N'(r.\text{x})\right]\right)$$

There is an obvious relationship between this function and the record type

$$\begin{bmatrix} \text{x} : \mathit{Ind} \\ \text{c} : N'(\text{x}) \end{bmatrix}$$

It is clear that the compositional function can be derived from the record type (but we will not provide a general definition of this derivation here due to space limitations). We will use the record type representation in the remainder of this paper.

In the Gloves example, after the father's utterance of "gloves", Naomi could use syntactic alignment to understand this term as a noun with the corresponding kind of compositional semantics:

$$\text{GloveCompSem} = \begin{bmatrix} \text{x} : \mathit{Ind} \\ \text{c}_{\text{glove}} : \mathit{glove'}(\text{x}) \end{bmatrix}$$

In this way, compositional semantics can be derived from corrective feedback in dialogue. However, compositional semantics of this kind does not reveal very much, if anything, about the details of word semantics unless we add ontological information.

The ontological semantics of a lexical expression describes how the concept associated with the expression relates to other relevant concepts in an ontology. Provided that Naomi learns from the interaction that gloves are also a kind of clothing, Naomi's glove class after the first utterance by the father is the following type

$$\begin{bmatrix} \text{x} & : & \mathit{Ind} \\ \text{c}_{\text{glove}} & : & \text{glove}'(\text{x}) \\ \text{c}_{\text{physobj}} & : & \text{physobj}(\text{x}) \\ \text{c}_{\text{clothing}} & : & \text{clothing}(\text{x}) \end{bmatrix}$$

The father's second utterance contains a partial but explicit definition of the ontology of gloves and mittens:

- Father: when they have fingers in them they are called gloves and when the fingers are all put together they are called mittens.

When integrating this utterance, Naomi may modify her take on the ontological semantics so that after this update the meanings for "glove'" will be:

$$\begin{bmatrix} \text{x} : \mathit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{x}) \\ \text{c}_{\text{physobj}} : \text{physobj}(\text{x}) \\ \text{c}_{\text{clothing}} : \text{clothing}(\text{x}) \\ \text{c}_{\text{handclothing}} : \text{handclothing}(\text{x}) \\ \text{c}_{\text{withoutfingers}} : \text{withoutfingers}(\text{x}) \end{bmatrix}$$

4 Perceptual type

In this paper, we will add a further aspect of meaning, namely *perceptual type* (or perceptual meaning).

For our current purposes, we will represent perceptual meaning as a record type specifying and

individual and one or more propositions indicating that the individual is of a certain perceptual type, i.e., that it has certain physically observable characteristics.

The word "glove′", for example, may be associated with a certain shape:

GlovePercType =
$$\begin{bmatrix} \text{x} : \text{Ind} \\ \text{c}_{\text{glove-shape}} : \text{glove-shape(x)} \end{bmatrix}$$

Propositions, i.e., types which are constructed with predicates, are sometimes referred to as "types of proof". The idea is that something of this type would be a proof that a given individual has a certain property. One can have different ideas of what kind of objects count as proofs. For subsymbolic aspects of meaning, we are assuming that the proof-objects are readings from sensors. Thus, we assume that GlovePercType above will be inhabited if the agent detects a glove-shaped object in the context.

5 Aspects of meaning in TTR

Since all aspects of meaning can be modified as a result of language use in dialogue, we want our account of semantic innovation and semantic updates to include several aspects of lexical meaning. In this section, we will show how multiple aspects of lexical meaning can be represented using TTR.

Each lexical meaning representation in the lexicon can be structured according to the above aspects of meaning.

$$[\text{glove}] = \begin{bmatrix} \text{comp} & : & \text{GloveCompSem} \\ \text{class} & : & \text{GloveClass} \\ \text{perc} & : & \text{GlovePercType} \end{bmatrix}$$

A simplified representation can be obtained by collapsing the distinctions between the different aspects of meaning.[2].

$\bigwedge[\text{glove}] = [\text{glove}].\text{comp} \wedge [\text{glove}].\text{class} \wedge [\text{glove}].\text{perc} =$
$$\begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{x}) \\ \text{c}_{\text{physobj}} : \text{physobj(x)} \\ \text{c}_{\text{clothing}} : \text{clothing(x)} \\ \text{c}_{\text{glove-shape}} : \text{glove-shape(x)} \end{bmatrix}$$

[2]We define the operator $\bigwedge$ as follows:

If $T = \begin{bmatrix} \ell_1 & : & T_1 \\ \vdots & & \\ \ell_n & : & T_n \end{bmatrix}$, then $\bigwedge T = T_1 \wedge \ldots \wedge T_n$.

6 Meaning accommodation

To show how meaning can be learnt by observations of language use, (Carey and Bartlett, 1978) set up an experiment based on nonsense words to mimic the circumstances in which children naturally encounter new words. The subjects were 3- and 4-year olds. To enable testing for learning effects, they used the nonsense word: "chromium" to refer to the colour olive.They use it in the normal (nursery school) classroom activity of preparing snack time. In the experiment, one cup painted olive, and another was painted red. The adult test leader says to the child "Bring me the chromium cup; not the red one, the chromium one.". All children picked the right cup; however, this could be done by focusing on the contrast "not the red one" without attending to the word "chromium".

The results indicated that a very low number of exposures (five) had influenced the child's naming of olive and had effected a lexical entry for "chromium" which in many cases included that it was a colour term, and in some cases knowledge of its referent. Some learning seems to occur after a single exposure, at least sometimes. Based on this, it was concluded that acquisition proceeds in two phases. Firstly, *fast mapping* resulting from one or a few exposures to the new word; this includes only a fraction of the total information constituting full learning of the word, and typically includes hyponym relations. Secondly, *extended mapping* over several months, by which children arrive at full acquisition, including the ability to identify and name new instances. The experiment demonstrates learning from exposure without corrective feedback, and shows that both ontological information and perceptual type can be learnt in this way.

In accommodating innovative meanings, the innovation can concern either a new word for the DP, as in the experiment with "chromium", or a known word which is used in a new way. In the latter case, the needs to identify a way of using an expression that deviates from the way that expression is usually used (in the DP's subjective experience).

Below, we will propose a formalisation of the notion of innovative use of a known linguistic expression, in terms of a relation between the context of utterance and the meaning of an expression. We will also provide an illustrative example which includes the semantic updates resulting from the accommodation.

7 Representing contexts in TTR

To represent individual dialogue participants' takes on contexts[3], we will use record types with manifest fields. This allows our context to be underspecified, reflecting the fact that an agent may not have a complete representation of the environment. For example, an agent may entertain some propositions without necessarily having a proof for them, or know of the existence of certain objects without having identified them.

In first language acquisition, learning of perceptual type typically takes place in contexts where the referent is in the shared focus of attention and thus perceivable to the dialogue participants, and for the time being we limit our analysis to such cases. For our current purposes, we assume that our DPs are able to establish a shared focus of attention, and we will designate the label "focobj" for the object or objects taken by a DP to be in shared focus.

We assume that a DP's take on the context may be amended during interpretation by checking the environment (Knoeferle and Crocker, 2006). That is, the take on the context against which an utterance is interpreted and evaluated may not be identical to the DP's take on the context prior to the interpretation process. We will not be going into details of incremental interpretation here, however.

8 Contextual interpretation for non-innovative uses

As a backdrop for our discussion of how to detect innovation, we will first show how "normal" contextual interpretation, in the absence of innovations, is assumed to work.

For our current purposes, we will not go into details of compositional analysis, but instead provide the resulting semantic representation of the meaning of a whole utterance. We will follow Kaplan in assuming this result to be a function from context to content. Parts of the meaning is *foregrounded*, and which parts are *backgrounded*. Background meaning (BG) represents constraints on the context, whereas foreground material (FG) is the information to be added to the context by the utterance in question. We can represent this either as a record or as a function:

$$\begin{bmatrix} \text{BG} = \ldots \\ \text{FG} = \ldots \end{bmatrix}$$

$\lambda t{:}\text{BG(FG)}$

The functional version takes as argument a record type t which is a subtype of the background meaning of the uttered expression (typically a context containing manifest fields representing objects in the environment and propositions about these objects), and returns a record type corresponding to the foreground meaning. Contextual interpretation amounts to applying this function to the context. After contextual interpretation, we assume that the content of the utterance in question is to be added to the information state (Traum and Larsson, 2003) of the interpreting agent, which we here will take to be identical the agent's take on the context[4].

To illustrate contextual interpretation, we will use a modified version of the "gloves" example, where Naomi simply observes an utterance by Father:

Modified "Gloves" example:

- (Naomi is putting on her new gloves)
- Father: Those are nice gloves!

In this section, we will assume that Naomi is familiar with the term "gloves" and assigns it the standard (in the community of adult speakers of English) lexical meaning.

$$\bigwedge_{\cdot}[\text{glove}]^{Naomi} = \begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{x}) \\ \text{c}_{\text{physobj}} : \text{physobj(x)} \\ \text{c}_{\text{clothing}} : \text{clothing(x)} \\ \text{c}_{\text{glove-shape}} : \text{glove-shape(x)} \end{bmatrix}$$

We assume that Naomi's take on the meaning of the sentence uttered by Father is the following (using record type notation):

$[\text{Those are nice gloves}]^{Naomi} =$

[3] Occasionally and somewhat sloppily referred to as "contexts" below.

[4] In a complete model including both information sharing, dialogue management and semantic coordination the information state would probably need to be more complex than this.

$$\left[\begin{array}{ll} & \left[\begin{array}{l}\text{focobj} : \textit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{focobj}) \\ \text{c}_{\text{physobj}} : \text{physobj}(\text{focobj}) \\ \text{c}_{\text{clothing}} : \text{clothing}(\text{focobj}) \\ \text{c}_{\text{glove-shape}} : \text{glove-shape}(\text{focobj})\end{array}\right] \\ \text{FG} = & \left[\text{c}_{\text{nice}} : \text{nice}'(\text{BG.focobj})\right]\end{array}\right]$$

Of course, we are skipping over a lot here, including a compositional analysis and TTR analyses of "those", "are" and "nice". We also ignore the complication that "gloves" refers to a pair of gloves. Furthermore, we assume that the deictic "those" picks out the object in shared focus of attention (focobj).

Now for the context. We assume that when interpreting Father's utterance, Naomi perceives her gloves to be in shared focus. Her take on the situation includes the fact that they are gloves, that they have a certain appearance, and a certain ontological status.

$$c^{Naomi} = \left[\begin{array}{lcl}\text{focobj=a} & : & \text{Ind} \\ \text{c}_{\text{glove}} & : & \text{glove}'(\text{focobj}) \\ \text{c}_{\text{physobj}} & : & \text{physobj}(\text{focobj}) \\ \text{c}_{\text{clothing}} & : & \text{clothing}(\text{focobj}) \\ \text{c}_{\text{glove-shape}} & : & \text{glove-shape}(\text{focobj})\end{array}\right]$$

This allows us to do contextual interpretation by applying (the function version of) $\dot{\bigwedge}$[Those are nice gloves]^{Naomi} to c^{Naomi}:

$$[\text{Those are nice gloves}]^{Naomi}_{c} = [\text{Those are nice gloves}]^{Naomi}(c^{Naomi}) = \left[\text{c}_{\text{nice}} : \text{nice}'(\text{focobj})\right]$$

This record type can then be used to extend the information state (context). Note that nice$'$ will now be predicated of the individual a labeled by focobj in the context.

9 Formalising innovation

This section provides a TTR analysis of the first part of meaning accommodation, i.e. detection of an innovative use. Detecting innovation when a word is completely novel is relatively trivial; one only needs to recognise that it is not a word one has heard before[5]. We will instead focus on the case where a known expression is used with a (subjectively) innovative meaning.

[5]This is not to say that *assigning a meaning* to the new word is in any sense trivial.

We want a formal notion of innovativeness which we can use to detect innovative uses of linguistic expressions. The underlying intuition is that the meaning of an expression should say something about the kind of context in which it can be (non-innovatively) used. But how, exactly?

Intuitively, contextual interpretation can go wrong in at least the following ways:

1. Some information presupposed by the expression contradicts some information in the context; we refer to this as *background inconsistency*.

2. Some content conveyed by the utterance of the expression contradicts something in the context; we refer to this as *foreground inconsistency*.

Formally, these intuitions can be captured as follows. An expression e is *innovative* in context c if there is a mismatch between e and c in either of the following ways[6]:

1. c is inconsistent with the backgrounded meaning BG of e, formally $[e].\text{BG}\dot{\bigwedge}c \approx \bot$ (background inconsistency)

2. the content $[e]_c$ of e in c is inconsistent with c, formally $c\dot{\bigwedge}[e]_c \approx \bot$ (foreground inconsistency)

This definition follows naturally from how contextual interpretation works. Recall that meaning can be seen as a function from context to content, where background meaning serves as a constraint in the context. The definition of innovation checks that it will be possible to apply the meaning-function to the context, by checking that the context is consistent with the constraints imposed by the backgrounded meaning, and that the resulting contextual interpretation will be consistent with the context.

In cases where an innovative use is detected, "normal" contextual interpretation breaks down and needs to be fixed, either by objecting to or questioning this usage[7], or by adapting one's

[6]Due to space limitations we do not formally define inconsistency here. For current purposes, it suffices to say that a record type is inconsistent (equivalent to $\bot$) if it contains a proposition and its negation.

[7]Background and foreground inconsistencies differ in how they can be rejected. While foreground inconsistencies can be rejected by saying simply "that's not true", background inconsistencies cannot be rejected in this way. A more

take on the meaning[8]. We refer to the latter as meaning-accommodation. Modification by *addition* or *deletion* of fields may be necessary to make the meaning non-innovative.

As an example of meaning accommodation that shows the relation between this kind of learning and the case of corrective feedback described above, we will again use the modified "gloves" example. Here, we wish to illustrate what happens when a previously known word is encountered with a different meaning. We therefore assume, for the sake of argument, that Naomi initially has a concept of gloves. We will assume that Naomi takes "gloves" as having a perceptual type distinct for that of "mittens". However, again for the sake of argument, we assume that she is mistaken as to the nature of this difference; for example, she may disregard the difference in shape and instead think that mittens and gloves have different textures (e.g. that gloves are shiny whereas mittens are woolly).

$\dot{\bigwedge}[\text{glove}]_{Naomi}$ =

$$\begin{bmatrix} \text{x} : \textit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{x}) \\ \text{c}_{\text{physobj}} : \text{physobj}(\text{x}) \\ \text{c}_{\text{clothing}} : \text{clothing}(\text{x}) \\ \text{c}_{\text{shiny-texture}} : \text{shiny-texture}(\text{x}) \\ \text{c}_{\text{handclothing-shape}} : \text{handclothing-shape}(\text{x}) \end{bmatrix}$$

That is, Naomi thinks that mittens and gloves both have a common shape, but that they differ in texture. This means that the meaning of Father's utterance will be

$[\text{Those are nice gloves}]^{Naomi}$ =

$$\begin{bmatrix} \text{BG} = & \begin{bmatrix} \text{focobj} : \textit{Ind} \\ \text{c}_{\text{glove}} : \text{glove}'(\text{focobj}) \\ \text{c}_{\text{physobj}} : \text{physobj}(\text{focobj}) \\ \text{c}_{\text{clothing}} : \text{clothing}(\text{focobj}) \\ \text{c}_{\text{shiny-texture}} : \text{shiny-texture}(\text{focobj}) \\ \text{c}_{\text{handclothing-shape}} : \text{handclothing-shape}(\text{focobj}) \end{bmatrix} \\ \text{FG} = & \begin{bmatrix} \text{c}_{\text{nice}} : \text{nice}'(\text{FG.focobj}) \end{bmatrix} \end{bmatrix}$$

When encountering Father's utterance, we take it that the relevant take on the context for evaluating and understanding the utterance is something like

c^{Naomi} =

$$\begin{bmatrix} \text{focobj=a} : \text{Ind} \\ \text{c}_{\text{physobj}} : \text{physobj}(\text{focobj}) \\ \text{c}_{\text{clothing}} : \text{clothing}(\text{focobj}) \\ \text{c}_{\text{woolly-texture}} : \text{woolly-texture}(\text{focobj}) \\ \text{c}_{\text{handclothing-shape}} : \text{handclothing-shape}(\text{focobj}) \\ \text{c}_{\text{not-shiny-texture}} : \text{not}(\text{shiny-texture}(\text{focobj})) \end{bmatrix}$$

The $\text{c}_{\text{not-shiny-texture}}$ field can either result from consulting the environment by checking whether a shiny texture cannot be detected on focobj, or by inference from the proposition in $\text{c}_{\text{woolly-texture}}$.

Now, according to our definition of innovation, Naomi will detect a background inconsistency in that [Those are nice gloves].BG $\dot{\bigwedge} c_{Naomi} = \bot$. The inconsistency of course stems from the presence of a proposition (shiny-texture(focobj)) and its negation in the combined record. Contextual interpretation will thus fail, since the meaning-function cannot be applied to the context[9].

Figuring out exactly how to modify "glove′" in the above example is non-trivial and potentially depends on a complex mix of different contextual factors and other clues. It is fairly clear to English speakers what Naomi *should* do: the perceptual type assigned to "glove′" needs to be altered by taking the difference in shape between gloves and mittens into account, and by disregarding the mistaken importance of difference in texture. Naomi would then arrive at the concept of "glove′" presented above in section 5, which would yield a meaning of Father's utterance which would be consistent with Naomi's take on the context (which would now, as a consequence, disregard texture but include that focobj is glove-shaped). In fact, both meaning and context would be as described in section 8.

Note that as a result of contextual interpretation, Naomi has not only interpreted Father's utterance but as an effect of fixing an context-meaning mismatch she has also arrived at a new potential lexical meaning of "glove′".

The different aspects of meaning of an expres-

appropriate reaction to a background inconsistency is to issue a clarification request regarding the faulty expression, in our example something like "glove?".

[8] Arguably, a further option should be included: to adapt one's take on the context.

[9] In this description, we are leaving out a complicating aspect of contextual interpretation, namely how background fields may be coerced to the foreground if they do not correspond to anything in the context (Cooper, 2005). If one tried to apply a function with background inconsistency to a context this would trigger coercion of the inconsistent material to the foreground (since it does not match anything in the context) which would result in a foreground inconsistency.

sion are not independent. For example, additions to ontological semantics may be echoed by additions to perceptual type. For example, if the interaction in our modified "gloves" example continued with Father (anticipating Naomi's problems with figuring out the difference between gloves and mittens) saying (as in the original example) "when they have fingers in them they are called gloves and when the fingers are all put together they are called mittens", Naomi could use this as a basis for constructing a "glove-detector" based on existing knowledge of the appearance of fingers. Nevertheless, learning to properly identify gloves may take some time and require many exposures to instances of gloves.

10 Conclusion and future work

The work presented here is part of a research agenda aiming towards a formal account of semantic coordination in dialogue. Here, we have focused on formalising a notion of (semantic) innovation which is central to meaning accommodation. We have provided an example involving the detection of a subjectively innovative use of an expression, and the adaptation of a lexical meaning to accommodate this innovation.

Apart from filling in the details missing from the present account (such as a full compositional analysis), future work on meaning accommodation includes providing general formal accounts of (1) how to determine when to do meaning accommodation in reaction to a detected innovative use (as opposed to rejecting it), and (2) deciding exactly how to modify a concept to accommodate an innovative use.

Acknowledgments

This research was supported by The Swedish Bank Tercentenary Foundation Project P2007/0717, Semantic Coordination in Dialogue.

References

Patrick Blackburn and Johan Bos. 2005. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. CSLI Studies in Computational Linguistics. CSLI Publications, Stanford.

S. E. Brennan and H. H. Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22:482–493.

Samuel Carey and E. Bartlett. 1978. Acquiring a single new word. *Papers and Reports on Child Language Development*, 15:17–29.

H. H. Clark and D. Wilkes-Gibbs. 1986. Refering as a collaborative process. *Cognition*, 22:1–39.

E. V. Clark. 2007. Young children's uptake of new words in conversation. *Language in Society*, 36:157–82.

Robin Cooper and Staffan Larsson. 2009. Compositional and ontological semantics in learning from corrective feedback and explicit definition. In Jens Edlund, Joakim Gustafson, Anna Hjalmarsson, and Gabriel Skantze, editors, *Proceedings of DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue*.

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3:333–362.

Thierry Coquand, Randy Pollack, and Makoto Takeyama. 2004. A logical framework with dependently typed records. *Fundamenta Informaticae*, XX:1–22.

P.G.T. Healey. 1997. Expertise or expertese?: The emergence of task-oriented sub-languages. In M.G. Shafto and P. Langley, editors, *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pages 301–306.

D. Kaplan. 1979. Dthat. In P. Cole, editor, *Syntax and Semantics v. 9, Pragmatics*, pages 221–243. Academic Press, New York.

Pia Knoeferle and Matthew W. Crocker. 2006. The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30(3):481–529.

Staffan Larsson. 2007. Coordinating on ad-hoc semantic systems in dialogue. In *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue*.

Richard Montague. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven. ed. and with an introduction by Richmond H. Thomason.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–226, April.

Matthew Saxton. 2000. Negative evidence and negative feedback: immediate effects on the grammaticality of child speech. *First Language*, 20(3):221–252.

David Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Ronnie Smith and Jan Kuppevelt, editors, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.

Generalized quantifiers and clarification content

Robin Cooper
University of Gothenburg
Gothenburg, Sweden
cooper@ling.gu.se

Abstract

Purver and Ginzburg introduce the Reprise Content Hypothesis (RCH) and use it to argue for a non-generalized quantifier approach to certain quantifiers. Here we will contrast their approach with an approach which employs a more classical generalized quantifier analysis and examine what predictions it has for possible clarifications and reexamine the data which Purver and Ginzburg present in the light of this.

1 Introduction

Ginzburg (forthcoming) and previous work (Purver and Ginzburg, 2004; Ginzburg and Purver, 2008) introduce the Reprise Content Hypothesis (RCH) and use it to argue for a non-generalized quantifier approach to certain quantifiers. RCH comes in two versions and is stated in Ginzburg (forthcoming) as

RCH (weak) A fragment reprise question queries a part of the standard semantic content of the fragment being reprised.

RCH (strong) A fragment reprise question queries exactly the standard semantic content of the fragment being reprised.

They argue for the strong variant and then use this to draw consequences for the semantic content of quantified noun phrases in general, claiming that this provides a strengthening of the constraints placed on semantic interpretation by compositionality.

In this paper, we will question this conclusion, arguing that a more classical generalized quantifier analysis, recast in terms of type theory with records, not only provides a more adequate coverage of the basic compositional semantics but also accounts in an explanatory way for the reprise clarification data that Purver and Ginzburg cite. We will first consider (in Section 2) the anatomy of generalized quantifiers and the latest version of the Purver and Ginzburg proposal presented by Ginzburg (forthcoming). We will then look at some theoretical possibilities for how generalized quantifiers might be clarified (Section 3). We will then review the data concerning the clarification of quantifiers that Purver and Ginzburg have presented (Section 4) concentrating mainly on the kinds of clarifications they involve whereas Purver and Ginzburg concentrated on the types of clarification requests. Finally, we will draw some general conclusions about the relationship between clarifications and quantifiers in Section 5.

2 The anatomy of generalized quantifiers

The anatomy of quantified propositions can be characterized using TTR (Cooper, 2005; Cooper, forthcoming) and the analysis of non-dynamic generalized quantifiers presented in (Cooper, 2004) as the type in (1).

$$(1) \quad \begin{bmatrix} \text{restr} & : & \mathit{Prop} \\ \text{scope} & : & \mathit{Prop} \\ \text{c}_q & : & q(\text{restr},\text{scope}) \end{bmatrix}$$

Here we use the idea from type theory that the intuitive notion of proposition is represented by a type (known under the slogan "propositions as types"). In particular we shall use a record type as in the TTR analyses we have been developing. The idea is that the proposition is "true" if there is something of the type and "false" if the type is empty. The first field represents the restriction of the quantifier (corresponding to the common noun phrase such as *thief*) which is required to be of type "property". We take *Prop* to be an abbreviation for the function type $\left[\text{x:}\mathit{Ind}\right] \rightarrow \mathit{RecType}$, that is, the type of functions from records with a field

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

labelled 'x' for an individual to record types (corresponding to the intuitive notion of proposition).

The second field represents the scope of the quantifier, also required to be a property. If the quantifier corresponds to a noun phrase in subject position the scope corresponds to a verb phrase such as *broke in here last night*.

The third field represents a constraint requiring that a certain quantifier relation q hold between the two properties. For example, if q is the existential quantifier relation (corresponding to the English determiner *a* or singular count *some*) then the relation will hold just in case there is an object which has both the restriction and the scope property. q(restr,scope) also represents a type. We can think of it as the type of witnesses for the quantifier relation holding between the two properties. So in the case of the existential a witness would be something which has both properties 'restr' and 'scope'. If there is no such object then this type will be empty.

An object will be of this record type if it is a record containing at least three fields with the labels in the type (labels may only occur once in a record or record type) and values of the types required by the record type. Thus if there is no witness for the quantifier constraint type 'c_q' then there will not be anything of the quantified proposition type of the form (1) either. A particular example of a quantified proposition will be a refinement of the type in (1). If we represent the property of being a thief informally as '*thief*', the property corresponding to *broke in here last night* as '*bihln*' and the existential quantifier relation as $\exists$, then the type corresponding to the proposition corresponding to *a thief broke in here last night* would be (2).

$$(2)\quad \begin{bmatrix} \text{restr='}\textit{thief}\text{'} & : & \textit{Prop} \\ \text{scope='}\textit{bihln}\text{'} & : & \textit{Prop} \\ c_{\exists} & : & \exists\text{(restr,scope)} \end{bmatrix}$$

where the type *Prop* has been restricted to be the singleton type which contains exactly the property '*thief*' in the restriction field and the property '*bihln*' in the scope field. Note that what makes this a crucially generalized quantifier approach to quantified propositions is the use of the quantifier relation which holds between two properties and not the use of abstraction over properties in the compositional treatment of noun phrase interpretations according to Montague's style (Montague, (1974), Chapter 8: 'The Proper Treatment of Quantification in Ordinary English'). So, for example, the content of the noun phrase *a thief* will be a function from properties to record types where the scope field has been abstracted over:

$$(3)\quad \lambda P{:}\textit{Prop}\ \left(\begin{bmatrix} \text{restr='}\textit{thief}\text{'} & : & \textit{Prop} \\ \text{scope=}P & : & \textit{Prop} \\ c_{\exists} & : & \exists\text{(restr,scope)} \end{bmatrix} \right)$$

It is normally an object like that represented in (3) that is considered as a generalized quantifier, following the presentation of generalized quantifiers in Barwise and Cooper (1981) as sets of sets. It is this view that Purver and Ginzburg seek to argue against. However, it must be emphasized that the essential component of generalized quantifier theory is the use of the relation between properties (or sets) to represent quantification. The use of the lambda calculus in (3) can be regarded as a kind of glue to get the compositional semantics to work out. (This is the kind of view of the lambda calculus as a glue language which is presented by Blackburn and Bos (2005).) If you have another way to engineer the compositional semantics then you could abandon the kind of lambda abstraction used in (3) but still use the generalized quantifier notion of relations between sets.

Now let us consider (4).

$$(4)\quad \begin{bmatrix} \text{q-params:} \begin{bmatrix} \text{x:}\{\textit{Ind}\} \\ \text{r:most(x,student)} \end{bmatrix} \\ \text{cont:left(q-params.x)} \end{bmatrix}$$

This a representation for *most students left* proposed by Ginzburg (forthcoming) in his TTR recasting of the Purver and Ginzburg approach to quantification. It requires that there be a set 'x' which is what Barwise and Cooper (1981) would call a witness for the quantifier 'most(student)', that is, some set containing most students. In addition it requires that the predicate 'left' holds (collectively) for that set (which we may interpret as the predicate 'left' holding individually of each member of the witness).[1] This analysis is,

[1] Note that the notion of witness for a quantifier introduced by Barwise and Cooper is different from the notion of witness for a quantified sentence which we discussed above. The set of witnesses for a quantifier is the set of objects which potentially could be witnesses for the whole quantified sentence. A witness for the sentence will be a witness for the quantifier, but a witness for the quantifier will not necessarily be a

then, also a generalized quantifier analysis. It differs from the previous one in that it emphasizes the witness set and uses a different relation between sets for the quantifier relation, namely a relation between a witness set and the set corresponding to what we called the restriction previously. The witness quantifier relation is 'most' in (4). This analysis works well for monotone increasing quantifiers. However, as Purver and Ginzburg (2004) point out, it is more problematic with monotone decreasing quantifiers since there you have to check the witness set against the restriction and the scope in a different way. In that paper they go through a number of different options for solving the problem, finally coming to a preference for treating monotone decreasing quantifiers as the negation of monotone increasing quantifiers. That suggests to me that the representation for *few students left* corresponding to the analysis in (4) should be something like (5).

$$(5)\quad \begin{bmatrix} \text{c:}\neg(\begin{bmatrix} \text{q-params:} \begin{bmatrix} \text{x:}\{Ind\} \\ \text{r:many(x,student)} \end{bmatrix} \\ \text{cont:left(q-params.x)} \end{bmatrix}) \end{bmatrix}$$

that is, something that requires that there is no set x containing many students such that x (collectively) left. Now the only way I can think of to engineer the compositional semantics to achieve (5) is to have something along the lines of (6) corresponding to the noun phrase.

$$(6)\ \lambda P{:}Prop\ (\begin{bmatrix} \text{c:}\neg(\begin{bmatrix} \text{q-params:} \begin{bmatrix} \text{x:}\{Ind\} \\ \text{r:many(x,student)} \end{bmatrix} \\ \text{cont:}P\text{(q-params.x)} \end{bmatrix}) \end{bmatrix})$$

but this involves exactly the Montagueesque lambda paraphernalia that Purver and Ginzburg wish to avoid. However, as before, if you have an alternative way of engineering the compositional glue then you can apply it here as well while still maintaining the anatomy of quantification based on the witness quantifier relation.

Perhaps more difficult is the fact that the Purver-Ginzburg analysis also has difficulties with non-monotone quantifiers such as *only students* or *an even number of students* where it is not so clear that the negation strategy is available.

witness for the sentence. However, this distinction seems difficult to tease apart when looking at the clarification data and we will ignore it below.

This separation of the glue function of the lambda calculus and the analysis of quantified utterances in terms of generalized quantifier relations between sets leads me to suppose that Purver and Ginzburg's objection is not so much to generalized quantifiers as such as to the use of Montague's lambda calculus based approach to compositional semantics. This leads me to go back and reconsider their data in terms of the original generalized quantifier relation. Whereas they focussed their attention mainly on the clarification request, we will focus ours mainly on the clarification itself.

3 Potential clarification requests and clarifications

We might expect clarifications corresponding to each of the three fields, that is, the restriction, the scope and the quantifier constraint. In the case of the quantifier constraint we might expect the quantifier relation to be clarified or the witness. We consider two kinds of clarifications: the responses given to noun phrase reprise clarification requests and non-reprise clarification requests relating to quantifiers. Responses to noun phrase reprise clarification requests are exemplified by examples like

(7) A: A thief broke in here last night
B: A thief?
A: *a.* my ex-husband, actually (*witness*)
b. burglar wearing a mask (*restriction*)
c. got in through the bedroom window (*scope*)
d. two, actually (*quantifier relation*)

Seeing as we are focussed on the clarification request *A thief?* whose content is such that the scope field is abstracted over we might expect the scope clarification above to be less acceptable than the others.

Examples of non-reprise clarification requests relating to a quantifier are

(8) A: Somebody broke in here last night
B: *a.* (not) your ex-husband? (*witness*)
b. burglar wearing a mask? (*restriction*)
c. got in through the bedroom window? (*scope*)
d. just one? (*quantifier relation*)

Note that the availability (or not) of the restriction clarification question here could be important for distinguishing between a theory where clarification options are based on the content (where the restriction field is available) and a theory where clarification options are based solely on syntactic constituents of the preceding utterance, where in this example there is no common noun phrase in the relevant noun phrase *somebody*.

Intuitively it seems that clarifications not corresponding to a witness for the quantifier or one of the three fields introduced by the quantifier are harder to interpret as a clarification of the quantifier. Consider

(9) A: Somebody broke in here last night
B: *a.* maroon?
b. maroon sweater?
c. police?
d. scar over the left eye?

It is hard to give examples of impossible dialogues since there is no notion of grammaticality as there is with single sentences. What we can examine is the most likely interpretation given what we gather about the context from what we know about the dialogue. (9a) seems hard to interpret at all unless, for example *maroon* is being used (innovatively) as a way of characterizing skin-colour, in which case it would be a clarification relating to the restriction. A natural way of interpreting (9b) would be as elliptical for *wearing a maroon sweater* which would in effect coerce it to be a clarification of the restriction. Depending on the political situation in the country the dialogue is about (9c) might be interpreted as a restriction clarification, i.e. *Was it the police who broke in?*, or as a very elliptical way of asking whether A called the police. This latter interpretation could be facilitated, for example, if A and B routinely talked about break-ins and had a checklist of questions which they normally asked, among them whether the police was called. In this case, of course, (9c) would not be a clarification of the quantifier. Finally, (9d) is most naturally interpreted as elliptical for *with a scar over the left eye*, making it as a clarification of the restriction.

A central question is to what extent similar facts can be observed about generalized quantifiers which are not reducible to standard "referential" quantifiers and whether different classes of quantifiers behave differently with respect to the availability of clarification interpretations. Consider

(10) A: most thieves are opportunists
`http://www.accessmylibrary.com/coms2/summary_0286-33299010_ITM`, accessed 18th January, 2010
B: most thieves?
A: *a.* successful ones (*witness*/*restriction*)
b. bide their time (*scope*)
c. 80%, actually (*quantifier relation*)

Here the witness and restriction clarifications appear to collapse since a witness set for the quantifier has to be a subset of thieves (i.e. the restriction) which contains most thieves. However, (10a) does appear to be ambiguous between an interpretation corresponding to "successful thieves are opportunists" (a witness reading) and "most successful thieves are opportunists" (a restriction reading). Thus while the form of the clarification is the same its interpretation is ambiguous between a witness clarification and a restriction clarification.

In all of these examples, it seems that potential clarifications relating to the scope are intuitively less likely as "clarifications", as opposed to "additional relevant information". This seems natural as the quantifier itself does not contain the scope and it is therefore difficult to see information about the scope as clarification of the quantifier as such as opposed to the sentence as a whole. Our prediction is thus that clarifications of quantified noun phrases will fall into one of the following three classes:

(11) **Predicted clarification classes for quantified noun phrases**

- witness clarifications
- restriction clarifications
- quantifier relation clarifications

4 Some data

In Section 3 we showed some predictions for clarification made by a generalized quantifier approach to NP interpretation. In this section we will look at the examples that have been presented in the literature by Purver and Ginzburg and see to what extent they provide examples of the kinds of clarification we have predicted. As expected the Purver-Ginzburg data divides into witness clarification, restriction clarifications and quantifier relation clarifications. The large majority of cases are restriction clarifications. Scope clarifications do not occur in their data. We give details of the relevant examples below.

4.1 Witness clarifications

(12)

Unknown:	And er they X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
Unknown:	Chorlton?
Unknown:	**Chorlton**, mhm, he examined me, erm, he, he said now they were on about a slide ⟨unclear⟩ on my heart. Mhm, he couldn't find it.

BNC file KPY, sentences 1005–1008
(Purver and Ginzburg, 2004)

(13)

Terry:	Richard hit the ball on the car.
	. . .
Nick:	What ball?
Terry:	**James [last name]'s football**.

BNC file KR2, sentences 862, 865–866 (Purver and Ginzburg, 2004)

Intuitively both of these examples appear to be witness clarifications, although one might argue that this status is unclear. (13) might be arguably a combination of a restriction clarification (*ball→football*) and a quantifier relation clarification if we analyze *James [last name]'s* as a determiner representing a quantifier relation. One might argue that (12) is also a restriction clarification if you have an analysis of the proper name *Chorlton* as something corresponding to "the person named Chorlton".

4.2 Restriction clarifications

The clear cases of restriction clarifications presented below exhibit a number of different strategies for relating the clarification to the clarification request or the original utterance which seem quite closely related to repair strategies that have been noted in the literature.

(14)

George:	You want to tell them, bring the tourist around show them the spot
Sam:	The spot?
George:	**where you spilled your blood**

BNC file KDU, sentences 728–730
(Purver and Ginzburg, 2004)

Here additional material is provided which is to be added as a modifier to the restriction. Often the entire noun phrase is repeated with the additional modifier inserted, as in the following examples:

(15)

Terry:	Richard hit the ball on the car.
Nick:	What car?
Terry:	**The car that was going past**.

BNC file KR2, sentences 862–864
(Purver and Ginzburg, 2004)

(16) Anon 1: In those days how many people were actually involved on the estate?
Tommy: Well there was a lot of people involved on the estate because they had to repair paths. They had to keep the river streams all flowing and if there was any deluge of rain and stones they would have to keep all the pools in good order and they would
Anon 1: The pools?
Tommy: Yes the pools. That's **the salmon pools**
Anon 1: Mm.

BNC file K7D, sentences 307–313
(Purver and Ginzburg, 2004)

(17) Eddie: I'm used to sa-, I'm used to being told that at school. I want you ⟨pause⟩ to write the names of these notes up here.
Anon 1: The names?
Eddie: **The names of them**.
Anon 1: Right.

BNC file KPB, sentences 417–421
(Purver and Ginzburg, 2004)

(18) Nicola: We're just going to Beckenham because we have to go to a shop there.
Oliver: What shop?
Nicola: **A clothes shop**. ⟨pause⟩ and we need to go to the bank too.

BNC file KDE, sentences 2214–2217
(Purver and Ginzburg, 2004)

(19) is different in that it is the dialogue participant who contributes the original clarification request who provides alternative restrictions. Note that in this case the restrictions do not correspond to a syntactic constituent in the original utterance (*nothing*).

(19) Anon 1: Er are you on any sort of medication at all Suzanne? Nothing?
Suzanne: No. Nothing at all.
Anon 1: Nothing? **No er things from the chemists and cough mixtures or anything** ⟨unclear⟩?

BNC file H4T, sentences 43–48
(Purver and Ginzburg, 2004)

In (20) we have a case where a modifier in the original utterance is replaced by a new modifier in the clarification, thus changing what was said non-monotonically, not merely further specifying what was said.

(20) Elaine: what frightened you?
Unknown: The bird in my bed.
Elaine: The what?
Audrey: The birdie?
Unknown: **The bird in the window**.

BNC file KBC, sentences 1193–1197
(Purver and Ginzburg, 2004)

The whole of the restriction can be replaced in this way.

(21) Mum: What it ever since last August. I've been treating it as a wart.
Vicky: A wart?
Mum: **A corn** and I've been putting corn plasters on it

BNC file KE3, sentences 4678–4681
(Purver and Ginzburg, 2004)

Even though a different noun is chosen to express the restriction it can nevertheless be a refinement of the original utterance. In (22) the natural interpretation is the director is a woman.

(22) Stefan: Everything work which is contemporary it is decided
Katherine: Is one man?
Stefan: No it is a woman
Katherine: A woman?
Stefan: **A director who'll decide**.

BNC file KCV, sentences 3012–3016
(Purver and Ginzburg, 2004)

(23) seems to be a case where the speaker is searching for the right noun to express the restriction.

(23) Unknown: What are you making?
Anon 1: Erm, it's a do- it's a log.
Unknown: A log?
Anon 1: **Yeah a book, log book**.

BNC file KNV, sentences 188–191
(Purver and Ginzburg, 2004)

The final restriction clarification example is a little difficult to classify.

(24) Richard: No I'll commute every day
Anon 6: Every day?
Richard: **as if, er Saturday and Sunday**
Anon 6: And all holidays?
Richard: Yeah ⟨pause⟩

BNC file KSV, sentences 257–261
(Purver and Ginzburg, 2004)

We have interpreted it as if it involves a discussion of whether the restriction *day* is to mean weekdays or all days of the week and whether it is to include holidays. An alternative analysis might classify this as a quantifier relation clarification, that is, a discussion as to whether it really is *every* day that is meant.

4.3 Quantifier relation clarifications

In the data that Purver and Ginzburg present there appear to be two clear examples of quantifier relation clarifications.

(25) Anon 2: Was it nice there?
Anon 1: Oh yes, lovely.
Anon 2: Mm.
Anon 1: It had twenty rooms in it.
Anon 2: **Twenty rooms?**
Anon 1: **Yes**.
Anon 2: How many people worked there?

BNC file K6U, sentences 1493–1499
(Purver and Ginzburg (2004) cite it without the last turn)

We included the final turn to strengthen the interpretation that it is the quantifier relation which is being clarified. It seems hardly likely that the restriction *rooms* is in need of clarification.

(26) Marsha: yeah that's it, this, she's got three rottweilers now and
Sarah: **three?**
Marsha: **yeah**, one died so only got three now ⟨laugh⟩

BNC file KP2, sentences 295–297
(Purver and Ginzburg, 2004)

5 Conclusion

The data that Purver and Ginzburg present seem to fit the predictions of the basic generalized quantifier anatomy quite well. Those cases whose classification is unclear still provide alternative analyses which are within the range of predictions of the analysis. What does this say about RCH? RCH is a hypothesis about the content of a reprise, not about the clarification in response to a reprise which is mainly what we have looked at. Nevertheless, it would seem that the response to the clarification question could provide clues as to how it can be interpreted. Perhaps this shows that even if we strip away the issues concerning the use of the lambda calculus as a glue language, the original generalized quantifier approach is only consistent with the weak version of RCH. But it is not clear to me that the Purver and Ginzburg approach fares any better with respect to the strong version of the RCH when you look at the clarifications themselves as opposed to just the clarification requests.

What conclusions can we draw from this? RCH is, of course, not just a hypothesis about the content of reprises. Purver and Ginzburg want to use it as a way to argue for what the compositional semantic content of quantified noun phrases should be in general. The fact that their analysis seems to encounter problems with quantifiers that are not monotone increasing thus represents a challenge. It seems unadvisable to introduce the RCH as a constraint on semantic interpretation in addition to compositionality if as a consequence you cannot cover all the basic compositional data. It probably is possible to find a more witness set oriented analysis for non monotonic increasing quantifiers (see for example the discussion of witness sets for monotone decreasing quantifiers in Barwise and Cooper (1981)) but it would probably involve a loss of generalization in the compositional semantics, namely the classical generalized quantifier analysis of all quantifiers in terms of relations between properties (or sets) no matter what their

monotonicity properties are.

One of the advantages of using TTR is that you get structured semantic contents. Instead of the unstructured sets and functions of classical model theoretic semantics, you get articulated record types with labels pointing to various components. What the clarifications discussed in this paper seem to show is that speakers pick up on these meaning components even when they are not represented by separate syntactic constituents. It seems to me that this is an important part of a semantic theory of dialogue which is perhaps being obscured if we adopt principles like RCH which seems to be pointing to contents which you cannot break apart.

Acknowledgments

This research was supported in part by VR project 2009-1569, Semantic analysis of interaction and coordination in dialogue (SAICD). I would like to thank Jonathan Ginzburg and Staffan Larsson for useful discussion. This material was presented at the Logic and Language Technology seminar in Gothenburg and I would like to thank the audience for lively discussion, particularly Dag Westerståhl for pointing out an embarassing error.

References

Jon Barwise and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219.

Patrick Blackburn and Johan Bos. 2005. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. CSLI Studies in Computational Linguistics. CSLI Publications, Stanford.

Robin Cooper. 2004. Dynamic generalised quantifiers and hypothetical contexts. In *Ursus Philosophicus, a festschrift for Björn Haglund*. Department of Philosophy, University of Gothenburg.

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3:333–362.

Robin Cooper. forthcoming. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

Jonathan Ginzburg and Matt Purver. 2008. Quantification, the reprise content hypothesis, and type theory. In Lars Borin and Staffan Larsson, editors, *From Quantification to Conversation*. University of Gothenburg, Gothenburg.

Jonathan Ginzburg. forthcoming. *The Interactive Stance: Meaning for Conversation*.

Richard Montague. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven. ed. and with an introduction by Richmond H. Thomason.

Matt Purver and Jonathan Ginzburg. 2004. Clarifying noun phrase semantics. *Journal of Semantics*, 21(3):283–339.

Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology

Florian Hahn and **Hannes Rieser***
CRC 673, *Alignment in Communication*, Bielefeld University

Abstract

The paper discusses how a gesture typology can be extracted form the Bielefeld Speech-And-Gesture-Alignment corpus (SAGA) making use of the annotated gesture morphology in the SAGA data. The SAGA corpus is briefly characterized. Using a portion of a MM dialogue, the interface between speech and gesture is shown focussing on the impact of gestures on lexical definitions. The interface demonstrates the need for working out types of gestures and specifying their semantics *via* a partial ontology. This is started setting up a "typological grid" for 400 gestures. It yields a hierarchy of *n*-dimensional single gestures and composites of them. A statistical analysis of the grid results is provided and evaluated with respect to the whole SAGA corpus. Finally, it is shown how the typological results are used in the specification of the speech-gesture interfaces for the MM dialogue.

Keywords: completion, gesture typology, gesture-speech interface, partial ontology, SAGA

1 Motivation, Dialogue Example from SAGA, Plan of Paper

Do referential and iconic gestures have a specifiable meaning and function in multi-modal dialogue? Questions like these are normally handled in a descriptive way with respect to arbitrarily collected empirical data, for example in the gesture research based on a semiotic tradition as initiated by Ekman and Friesen (1969) and carried on in McNeill (1992) and Kendon (2004). We have built up a corpus of multi-modal data, the Bielefeld Speech and Gesture Alignment corpus, SAGA (see Lücking et al. 2010), completely annotated for referential and iconic gestures to deal with them in a systematic way. Below we give a short characterization of SAGA. We investigate several topics with respect to SAGA, focussing on gestures co-occurring with noun-phrases in MM dialogue:

(a) the types of gestures used and their function, for example production of a one-dimensional line,

(b) the semantic value a gesture represents like being the boundary line of an object,

(c) how the semantic value interfaces with an accompanied natural language expression such as *the church-window* and,

(d) how a multi-modal meaning can be built up from the meaning of the words and the gesture's meaning in a compositional way,

(e) how MM meanings behave in dialogue.

We want to answer question (b), the semantic value represented, serving as a precondition for answering (c) to (e), hence the interest in systematic typology. At the outset, we present a dialogue example (Fig. 2) being about two churches and their windows (Fig. 1) which illustrates the function of gesture and which will serve as our reference datum throughout the paper.

The specific SAGA setting for the example used has the following caracteristics: Two

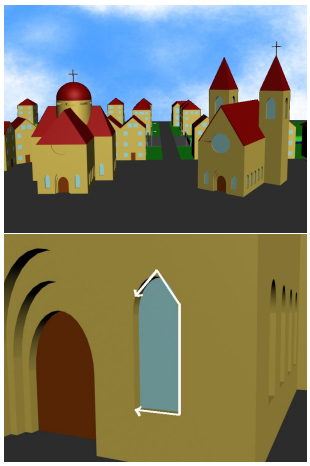


Figure 1: Looking at the VR-representation of the churches involved. The VR-stills show the two churches mentioned in the dialogue passage of Fig. 2. The Router's idea that the windows are gothic seems to come from the shadowing in the VR window representation. Anyway, the gesture used to depict the windows is fairly correct. This is indicated by the merge of the VR-window and the Router's gesture in the bottom still.

*Names in alphabetical order, 'MM' abbreviates 'multi-modal'. Corresponding author: Hannes Rieser, Hannes.Rieser@Uni-Bielefeld.de

R1 R2 R3

Dialogue Part 1

Router-Speech:	Beide	Kirchen	haben	diese	typischen	Kirchenfenster,	halt	unten
	Both	churches	have	these	typical	church-windows,	you know,	at the bottom
Router-Gesture:						Hedging	R1	R1
Follower-Speech:								
Follower-Gesture:								

Router-Speech:	eckig,	nach oben	dann	halt	so		gotisch	zulaufend.
	cornered,	upwards	moreover	you know	so		gothically	pointy
Router-Gesture:	R1	R2	R2	R2	R3			
Follower-Speech:						gotisch		
						gothic		
Follower-Gesture:			Nodding				Nodding	

Figure 2: The Router describing the church-windows to the Follower with the crucial bend at the top (R3).

agents, a Router, wearing body trackers, describing a route through a VR town and a Follower, trying to take up his description, face each other. The Router describes a landmark, two churches with gothic windows. The dialogue passage in Fig. 2 contains his multi-modal description of the church-windows. We present the German wording, the English translation and the stills showing the gestures in the way they accompany the words; we will rely on the English translation in the rest of the paper. The churches involved are depicted in Fig. 1.

What can we observe in the dialogue example? We treat the Router's contribution and the Router-Follower interaction.

The Router's Contribution: The Router starts gesturing parallel to wording *church-window*. Intuitively, he marks with his left hand (LH) the onset of an object and draws a line with his right hand (RH). The two-handed gesture R1 depicts some sort of corner. This goes on until the production of *upwards*. Concurrently with *upwards* his RH goes back to the onset held with LH and moves up, generating R2. Parallel to *so* the Router draws a hook R3, the top of a church-window.

The Router-Follower-Interaction: The Router's drawing is faster than his speech production, in addition, the German *halt* + *so*, engl. *you know* + *so* acts as a hesitation signal. Here the Follower comes in with her completion *gothic*. This is repaired by the Router extending it with *gothically pointy* (cf. Fig. 2).

So much for the data. We'll discuss the empirical findings from the point of view of corpus-based gesture typology and partial ontology. Section 2 provides background information concerning SAGA. Then we come back to the dialogue example again, investigating what we need for explaining speech gesture alignment in the dialogue (Section 3). Three statistics sections follow, one on the methodology of the typological grid (Section 4), the other on the statistics of it giving the frequency of types of gestures such as pointings, lines etc. in the grid (Section 5), and the third one an evaluation on how the statistical results for the grid can be generalized for the whole SAGA corpus (Section 6). Finally, we describe how the typology is used to provide explanations for the MM

dialogue part in Fig. 2 (Section 7).

2 Background on SAGA

The SAGA corpus contains 25 route-description dialogues taken from three camera perspectives. The setting comes with a driver, called "Router" "riding in a car" through a VR-setting, passing five landmarks connected by streets. After his ride the Router relates his experience in detail to a Follower supposed to organise her own trip following the Router's instructions. We collected video and audio data for both participants, for the Router in addition body movement tracking data due to markers on head, wrist, and elbow and eye-tracking data. The tracking data generated traces in Euclidean space and provided exact measurements for positions of head, elbow, and wrist. The gestures in the dialogues have all been annotated, the annotation predicates used like *indexing, modelling, shaping etc.* were rated.

The rating of these predicates is discussed in (Lücking et al. 2010). An example of a partial gesture annotation is given in Appendix, Fig. 3. The SAGA corpus contains ca. 6000[1] gestures. Roughly 400 gestures have been investigated in order to establish the typological grid described below.

3 What we Need for Explaining Speech-Gesture-Alignment in the Dialogue Example

3.1 The Router's Gestures

Consider Fig. 2. In R1, the Router's LH marks the left side of an object. Then he draws a horizontal line in three-dimensional space. R2 marks a vertical line starting at the corner set by the horizontal line and the left hand side producing a lower right angle. The vertical line ends in a fairly sharp bend. The only completely designed object is the bottom left "corner", all the other depictions are incomplete or underspecified.

3.2 Notes on the Speech-gesture Interface

We assume that gestures can be equipped with meaning, as a rule with a partial one and that gesture meaning interfaces with verbal meaning on different structural levels. Consequently, some gestures go with word meanings, others with the meaning of constituents, the meaning of dialogue acts, of rhetorical relations and so on. In previous papers we have shown how these various interfaces can be constructed (Rieser (2004, 2010), Rieser and Poesio (2009)) using type logics and compositional DRT. In this paper we will remain at the lexical and syntactic level. So, what we have to do is to observe the alignment of speech and gesture in these three cases all bound up with *church-window*:

Case 1:	you	know,	at	the	bottom	cornered
			R1	R1		R1
Case 2:	upwards,		moreover,		you know	
			R2		R2	R2
Case 3:	so					
	R3					

We consider the two occurrences of *you know* as meta-communicative acts[2] which can be inserted in multi-modal face-to-face communication before any complex constituent. *You know* can e.g. be taken as an attention-securing device, asking for acknowledgement or as focussing the relevant encyclopedic knowledge, whichever, it should be separated from the constituents contributing to the information describing the relevant event, i.e. in our case the Router's ride. So we have to care for the speech-gesture alignment of the rest.

3.3 Fine Grained Word Meaning, Partial Ontology, and Construction Meaning

After these considerations we turn to the details, first to the syntax of the router's contribution. The relevant part is shown in Appendix, Fig. 4. We have an N followed by two attribute-phrases AttrPhrs conjoined by *moreover*. The construction in Appendix, Fig. 4 is incomplete: It closes with an AdjPh where only the Adv *so* is realized. A completion is produced by the Follower. It is the Adj *gothic* which completes the AdjPh, see Fig. 4 for the cooperatively produced attribute. One should be aware of the fact that the scope of the completion is the "gappy" attr-phrase under construction by the Router. Roughly, we have to fuse in the end the meanings of *cornered at the bottom* and R1, *upwards* and R2, and *so* and R3. Below we provide lexicon definitions for *church-window*, and the morphology of the gestures R1, R2, R3 in-

[1] Due to filling gaps in our annotation and due also to re-annotation of material having already been annotated, both of which change the segmentation of gestures, the figures given for the number of gestures in SAGA have changed somewhat, roughly from 5000 plus to 6000.

[2] A paradigm which focuses on "inter-leavings" of meta-communicative and base-line material is Ginzburg (2010).

volved with their respective partial ontologies associated (*cf.* ch. 7).

Due to limits of space we concentrate in this paper on the lexicon definition of *church-window* and how it interfaces with the gestures R1-R3 without going into much detail.

Church-window: $\textit{church-window}(w) := \textit{window}(w) \wedge \textit{part-of}(w, wa) \wedge \textit{wall}(wa, ch) \wedge \textit{church}(ch) \wedge \textit{lower-part-of}(lp, w) \wedge \textit{middle-part-of}(mp, w) \wedge \textit{upper-part-of}(up, w) \wedge lp \oplus^{3} mp = cub \wedge \textit{cuboid}(cub) \wedge \textit{base}(b1, cub) \wedge \textit{breadth}(br, b1) \wedge \textit{breadth}(br, s1) \wedge \textit{side}(s1, lp) \wedge \textit{upper-part-of}(up, w) \Longleftrightarrow (((\textit{prism}(up) \wedge \textit{pointed}(up) \wedge \textit{acute}(up)) \vee (\textit{cylindric-section}(cls, up) \wedge \textit{round}(cls)))$.

A church-window *w* is part of a church *ch's* wall *wa* having a lower *lp*, a middle *mp* and an upper part *up*; the lower and the middle part form a cuboid *cub* with a base *b1* of a certain breadth *br* and a *side1*, part of *lp*, of the same breadth, the upper part *up* being either a pointed prism or a cylindric section. This gives us two versions of a church window, a gothic style one and a round arched one as depicted in Fig. 5 and Fig. 6, respectively.[4]

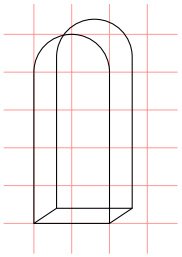


Fig. 5: The church-window as depicted by the right term of the disjunction $(\textit{cylindric-section}(cls, up) \wedge \textit{round}(cls))$

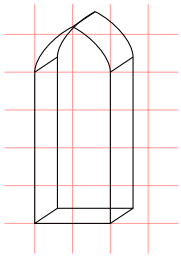


Fig. 6: The church-window as depicted by the left term of the disjunction $(\textit{prism}(up) \wedge \textit{pointed}(up) \wedge \textit{acute}(up))$

Fig. 7 represents the interface of speech and gesture in dialogue part 1. The arrows outside the pictures pointing towards the lexicon definitions indicate that gesture content operates on lexical content. In order to model that we would need the most basic gestural features and how they compose to make up gestures and their content. However, are there basic gesture features and complex ones observed by one speaker or even all speakers at all?

4 The Methodology of the Typological Grid

The question we ended up with in the previous section is: Are the gestures observed, sides, lines, the three-dimensional entities arising, arbitrary tokens, sporadic whims of the CPs' or are these systematically used at least throughout one datum (here video-film V5) by two agents or throughout the whole SAGA corpus by many or even all agents? In order to investigate both these typological questions we have set up a so-called typological grid for the datum V5 in the following way: intuitively, gestures build a space, consisting of simple and more complex gesture-morphological entities. The most fundamental entities we have are the individual annotation predicates like *hand-shape*, *wrist-movement* or *palm-direction*. For example, for the horizontal line starting at the left corner of the object in the example (see Fig. 2, R1 - R3, LH), we need the individual predicates *hand-shape*, *wrist-movement to the right* and *palm-direction facing down* or *left*. These are atomic features of the gesture space represented by AVM-matrices (App., Fig. 9). Only taken together do these single bits of information describe a horizontal line, again represented by a matrix (App., Fig. 10). The atomic features taken together form the most basic stratum of the gestural space (App., Fig. 17). They are derived from rated annotation predicates. Next in terms of complexity are the clusters: Clusters are bundles of functionally connected features, cf. section 5.2 for further motivation. 0-dimensional entities are introduced *via* the *indexing* annotation predicate. 1-dimensional entities originate from annotation predicates *drawing* or *modelling* and from wrist-movements. 2-dimensional entities, i.e. locations, regions and "more geometrical" 2D-objects are founded on *indexing*, *drawing* and *shaping*. 3-dimensional entities are based on *placing*, *shaping*, *drawing* or *modelling*.

Lines come with different bends and directions. They form the one-dimensional layer below the

[3] '$\oplus$' indicates mereological addition

[4] One of the reviewers is right in assuming that the lexicon definition is too near to the corpus data. A way out would be to leave the upper section of the church-window underspecified as to shape and to assume that it is the Router's gesture which resolves the underspecification with respect to *pointed* and *acute*. In Rieser (2010b) it is shown that the two agents resolve the underspecification in different ways, the Follower prefers the cylindric shape version which leads to a local inconsistency in the dialogue. The strategy of using underspecification techniques is different from the one followed in this paper which attributes priority to verbal information and investigates how the gesture "catches up" with it.

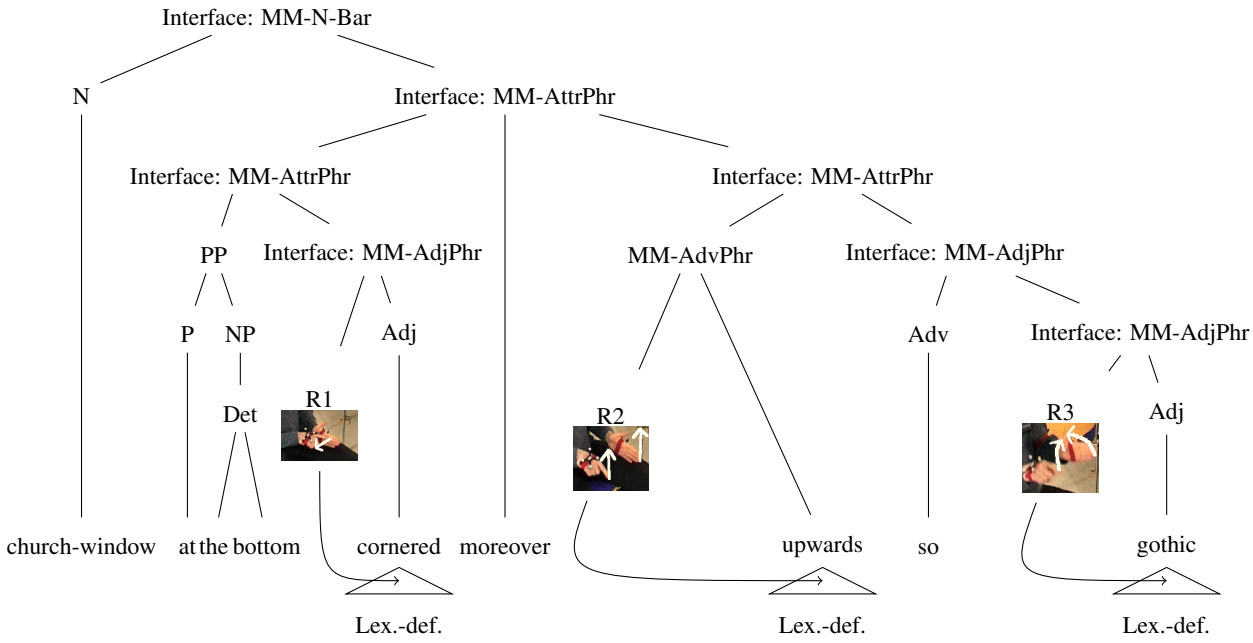


Figure 7: Schedule for the MM contributions and their interfaces. The interface points are at the level of lexical semantics and lead to MM phrases.

feature and cluster layers. The 0-dimensional entities represented by demonstrations have no spatial extension (App., Fig. 11). Furthermore, there are two-dimensional entities like rectangles, squares and so on (App., Fig. 12), followed in complexity by three-dimensional entities such as cuboids (App., Fig. 13). An interesting empirical fact is that we get composites of *n*-dimensional entities. The Router's gestures R1-R3, for example, form a composite of an object's left side, a single horizontal line and a composite line consisting of a vertical part and the bend. The composite depicts partial information about the upper part of a gothic church-window. There are many composites in the V5 datum, the functionally most conspicuous ones being the following: line touching circle orthogonally from the outside (Fig. 14), horizontal and orthogonal line meeting (Fig. 15), two three-dimensional objects held and set into relation to a third one introduced earlier on (Fig. 16).

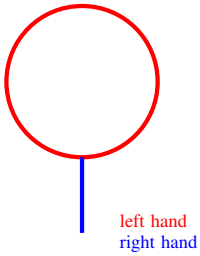


Fig. 14: Composite of two-dimensional and one-dimensional entity. LH holding a round object, RH drawing the path touching the circle.

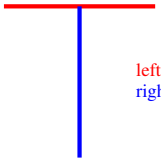


Fig. 15: Composite of two lines: LH modelling line orthogonal to path drawn continuously by RH. The point of contact produced is on the orthogonal line.

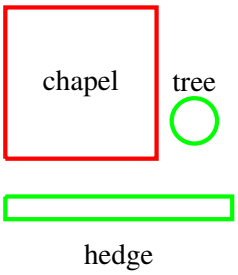


Fig. 16: Composite of 3 three-dimensional entities. LH holding a three dimensional entity (chapel) and RH placing the tree and shaping the hedge in front of the chapel.

5 Statistics of the Typological Grid

Out of the grid data the statistics shown in table 1 was computed. First we have a look at the differences in gesturing between the Router and the Follower. Consider the router's RH: It turns out that the Router mostly uses lines in RH (48%). In the second place come 2D-objects (locations, regions, circles, rectangles etc., 28%), three-dimensional objects in RH are next (prisms, cuboids, cylindroids, spheres etc., 15%) followed by 0-dimensional entities (abstract objects, 9%). The ranking of gestures for the Router's LH is as follows: 0-dimensional entities (7 %) $<$ one-dimensional entities (22%) $<$ three dimensional entities (34%) $<$ two dimensional entities (%

37). In the Follower's RH one-dimensional gestures (lines, 40%) dominate. To a lesser extent he uses two-dimensional (locations, 31%), 0-dimensional (abstract objects, 10%) and three-dimensional (prisms, spheres, 7%) gestures. With his LH he only shapes 0-dimensional objects (abstract objects, 50 %) and three-dimensional ones (prisms, spheres, 50%).

5.1 Interpretation: The Global Picture

Generally speaking, the Router concentrates on depicting routes, regions and locations as well as objects as (parts of) landmarks. Composites consisting of $n \geq 2$ gestures provide the possibility to "hold" the landmarks and sketch the route to them: at the same time both, landmark and route are relationally placed in Router's gesture space (see Figs. 14-16). Interestingly, the Follower sets up his interactive map using one-dimensional gestures most of the time. In other words, he concentrates on representing routes. With both, Router and Follower, the RH is dominating when gesturing (cf. table 1). The Router uses far more two-handed composites than the follower. He populates gesture space with more objects than the Follower does (cf. table 1). As a consequence, his gesture space embodies more information than the Follower's.

5.2 Interpretation: The Role of Features in Detail

The five features, *HandShape*, *BOHDirection*, *PalmDirection*, *WristPosition*, and *WristMovementDirection* are most frequently used by both Router and Follower in their LHs and RHs, respectively. Hence, the annotationally motivated grouping of the features *HandShape*, *BOHDirection* and *PalmDirection* into one *FeatureCluster* at the outset of the typology work (cf. Figs 9 - 13 in App.) gets statistical support. At the same time the large number of *WristPosition* features and *WristLocationMovementDirection* features motivates the set up of clusters for *WristPosition* and *WristMovement* respectively. Both Router and Follower predominantly use their RHs, to be inferred from the greater number of feature clusters there (App., Table 3).

6 Evaluation: Generalizing the Statistical Results for the Grid for the whole SAGA Corpus

The grid material provides a hierarchy of *n*-dimensional gesture categories. How can we use the grid results in establishing the overall SAGA statistics? So far, we can set up the following hypotheses:

1. The grid categories can be used for other data, even for those which do not belong to the SAGA corpus.[5]

2. We know the AVMs for e.g. 0-dimensional objects, one-dimensional objects etc. in detail which we have to look for in annotated material, presupposing, of course, similar annotation conventions. The assumption is that we'll get similar hierarchies for other (even new) data as we have in the grid, i.e. 0-dimensional, one-dimensional etc. objects, depending, of course, on the task. If we had a task dealing with planar objects only, we would presumably have not so many 3-dimensional gestures.

   A cursory investigation of the variables in the rest of the SAGA corpus has shown that this is true. We do indeed have objects of the *n* dimensions established for V5 in most of them. However, most probably, V5 shows the hierarchy in the most explicit way, in other films some layers seem to be missing, e.g. there are no 2- and 3-dimensional entities for the Follower in V1.

3. Concerning the video-datum we started from we can investigate whether the speech-gesture ensembles of the Router and the Follower are structured in a similar way. This is important for research into inter-personal alignment and the role of gesture in interaction.

[5] In reply to a question of one of the reviewers: So far, we are sure that this hypothesis is true of the other video films in the SAGA corpus but we have not tested it with respect to unseen data from different MM corpora. However, the hierarchy extracted from the SAGA data is very general, leading to the assumption that whenever a CP has a kind of linear information he can use a "line"-gesture and similarly for the other dimensions. A restriction concerning generalisation we presumably face could be due to the SAGA domain of concrete n-dimensional objects and routes between them. Whereas we expect that the hierarchy can be used for geometrical objects, CPs discussing sets, functions and λ-terms could well use different gestures.

Table 1: Statistics of the gesture morphology in the typological grid datum ordered by dimensions.

0-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	14	9	5	7	–	–	19	5
Follower	10	22	4	50	–	5	19	27

1-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	75	48	16	22	7	5	103	28
Follower	18	40	–	0	3	0	21	30

2-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	44	28	27	37	–	24	95	25
Follower	14	31	–	0	–	2[1]	16	23

3-Dim

	RH	RH%	LH	LH%	Composites	TWH Composites	total	total%
Router	24	15	25	34	–	67	116	31
Follower	3	7	4	50	–	–	7	10

Mixed-Composites

	OH	TWH	total	total%
Router	–	40	40	11
Follower	–	8	8	10

Totals

	total RH	total LH	total OH Comp.	total TWH Comp.	total
Router	157	73	7	136	373
Follower	45	8	3	15	71

[1] Note: Composites can occur without any corresponding single or one- handed gestures. In that case the composites can't be reduced to single or one-handed gestures. Therefore in this column we have TWH Composites but no LH gesture.

Here we have examples which show that such an investigation makes sense.

4. We can investigate whether the speech-gesture ensembles of other agents and of other pairs of agents in the corpus are structured in a similar way.

5. One can use the partial ontology set up for an *n*-dimensional gesture of the grid for a selected arbitrarily *n*-dimensional one from the rest of the corpus and test whether the former is adequate.

   This has been confirmed for lines and cuboids.

7 Application of Typology: Interfacing Verbal Meaning and Gestural Meaning in MM Dialogue

So far we have seen the following: A dialogue passage with gestures co-occurring with speech and the gesture typology for one complete datum, V5, which gives us a hierarchy of gestures ranging from 0-dimensional entities to *n*-ary composites. The issue we face now is "How can we use the typology in explaining the meaning of multimodal discourse?" We associate the gestures R1-R3 with their descriptions in terms of gesture morphology, both attributes and values. Attributes and values, for example *HandShape*-LH and *Bspread* in R1 are fused into a new attribute where the original value is still "visible" as a suffix. This new attribute is given a stipulated semantic value *side(s1)*

$\wedge$ *of(s1, z)* $\wedge$ *brdth(br, s1)*, in terms of the grid hierarchy, a two-dimensional entity, a side *s1* of something z having a breadth *br*. In a similar manner, the *PathofWrist*-attribute is associated with some length *l1* and the *TwoHandedConfiguration* with a right angle. Of course, not every information contained in the stipulation is derived from the typology but the typology serves as a precondition for the partial ontology.

R1	
HandShape-LH-Bspread	$side(s1) \wedge of(s1, z) \wedge brdth(br, s1)$
PathofWrist-RH-LINE>LINE>LINE	$length(l1, bb) \wedge of(bb, z)$
TwoHandedConfiguration-RFTH>BHA>RFTH>BHA	$right\text{-}angle(s1, ra, bb)$

brdth abbreviates breadth. LH signs a side *s1* of an object *z*. So, there is something z of which *s1* is a side. RH provides a path *l1* emerging from the side *s1* to the right. Both hands produce the right angle of an object *z* shaped by the side *s1* and the planar object *bb*. Intuitively, we have depicted an object *z* with a corner formed by the planar object *bb* and the side *s1*.

R2	
HandShape-LH-Bspread	$side(s1) \wedge of(s1, z) \wedge brdth(br, s1)$
PathofWrist-RH-LINE	$height(h1, ss2)$
WristMovementDirection-RH-MU	$height(h1, ss2)$

LH continues to hold the side *s1*. RH signs the height *h1* of an object *ss2*. The second and the third attribute-value pair provide the same information.

Note that it would be incorrect to identify variables *s1* and *ss2*.

R3	
HandShape-LH-*Bspread*	$side(s1) \wedge of(s1, prr) \wedge brdth(br, s1)$
PathofWrist-RH-*LINE>LINE*	$edge(e1, prr) \wedge edge(e2, prr)$
WristMovementDirection-RH-*MR/MU>MD/MR*	$angle(e1, a, e2) \wedge acute(a)$

The side *s1* is still held by LH. The RH produces two edges *e1* and *e2* of an object *prr* which form an acute angle at the top.

Linking up R1-R3 with the *church-window* Property:

The overall speech-gesture meaning integration can be derived from Fig. 7. We cannot show that in detail here. The final structure of the MM meaning of the N-Bar construction plus the accompanying gestures is:

$window(w) \wedge$ *part-of*$(w, wa) \wedge wall(wa, ch) \wedge church(ch) \wedge$ *lower-part-of*$(lp, w) \wedge$ *middle-part-of*$(mp, w) \wedge$ *upper-part-of*$(up, w) \wedge lp \oplus mp = cub \wedge cuboid(cub) \wedge base(b1, cub) \wedge breadth(br, b1) \wedge side(s1, lp) \wedge$ (*upper-part-of*$(up, w) \Leftrightarrow (prism(up) \wedge pointed(up) \wedge acute(up))) \wedge$ $\underline{at(w, lp)} \wedge \underline{\text{lower-part-of}(lp, w)} \wedge \underline{side(s12, lp)} \wedge$ $\underline{\text{right-angle}(b1, ra, s12)} \wedge \underline{length(l1, s12)} \wedge \underline{height(h1, ss2)}$.

Comparing this result with the lexicon-entry for *church-window* introduced in 3.3 shows that the lexicon-entry remained consistent. Only the portion marked by underlining is additional information. It contains the information needed for the lower part of the church-window. So we see that iconic gestures can highlight which aspect of word meaning is intended from the CP's point of view, thus supporting parts of the more analytic lexical definition. This is especially clear from the disjunction: Only the "pointed" and "acute" version is consistent with the gesturing.

Acknowledgements

Work on this paper has been carried out in the project B1, *Speech-gesture Alignment*, of the CRC *Alignment in Communication*, Bielefeld University. Support by the German Research Foundation is gratefully acknowledged. We also want to express our thanks to our fellow researchers Kirsten Bergmann, Stefan Kopp and Andy Lücking, as well as to three anonymous SEMdial reviewers.

References

Bergmann, K., Fröhlich, C., Hahn, F., Kopp, St., Lücking, A. and Rieser, H. 2007. *Wegbeschreibungsexperiment: Grobannotationsschema.* Bielefeld Univ.

Bergmann, K., Damm, O., Fröhlich, Hahn, F., Kopp, St., Lücking, A., Rieser, H. and Thomas, N. 2008. *Annotationsmanual zur Gestenmorphologie* Bielefeld Univ.

Ekman, P. and Friesen, W. 1969. The repertoire of nonverbal behavior: categories, origins, usage and coding. *Semiotica 1*, pp. 49-98

Kendon, A. 2004. *Gesture: Visible Action as Utterance.* CUP.

Kita, S. (Ed.) 2003. *Pointing: Where Language, Culture and Cognition Meet.* Lawr. Erlb.

Lücking, A., Bergmann, K., Hahn, F., Kopp, St., Rieser, H. 2010. The Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). Submitted for LREC 2010.

McNeill, D. 1992. *Hand and Mind. What Gestures Reveal About Thought.* Univ. of Chic. Press

Ginzburg, J.: 2010. *The Interactive Stance. Meaning for Conversation*. MS, King's College London, (to appear).

Poesio, M. and Rieser, H.: 2010. An Incremental Model of Anaphora and Reference Resolution Based on Resource Situations. (Submitted to DD, SI on Incrementality).

Poggi, I. 2002. From a Typology of Gestures to a Procedure for Gesture Production. In: Wachsmuth, I. and Sowa, T. (eds.), *Gesture and Sign Language in Human-Computer Interaction*. LNAI 2298, pp. 158-168

Rieser, H. 2004. Pointing in Dialogue. In *Proceedings of Catalog* '04, pp. 93-101.

Rieser, H.: 2010a. On Factoring out a Gesture Typology from the Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). In Kopp, St. and Wachsmuth, I. (Eds.), *Proceedings of GW 2009,* Springer: Berlin, Heidelberg, pp. 47-60.

Rieser, H.: 2010b. How to Disagree on a Church Window's Shape Using Gesture. In: *Proceedings of SEMdial 2010.*

Rieser, H. and Poesio, M.: 2009. Anaphora and Direct Reference. Empirical Evidence from Pointing. In: P*roceedings of DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue.* Stockholm, Sweden, pp. 35-43.

Appendix

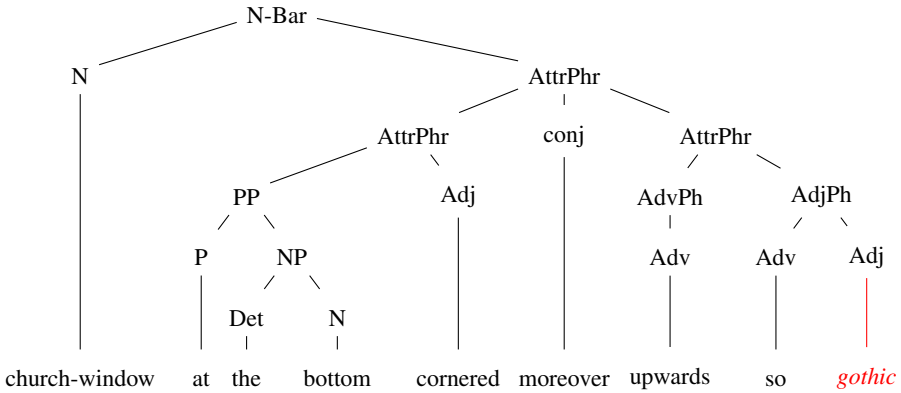


Figure 4: Represents the completed NP with the Follower's *gothic*, triggered by the Router's gesture, inserted. Observe that *in toto* we have a cooperatively produced AdjPh.

Table 2: Non-zero-valued features used in the typological grid.

Features	**Router**		**Follower**	
	LH	RH	LH	RH
PathOfWrist	3	3	2	3
WLMDirection	10	14	2	9
WLMRepetition	2	3	1	2
WristDistance	2	3	2	2
WristPosition	9	13	11	18
BackOfHandDirection	11	11	4	10
BOHMDirection	5	5	1	1
BOHMDRepition	1	1	1	1
PathOfBOH	3	3	1	1
PalmDirection	14	13	5	12
PDMDirection	3	3	1	1
PDMRepitition	1	1	1	1
PathOfPalm	2	2	1	1
HandShape	22	22	4	12
HSMRepitition	1	2	1	1
HSMDirection	1	6	1	1
PathOfHandshape	1	2	1	1
TemporalSequence	1	2	1	1
TWH Features	both hands		both hands	
TWH-Configuration	13		9	
TWH-Movement	6		3	

Table 3: Clusters used in the typological grid.

Cluster	**Router**		**Follower**	
	LH	RH	LH	RH
PMovement	3	2	1	1
HSMovement	1	4	1	1
BOHMovement	4	6	1	1
FeatureCluster	71	109	6	33
WristMovement	26	50	2	16
WristPosition	34	52	12	30
TWH Cluster	both hands		both hands	
TWH-Cluster	35		19	

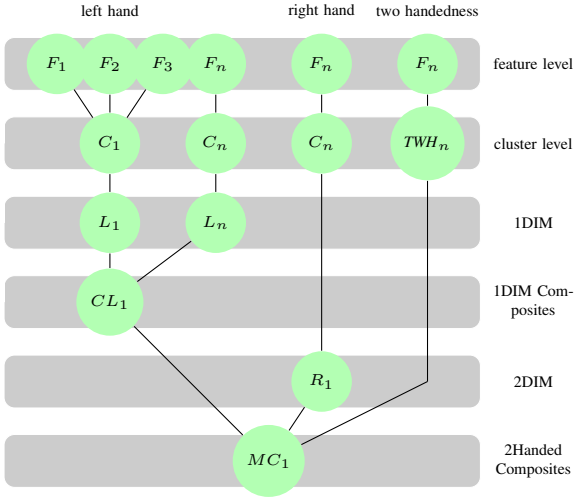


Figure 17: Section of gesture hierarchy (simplified).

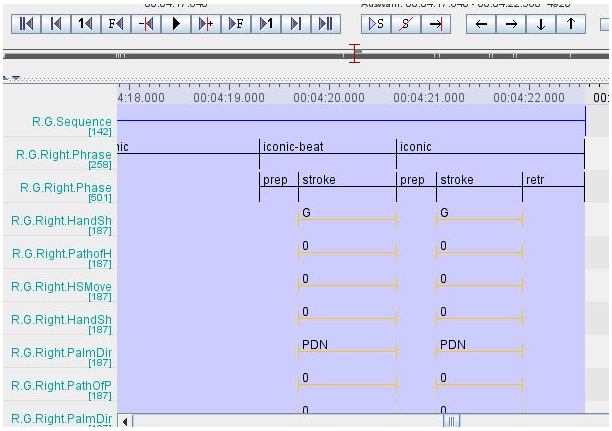


Figure 3: Example of the gesture annotation. It partially represents R1 and R2 from Fig. 2.

R-Line-RH		
R-FeatureCluster-RH-1a	*R-FeatureCluster-RH-1a-cat*	
	HandShape	*G*
	PalmDirection	PDN
	BackOfHandDirection	BAB
R-FeatureCluster-RH-2a	*WristMovement-RH-1a-cat*	
	PathofWrist	*Line>Line>Line*
	WristLocationMovementDirection	*MR>ML >MR*
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3a	*WristPosition-RH-1a-cat*	
	WristPos	CLW
	WristPosDist	DEK

HandShape	*G*

Figure 9: *Feature HandShape* and its value *G*.

Figure 10: Gesture representing a horizontal line.

R-Direction-G212-RH		
R-FeatureCluster-RH-1a	*R-FeatureCluster-RH-1a-cat*	
	HandShape	*G*
	PalmDirection	PDN
	BackOfHandDirection	BAB/BTL
R-FeatureCluster-RH-2a	*WristMovement-RH-1-cat*	
	PathofWrist	$\emptyset$
	WristLocationMovementDirection	$\emptyset$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3a	*WristPosition-RH-1-cat*	
	WristPos	CUP
	WristPosDist	DEK

Figure 11: 0-dimensional entity direction.

R-Rectangle-LH		
R-FeatureCluster-LH-1c	*R-FeatureCluster-LH-1a-cat*	
	HandShape	*B spread*
	PalmDirection	*PTR*
	BackOfHandDirection	BAB
R-FeatureCluster-LH-2a	*WristMovement-LH-1a-cat*	
	PathofWrist	$\emptyset$
	WristLocationMovementDirection	$\emptyset$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-LH-3c	*WristPosition-LH-1a-cat*	
	WristPos	CLL
	WristPosDist	DEK

Figure 12: 2-dimensional entity rectangle or side.

R-Cuboid-G103-RH		
R-FeatureCluster-RH-1c	*R-FeatureCluster-RH-1az-cat*	
	HandShape	*small C*
	PalmDirection	*PAB*
	BackOfHandDirection	*BUP*
R-FeatureCluster-RH-2a	*WristMovement-RH-1ax-cat*	
	PathofWrist	$LINE > LINE > LINE$
	WristLocationMovementDirection	$MF > MB > MF$
	WristLocationMovementRepetition	$\emptyset$
R-FeatureCluster-RH-3c	*WristPosition-RH-1q-cat*	
	WristPos	*CC*
	WristPosDist	$DCE > DEK$

Figure 13: 3-dimensional entity cuboid.

Early Interpretation by Implicature in Definite Referential Descriptions

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam

Abstract

This paper discusses the processes by which dialogue participants incrementally compute the conveyed meaning of referential descriptions. I show that some of the phenomena observed in the psycholinguistics literature can be accounted for by theories of conversational implicature that exploit ingredients from computational models of Referring Expression Generation. The result is a system that computes inferences incrementally, from partial utterances, without need to reason with hypothesized complete descriptions.

1 Introduction

It is widely accepted that utterances in conversation often communicate information that goes beyond the conventional meaning conveyed by the linguistic expressions used. A good deal of the processes that help dialogue participants reduce the gap between the conventional meaning of utterances and the enriched meanings actually intended by speakers has to do with *inference*—a notion which is at the core of Gricean pragmatics and the theory of conversational implicature (Grice, 1975; Grice, 1989). It is also generally agreed that in conversation speakers and addressees produce and understand language incrementally, in (at least) a word-by-word fashion rather than in one go once, say, the end of an utterance has been reached. The question thus arises as to whether theories of conversational implicature (which, as most semantic and pragmatic theories, were originally designed to operate at the utterance level) can be accommodated within the *incremental turn*.

In this paper I look into how dialogue participants incrementally compute the intended meaning of referring descriptions (i.e. their intended referent). I show that a slightly modified version of Hirschberg (1985)'s computational theory of scalar implicature that allows us to compute implicatures at the sub-utterance level can account for incremental effects observed in psycholinguistic experiments.

I start by giving an overview of previous work on the incremental processing of referential descriptions regarding both resolution and generation. In Section 3 I survey experimental results that indicate that information that goes beyond conventional semantic meaning is used incrementally at the sub-utterance level. To account for the pragmatic inferences observed, I explore an account that combines ingredients from generation models with a theory of scalar implicature. After introducing the rudiments of such a theory and some basic semantic processing in Section 4, I sketch my proposal in Section 5 and apply it to some examples.

2 Definite Referential Descriptions

Referential descriptions in the sense of Donnellan (1966)—i.e. definite descriptions that serve the purpose of letting the addressee identify a particular entity out of a set of entities assumed to be in the current focus of attention—have been studied extensively in dialogue research, specially in the context of *referential matching tasks* such as those used in the psycholinguistic experiments of Krauss and Weinheimer (1966) and Clark and Wilkes-Gibbs (1986). Work within the collaborative model of grounding put forward by Clark and colleagues (Clark, 1996) has emphasized the fact that the form and meaning of these descriptions often depends on *historic* aspects such as conceptual pacts established with specific conversational partners during the course of interaction (Brennan and Clark, 1996). Thus these approaches have mostly focused on *subsequent mentions*, i.e. descriptions that refer back to entities mentioned previously, or that refashion earlier descriptions that were not

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

grounded (Clark and Wilkes-Gibbs, 1986).

In the present paper I concentrate on the arguably simpler case of *first mention* referential descriptions. The main reason for this is that experiments that study incremental processing at the sub-utterance level—which is our focus here—have essentially investigated first mention descriptions in simple tasks, where the referring goal can be achieved with little interaction,[1] by means of a single referential description. Several approaches have investigated this kind of descriptions from the generation and the resolution perspective. In the remainder of this section, I summarize some of them and then give an overview of some experimental results reported in the psycholinguistics literature.

2.1 Resolution

A good deal of computational work on incremental interpretation of referring descriptions—such as the early proposals of Mellish (1985) and Haddock (1989) as well as more recent approaches such as Schuler (2003)—models incremental reference resolution as a symbolic process of constraint satisfaction, where predicates are associated with sets of constraints. The core idea is that, as a referring expression is processed from left to right, the constraints introduced by each predicate in the expression progressively narrow down the set of potential referents. Consider, for instance, the description '*the black wooden chair*'. Here processing '*black*' would eliminate from the set of potential referents those elements in the context that are not black; processing '*wooden*' would narrow down that set further to the subset of black elements that are made of wood; while finally processing '*chair*' would pick up the chairs among the black wooden elements.

These resolution models thus focus on computing semantic denotation in a model-theoretical way, largely ignoring any aspects related to pragmatics and implicature. However, as we shall see in Section 3, hearers can incrementally use pragmatic information that goes beyond conventional interpretations to identify referents at stages where there is semantic ambiguity, thus speeding up the process of establishing the speaker's meaning. Constraint-based models can naturally be enriched with pragmatic constraints, as done, for instance, by DeVault and Stone (2003), who complement the conventional content with goal and and plan-related inferences. I shall follow a similar approach and extend a constraint-based semantic model with scalar implicatures.

A different trend of approaches (e.g. Roy (2002) and Schlangen et al. (2009)) explore probabilistic models of reference. In this case, the referential potential of linguistic expressions is learned from data by exploiting statistical correlations between the linguistic expressions used and particular referential configurations (i.e. the context of utterance with its set of potential referents and knowledge about which referent is the intended one). Models of this sort thus do not make a clear distinction between semantics and pragmatics: They capture expectations about linguistic meaning and cooperative behaviour implicitly and with the same mechanisms.

2.2 Generation

The Generation of Referring Expressions (GRE) is one of the key areas within the field of Natural Language Generation. Traditionally, researches in this area have attempted to generate *minimal descriptions*, i.e. the shortest possible descriptions that succeed in uniquely identifying the intended target referent in a given context. For instance, if the context includes one black chair and one brown table, the description '*the black chair*' is not minimal, while the description '*the chair*' is. The idea is that minimal descriptions are consistent with Grice's Maxim of Manner, in particular with the Sub-maxim of Brevity: "be brief; avoid unnecessary prolixity" (Grice, 1975). Since the shorter description '*the chair*' succeeds in identifying the referent, the presence of the predicate '*black*' is considered *redundant* and hence susceptible of generating false implicatures by violating the maxim of Brevity.

The problem with this approach is that it does not take into account the fact that interpretation is a continuous process and ignores the possibility of reasoning incrementally: whether a predicate is considered redundant or not is determined by reasoning with *complete* descriptions. The predicate '*black*' is considered redundant in the complete description '*the black chair*' because there is an alternative, shorter, complete description that does without it. This view thus misses the point that what may count as redundant upon completion of

[1] Although see Brown-Schmidt et al. (2005) for a promising attempt to use eyetracking methods in interactive settings.

an utterance can be informative during incremental processing.

The seminal work of Dale and Reiter (1995) addresses precisely this issue. The Incremental Algorithm proposed by these authors uses properties in a predefined *preference order* incrementally, as long as they have *discriminatory power*, i.e. as long as they rule out some distractors—elements that are not the intended referent—from the context set. For instance, in our earlier example, '*black*' has incremental discriminatory power because at the point when the modifier is uttered it rules out the table, which is not black. Thus, if colour is a particularly salient property (it is ranked high in the preference order), the use of a colour predicate such as '*black*' can help the hearer to more easily identify the intended referent—a point also made by Grosz (1981): "*A speaker should be redundant only to the degree that redundancy reduced the total time involved in identifying the referent.*" The optimization of property preference orderings and the use of properties that are redundant *a posteriori* but that do have incremental discriminatory power are issues that are actively being investigated in GRE research (Viethen et al., 2008; Krahmer et al., 2008).

3 Psycholinguistic Evidence

In a series of experiments, Sedivy and colleagues (Sedivy et al., 1999; Sedivy, 2003) have used the eye-tracking paradigm to investigate how humans interpret instructions that contain referential descriptions with different types of modifiers, including colour (e.g. *black*, *red*), material (e.g. *plastic*, *wooden*), and scalar adjectives (e.g. *tall*, *big*). In these experiments, subjects wearing a head-mounted eye-tracker are shown an array of objects and are asked to pick up one of them with instructions such as '*Pick up the plastic spoon*'. Since subjects direct their gaze towards potentially referred objects, the precise information provided by the eye-tracker offers direct evidence about the alternative referents that are being considered by a subject at precise points in time and about the point at which a commitment to an interpretation is made.

The displays used in the experiments were such that upon hearing the adjective, more than one referent was possible while some referents could be discarded. For instance, one sample scenario included two objects that were made of plastic—the target spoon and a comb—plus a tie and a bulb as distractors. As expected, upon hearing the adjective subjects looked at the two objects that matched its semantic content (the spoon and the comb). The more interesting result was that in scenarios where a contrasting object was also in the display (e.g. a metal spoon) a preference for the target plastic object could be observed before the head noun was uttered, regardless of the semantic indeterminacy. That is, in these scenarios the identification of the target object took place *earlier*, at a point when the ongoing utterance was still semantically ambiguous. Experiments also show that this effect, let us call it *"contrastive bias"*, does not obtain with all kinds of modifiers: hearers did not show this kind of bias when interpreting colour adjectives, while they did for descriptions with material and scalar adjectives.[2]

With regard to an experiment that included also production, Sedivy (2003) reports that what distinguished modifiers that gave rise to a contrastive bias from those that did not was the frequency with which that kind of modifier was spontaneously generated to describe an object in a context where modification was not required for unique identification. That is, modifiers such as colour adjectives that are often produced "redundantly"—redundantly *a posteriori*, but that may still help to reduce the search space for the hearer incrementally—do not lead to a contrastive bias, while those that are typically used only when they are needed for unique identification of the referent are understood as such and hence give rise to a contrastive inference.

This latter result seems to indicate that the observed contrastive bias could be successfully modelled by a probabilistic approach that is sensitive to the statistical correlations in the data. However, a follow-up experiment demonstrated that this is not entirely trivial. Grodner and Sedivy (forthcoming) found that when subjects were explicitly told beforehand that the speaker suffered from an impairment leading to linguistic deficits, they did not show evidence of a contrastive bias regardless of the statistical patterns in the data.[3] This seems

[2] Throughout the paper, I will ignore scalar adjectives and exemplify my points with colour and material modifiers. The special features of gradable adjectives, which I have addressed elsewhere (Fernández, ms), are not critical for the approach presented here.

[3] Results of a similar nature regarding disfluencies are reported by Arnold et al. (2007): subjects who hear disfluent descriptions infer that the referent is difficult to describe;

to indicate that the explanatory power of statistical regularities is limited.[4]

In her discussion of the experimental results I have surveyed, Sedivy (2007) seems to appeal to an idea of redundancy that—in line with the point made in Section 2.2—relies on reasoning with hypothesized complete descriptions: "*Thus, when the display contained a referent for which the use of a modifier was communicatively motivated, people showed a preference for this referent compared to displays in which there was no clear reason to refer to the same target referent using a modifier*". Here I shall adopt a different perspective and show that these results, as well as the differences observed for different types of modifiers, can be accounted for by combining elements from computational models of Referring Expression Generation with ingredients from standard theories of scalar implicature to arrive at a system that operates incrementally, without the need to reason with complete descriptions.

4 Preliminary Notions

Before moving on to sketch an account of the effects I have described, I shall first introduce the rudiments of scalar implicature theories, the notation I will use, and the basic process of incremental semantic interpretation I assume.

4.1 Scalar Implicature

Scalar implicature is a kind of conversational implicature (Grice, 1975) whose computation is dependent upon the identification of some salient relation that orders a concept referred to in an utterance with other concepts of the same type. The idea is that the use of an expression in the scale (i.e. the ordering) implicates that (the speaker believes that) the other expressions in the scale (often considered stronger) do not apply.

Following the seminal work of Horn (1972) and others such as Gazdar (1979), Hirschberg (1985) proposes a theory of scalar implicature that specifies the conditions under which a speaker may license a scalar implicature and a hearer may infer it. In her theory, scalar implicatures are calculated from surface semantic representations of complete utterances by (1) identifying a potential scalar sub-formula in the logical form, (2) identifying the scale or scales that this subformula belongs to, and (3) inferring negative implicatures for alternate and higher values in the scale(s).

I shall essentially follow this approach,[5] but allow for the possibility of incrementally computing scalar implicatures from sub-formulas as they become available during incremental processing, and for the possibility of computing the implicatures from formulas other than those strictly corresponding to the logical forms of utterances (this should become clearer in Section 5.2).

As extensively discussed by Hirschberg (1985), scales can be of several types. I consider two types of scales: alternate scales $\{\sigma, \sigma', \ldots\}$ containing scalar expressions that are not ordered but simply contrast with each other, and linear scales $\langle\sigma, \sigma', \ldots\rangle$ containing expressions that are linearly ordered according to some suitable relation.

The inference rule for scalar implicature I assume is the following, where ψ and $\psi[\sigma/\sigma']$ are identical except for the fact that all occurrences of expression σ in ψ have been substituted by σ' in $\psi[\sigma/\sigma']$:

(1) SCALAR IMPLICATURE INFERENCE RULE
Given a formula ψ, a scalar expression σ in ψ, and a scale S that includes σ:
$\forall\sigma'.((\sigma' \in S \wedge \sigma' > \neq \sigma) \rightarrow \neg\psi[\sigma/\sigma'])$

We can use the rule in (1) to compute, for instance, the scalar implicatures inferred from the utterances in (2) and (3) (indicated by $\rightsquigarrow$) by considering, in the standard way, that ψ corresponds to the semantics of the whole utterance.

(2) Some people left the party early.
a. S : $\langle$`all`, `some`$\rangle$
b. $\rightsquigarrow$ Not all people left the party early.

(3) A: Do you have apple juice?
B: I have grape, tomato or bloody mary mix.
a. S : {`grape`, `tomato`, `bloodymm`, `apple`}
b. $\rightsquigarrow$ I don't have apple juice.

4.2 Domain Representation

I model the domain in a way similar to how input databases are modelled in GRE systems, i.e. characterising entities in terms of attributes and values.

such inferences are however cancelled when subjects are told speakers suffer from a linguistic impairment.

[4] Or perhaps that an adequate probabilistic model should also take into account the cooperativity of the speaker.

[5] I employ a simplified version of the original approaches, which amongst other things ignores epistemic operators, but which suffices to illustrate the points that occupy us here.

I assume a first-order-logic system with all the usual logical symbols, including equality $=$ and a top $\top$ symbol; a domain of entities U; a set of relational symbols A corresponding to attributes (such as `colour`, `material`, `type`); and a set of constant symbols V corresponding values (such as `blue`, `red`, `plastic`, `metal`, `spoon`, `comb`). A function $Val : A \to \mathcal{P}(V)$ assigns to each attribute a set of appropriate values. Variables e, e' range over elements in U, variables $\texttt{att}, \texttt{att}'$ over elements in A, and variables $\texttt{val}, \texttt{val}'$ over elements in V. The interpretation function assigns to each $\texttt{att} \in A$ a relation $U \times Val(\texttt{att})$. I write $\texttt{att}(e) = \texttt{val}$ (instead of the more standard notation for relations $\texttt{att}(e, \texttt{val})$) to express that element e is related to value `val` by attribute `att`.

4.3 Incremental Interpretation

I shall use formulas $\texttt{att}(e) = \texttt{val}$ as logical forms of adjectives and nouns in definite referring descriptions. For instance, '*red*' will be interpreted as $\texttt{colour}(e) = \texttt{red}$; the extension of this expression is then the set of red elements in the context.

The semantic interpretation of a description such as '*the red plastic cup*' then proceeds as shown in (4). I use the symbol $\top$ to initialise the existentially quantified formula introduced by the definite article.[6]

(4) t_0 The t_1 red t_2 plastic t_3 cup t_4
$t_1 : \exists e.\top$
$t_2 : \exists e.\texttt{colour}(e) = \texttt{red}$
$t_3 : \exists e.\texttt{colour}(e) = \texttt{red} \wedge \texttt{material}(e) = \texttt{plastic}$
$t_4 : \exists e.\texttt{colour}(e) = \texttt{red} \wedge \texttt{material}(e) = \texttt{plastic} \wedge \texttt{type}(e) = \texttt{cup}$

This incremental process is in line with the symbolic constraint-based approaches to incremental reference resolution described in Section 2.1. Expressions add constraints that incrementally narrow down the set of potential referents. My aim is to complement this semantic process with default pragmatic inferences that can be computed using the ingredients of scalar implicature theories. I turn to this in the next section.

[6] I do not include the presupposition of unique existence in the semantic representation.

5 Early Interpretation by Implicature

In this section I show how we can use the main elements of the theory of scalar implicature I have sketched to account for the early interpretation effects observed in the psycholinguistic experiments described in Section 3.

5.1 Scales

Determining what is a possible salient scale in a given situation can be a tricky issue. In general expressions within an ordering share a common type (e.g. they are quantifier expressions, of juice types). But even so, it is not trivial to determine which elements of the relevant type can be considered part of a scale that is salient for both speaker and hearer. This point is emphasized by Benotti and Traum (2009) in their account of comparative implicatures, where they opt for deriving the simplest possible scale $\langle \texttt{no}, \texttt{yes} \rangle$ for scalar adjectives such as '*safe*' in comparative constructions.

I take *attributes* and *values* to be scalar expressions, i.e. expressions that can be associated with a scale of related concepts. Values give rise to alternate scales $\{\texttt{val}, \texttt{val}', \ldots\}$ containing different values relevant for one attribute, while attributes take part in ordered scales $\langle \texttt{dim}, \texttt{dim}', \ldots \rangle$ that rank several attributes according to salience. Here I shall directly borrow the notion of "property preference ordering" (or "list of preferred properties" (Dale and Reiter, 1995)) from REG systems and assume that the hearer's representation of the contextual domain (like the generator's) includes such a scale of attributes. I will however make minimal assumptions about the elements that belong to a particular scale. By definition, salient scales include the triggering scalar expressions that have been overtly uttered. For instance, an utterance of say '*plastic*' with logical form $\texttt{material}(e) = \texttt{plastic}$ can give rise to two scales:

(5) a. $S_1 : \{\ldots, \texttt{plastic}, \ldots\}$
b. $S_2 : \langle \ldots, \texttt{material}, \ldots \rangle$

The question is then how these scales that are assumed to be part of the common ground of speaker and hearer are further populated. Since we are concerned with a visually shared situation, I will define default rules for including additional expressions into a scale that rely only on properties of the shared visual context.[7] In particular,

[7] Of course other aspects may render expressions salient

we assume that if an expression is *witnessed* in the shared visual context it can enter a contextual scale. Intuitively, a value such as `red` is witnessed if the visual context includes an entity that is red. Similarly, an attribute such as `colour` is witnessed if the context includes an entity for which that attribute is applicable. As mentioned, attributes take part in ordered scales—preference orderings. Preferred attributes are those that are more salient. Here I equate salience with ease of perception: in a shared visual context, an attribute is salient if it can be easily perceived (its values easily discriminated) by both speaker and hearer. I use $>_{sal}$ to denote the preference ordering.

The following scale construction inference rules make more precise what has just been explained informally. Note that for ordered scales, the only attributes that are relevant are those that are more salient than the attribute evoked by the utterance itself.[8]

(6) SCALE CONSTRUCTION DEFAULT RULES
Given a (sub-)utterance u with semantics $\texttt{att}(e_i) = \texttt{val}$, and potential scales $S_1 : \{\ldots, \texttt{val}, \ldots\}$ and $S_2 : \langle \ldots, \texttt{att}, \ldots \rangle$:
a. $\forall e_j\ \texttt{val}'.\texttt{att}(e_j) = \texttt{val}' \wedge \texttt{val}' \neq \texttt{val} \wedge e_i \neq e_j \rightarrow \texttt{val}' \in S_1$
b. $\forall e_j\ \texttt{att}'.\texttt{att}'(e_j) = \texttt{val} \wedge \texttt{att}' >_{sal} \texttt{att} \rightarrow \texttt{att}' \in S_2$

These two types of scales—alternate value scales and scales of preferred attributes—are of a rather different nature and I shall asume that they are used in different ways by the SCALAR IMPLICATURE INFERENCE RULE given in (1) above. The implicatures inferred from value scales are the classic scalar implicatures computed from alternate scales such as that shown in (3). These scalar implicatures are standard in the sense that they can be computed from input formulas ψ that directly correspond to logical forms of (sub-)utterances without any further assumptions.[9]

5.2 Contrastive Inferences as Scalar Implicatures

We will now look into how the machinery we have in place allows us to infer scalar implicatures that can account for the contrastive bias that helps the hearer identify the intended referent at an early stage, regardless of the semantic ambiguity. The intuitive idea I want to explore is that these scalar inferences are only drawn if the attribute evoked by a sub-utterance has *incremental discriminatory power*, as introduced in Section 2.2—that is, if the context includes another element that has a different value for that attribute (and that hence would get eliminated from the set of potential referents). More formally, an expression such as '*plastic*' in a partial utterance such as '*the plastic...*' with semantics (7a) has discriminatory power if (7b) is supported by the context (where $e \neq e'$):

(7) a. $\exists e.\texttt{material}(e) = \texttt{plastic}$
b. $\exists e'.\texttt{material}(e') \neq \texttt{material}(e)$

The SCALAR IMPLICATURE INFERENCE RULE (repeated below for convenience) can then be exploited to enrich (7b).

(8) SCALAR IMPLICATURE INFERENCE RULE
Given a formula ψ, a scalar expression σ in ψ, and a scale S that includes σ:
$\forall \sigma'.((\sigma' \in S \wedge \sigma' >\!\neq \sigma) \rightarrow \neg\psi[\sigma/\sigma'])$

The clause in the matrix of (7b) ($\texttt{material}(e') \neq \texttt{material}(e)$) acts as input formula ψ for the rule, while the attribute `material` instantiates σ. Since the inferred implicatures ($\neg\psi[\sigma/\sigma']$) characterise both the intended referent and the additional element required for the expression to have discriminatory power, they fall under the scope of both quantifiers introducing these elements:

(9) $\exists e.\texttt{material}(e) = \texttt{plastic} \wedge$
$\exists e'.\texttt{material}(e') \neq \texttt{material}(e) \wedge$
$\rightsquigarrow \neg\,[\sigma'(e') \neq \sigma'(e)] \ \wedge$
$\rightsquigarrow \neg\,[\sigma''(e') \neq \sigma''(e)] \wedge$
$\ldots$

In order to illustrate how this works with a concrete example, let us consider the sample scenario described in Section 3. Recall that the visual context contains at least three elements: a plastic spoon, a plastic comb, and a tie—let's assume the latter is made of silk. Let's also assume that in this context, object `type` is a more salient attribute (e.g. can be more easily perceived) than `material` and that therefore upon processing the fragment '*the plastic...*' the preference scale $\langle \texttt{type}, \texttt{material} \rangle$ is evoked. The conventional meaning of this fragment is shown in (10a). There

for both speaker and hearer.

[8] Note that these are only *default* rules. Certainly elements can be part of a salient scale for less overt and immediately accessible reasons.

[9] Besides those that allow us to construct the relevant scale.

are two witnesses that make this formula true in the current context—the spoon and the comb—and hence there is semantic ambiguity. However the modifier has incremental discriminatory power since the context is consistent with (10b). Now, if the hearer does not have reasons to believe that the speaker is not cooperative (or capable of being so), the SCALAR IMPLICATURE INFERENCE RULE can be used to enrich (10b) with further inferences, as long as they are supported by the context. Given the preference scale $\langle \texttt{type}, \texttt{material} \rangle$, the rule can generate the implicature in (10c), which is equivalent to $\texttt{type}(e') = \texttt{type}(e)$.

(10) The plastic ...
  a. $\exists e.\texttt{material}(e) = \texttt{plastic} \wedge$
  b. $\exists e'.\texttt{material}(e') \neq \texttt{material}(e) \wedge$
  c. $\rightsquigarrow \neg[\texttt{type}(e') \neq \texttt{type}(e)]$

A context with an additional, contrasting element such as a metal spoon would support the implicature in (10c) and thus would allow the hearer to disambiguate the semantics in favour of the plastic spoon (since the only assignment that makes (10) consistent with the context is one where the plastic spoon is assigned to e). Thus, we are able to predict the *contrastive bias* reported by Sedivy and colleagues and to account for the fact that hearers are able to make predictions about potential referents incrementally, at a point when the ongoing utterance is still semantically ambiguous, without need to reason with hypothesized complete descriptions.

Resorting to scales akin to the property preference orderings typically used in GRE models also allows us to account for the differences observed with different kinds of modifiers. We can explain this by appealing to the relative position of different types of modifiers within the preference scale, specially with respect to the attribute type.[10] If an attribute is highly prominent and there is no other attribute higher up in the preference scale, then it will not give rise to the contrastive implicature, even though it may still have discriminatory power. It seems reasonable to assume that in shared visual situations often colour is very salient, arguably even more salient than the object type (11c). Thus, as observed by Sedivy, in a context with, say, a red plate, a red cup, and a blue cup, upon hearing '*Pick up the red...*' hearers do not exhibit any contrastive bias (i.e. no preference for the red cup is observed).

(11) ...the red ...
  a. $\exists e.\texttt{colour}(e) = \texttt{red}$
  b. $\exists e'.\texttt{colour}(e') \neq \texttt{colour}(e)$
  c. $S : \langle \texttt{colour}, \texttt{type} \rangle$

In this case, the scalar implicature is not inferred since there is no higher-ranked attribute that would license the application of the SCALAR IMPLICATURE INFERENCE RULE.

[10] At least judging from the limited experimental conditions tested by Sedivy and colleagues

6 Conclusions

In this paper I have discussed how pragmatic inferences can be exploited during incremental interpretation in the resolution of first-mention referential descriptions. I have concentrated on contrastive effects observed in the psycholinguistics literature and have sketched a proposal that combines elements from computational models of Referring Expression Generation with ingredients from standard theories of scalar implicature. The result is a system that operates incrementally on partial utterances, without need to reason with complete descriptions. Clearly, the main burden of the approach is the determination of the preference scale—a problem that REG models also face. Although the present paper offers only a preliminary account, it hopefully contributes to opening the door for investigating further how pragmatic theories can meet the challenges imposed by the incremental nature of language use in dialogue.

Acknowledgements

I am grateful to the anonymous SemDial reviewers for their comments. This research has been supported by a VENI grant funded by the Netherlands Research Council (NWO).

References

Jennifer E. Arnold, Carla L. Hudson Kam, and Michael K. Tanenhaus. 2007. If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology Learning Memory And Cognition*, 33(5):914.

Luciana Benotti and David Traum. 2009. A computational account of comparative implicatures for a spoken dialogue agent. In *Proceedings of the Eight International Conference on Computational Semantics*, pages 4–17, Tilburg, The Netherlands, January.

Susan Brennan and Herbert Clark. 1996. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology*, 22(6):1482–1493.

S. Brown-Schmidt, E. Campana, and MK Tanenhaus. 2005. Real-time reference resolution in a referential communication task. In JC Trueswell and MK Tanenhaus, editors, *Processing world-situated language: bridging the language-as-action and language-as-product traditions*, pages 153–172. MIT Press.

Herbert Clark and Donna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

Herbert Clark. 1996. *Using language*. Cambridge University Press.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 18:233–266.

David DeVault and Matthew Stone. 2003. Domain inference in incremental interpretation. In *Proceedings of ICoS 4: Workshop on Inference in Computational Semantics*, pages 73–87, Nancy, France. INRIA Lorraine.

Keith S. Donnellan. 1966. Reference and definite descriptions. *The philosophical review*, 75(3):281–304.

Raquel Fernández. ms. Semantics and Pragmatics in Incremental Reference Resolution: The Case of Relative Gradable Adjectives. Unpublished manuscript.

Gerald Gazdar. 1979. *Pragmatics: Implicature, Presupposition, and Logical Form*. NewYork: Academic Press.

H. Paul Grice. 1975. Logic and conversation. In D. Davidson and G. Harman, editors, *The Logic of Grammar*, pages 64–75. Dickenson, Encino, California.

Paul Grice. 1989. *Studies in the Way of Words*. Harvard University Press.

David Grodner and Julie Sedivy. forthcoming. The effect of speaker-spefici information on pragmatic inferences. In N. Pearlmutter and E. Gibson, editors, *The Processing and Acquisition of Reference*. MIT Press, Cambride, MA.

Barbara Grosz. 1981. Focusing and description in natural language dialogues. In A. K. Joshi, I. Sag, and B. Webber, editors, *Elements of Discourse Understanding*. Cambridge University Press.

Nicholas J. Haddock. 1989. Computational models of incremental semantic interpretation. *Language and Cognitive Processes*, 4(3):337–368.

Julia Hirschberg. 1985. *A Theory of Scalar Implicature*. Ph.D. thesis, University of Pennsylvania, Computing and Information Sciences.

Laurence Horn. 1972. *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, University of California, Los Angeles.

Emiel Krahmer, Mariët Theune, Jette Viethen, and Iris Hendrickx. 2008. GRAPH: The costs of redundancy in referring expressions. In *Proceedings of the 5th International Conference on Natural Language Generation*, Salt Fork OH, USA.

R.M. Krauss and S. Weinheimer. 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3):343–346.

Christopher S Mellish. 1985. *Computer Interpretation of Natural Language Descriptions*. Chichester: Ellis Horwood.

Deb Roy. 2002. Learning visually grounded words and syntax for a scene description task. *Computer speech and language*, 16(3):353–385.

David Schlangen, Timo Baumann, and Michaela Atterer. 2009. Incremental reference resolution: The task, metrics for evaluation, and a bayesian filtering model that is sensitive to disfluencies. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue*, London, UK, September.

William Schuler. 2003. Using model-theoretic semantic interpretation to guide statistical parsing and word recognition in a spoken language interface. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics (ACL)*, pages 529–536, Sapporo, Japan.

Julie Sedivy, Michael Tanenhaus, Craig Chambers, and Gregory Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71:109–147.

Julie Sedivy. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of psycholinguistic research*, 32(1):3–23.

Julie Sedivy. 2007. Implicature during real time conversation. *Philosophical Compass*, 2/3:475–496.

Jette Viethen, Robert Dale, Emiel Krahmer, Mariet Theune, and Pascal Touset. 2008. Controlling redundancy in referring expressions. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC' 08)*, pages 950–957.

On the distributed nature of mutual understanding

Dale Barr
University of Glasgow

Abstract

Mutual understanding is one of the most important topics in the study of language use; it is also one of the most perplexing. How is it that people understand one another given the fundamentally ambiguous nature of communication? Why do language users find conversation so effortless and unproblematic, while theories of language use suggest that it requires inordinately complex processes and representations? I will suggest that such paradoxes arise out of a tendency to localize processes of mutual understanding in the minds of individual language users. Instead, I will suggest that the work of mutual understanding is distributed more broadly, over individual, interactional, and cultural levels of language use. This approach offers a new way of understanding of the functional significance of certain psycholinguistic phenomena, such as apparent failures of perspective taking in referential communication.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Relevance for Dialogue

Jonathan Ginzburg
King's College, London
London, UK
jonathan.ginzburg@kcl.ac.uk

Abstract

Relevance in the sense of *conversational coherence* is the most fundamental notion for research on dialogue. It is the cornerstone of theories of dialogue in the same way that *grammaticality* is to syntax. In this paper I restrict attention to relevance relating a query to a possible (felicitous) response. Still, even restricted to this domain, attempts at a comprehensive characterization of relevance, difficult as they undoubtedly are, are few and far between. Indeed, most existing accounts are intrinsically restricted in their ability to scale up. I offer a number of arguments for the need for a notion of relevance internalized in some way within the theory of meaning: relevance seems to underpin certain types of clarification questions and *lack of conversational relevance* seems to underpin the inference that one does not wish to address a prior utterance. I sketch an account of relevance within the dialogue theory KoS underpinned by Type Theory with Records.

1 Introduction

Relevance in the sense of *conversational coherence* is the most fundamental notion for research on dialogue. It is the cornerstone of theories of dialogue in the same way that *grammaticality* is to syntax. Indeed (Turing, 1950) proposed that the ability to evince relevance in approximately this sense could be a plausible test for intelligence. In what follows, I restrict attention to relevance relating a query to a possible (felicitous) response.[1] This is, in part, due to obvious considerations of space, but also because this is a domain where considerable work has been done on one component of the problem. Still, even restricted to this domain, attempts at a comprehensive characterization of relevance, difficult as they undoubtedly are, are few and far between. Indeed, as I will explain below, most existing accounts are intrinsically restricted in their ability to scale up.

Beyond this, one issue to consider is whether there really is a need for a single notion of relevance—whose restriction to query moves we discuss here, internalized in some way within the theory of meaning.[2] There is at least one substantive argument for internalizing relevance, as well as some methodological motivation. The substantive argument is that a unitary notion of conversational relevance seems to underpin certain types of clarification questions—ones that arise when the coherence of an utterance seems unclear, as in ((1)a).[3] Similarly, as Grice famously pointed out, *lack of conversational relevance* seems to underpin the inference that one does not wish to address a prior utterance, as in ((1)b):

(1) a. Marjorie: Don't touch that cos she hasn't had it yet.

[1] For a more detailed account see (Ginzburg, 2011).

[2] I use 'theory of meaning' to avoid boring territorial disputes between semantics and pragmatics. I attempt, nonetheless, to be reasonably explicit as to whether components of the theory of relevance refer to public context or to agent–internal parameters, which is essentially how I view the distinction.

[3] An empirical caveat is in order here. 'What do you mean' is clearly NOT a purpose built CR for querying the relevance of an utterance. In practice, the vast majority of 'what do you mean' CRs, at least in the BNC, seem to be about literal content, NOT about coherence:

(i) Anon 6: No, there's nobody here much Richard: What do you mean there's nobody here, it's packed.

(ii) Cassie: You did get off with him? Catherine: Twice, but it was totally non-existent kissing so Cassie: What do you mean? Catherine: I was sort of falling asleep.

Pretheoretically this is perhaps not surprising, given that (perceived) complete lack of coherence is rare; whereas indeterminacy of content is a consequence of lexical and phrasal context dependence, but a detailed explanation is surely an important desideratum.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Dorothy: Does she eat anything? Marjorie: What do you mean? (British National Corpus (BNC))

b. Dr. Grimesby Roylott: My stepdaughter has been here. I have traced her. What has she been saying to you?
Sherlock Holmes: It is a little cold for the time of the year.
Dr. Grimesby Roylott: What has she been saying to you?
Sherlock Holmes: But I have heard that the crocuses promise well. ('The Speckled Band', Sir Arthur Conan Doyle, *The Adventures of Sherlock Holmes*, John Murray, London.)

The methodological argument is that characterizing relevance pushes theories of dialogue to be concrete, forcing them to be precise about the range of propositions they characterize as answers and to offer sources of relevance to utterances whose relevance as an answer they do not underpin. It also enables one to operationalize the notion of relevance for use in corpus studies and for other computational work.

In this paper I sketch an account of relevance within the dialogue theory KoS (Ginzburg, 1994; Ginzburg and Cooper, 2004; Larsson, 2002; Purver, 2006; Fernández, 2006; Ginzburg, 2011; Ginzburg and Fernández, 2010). The basic intuition is that relevance of an utterance relative to an agent's information state amounts to the possibility of integrating the utterance into the information state; though as we will see this basic intuition needs to be refined to deal with cases like ((1)b). I start by offering a more or less theory neutral characterization of relevance, suggesting the need to encompass (in approximate order of theoretical difficulty)

- *q*(uestion)-specificity—this includes both answerhood and some sort of dependence or entailment relation between questions,
- metadiscursive relevance (a notion that underwrites utterances like "I don't know" and 'I don't want to talk about this.")
- genre-based relevance, the latter much studied in AI work on dialogue
- metacommunicative relevance, a notion that underwrites clarification interaction.

Thus, defining relevance involves interplay between semantic ontology, grammar, and interaction conventions. Various frameworks where relevance merely ties in the content of utterances (e.g. (Groenendijk and Roelofsen, 2009; van Benthem and Minica, 2009)) and even (Asher and Lascarides, 2003), which has admirably wide coverage, seem intrinsically unable to scale up to deal with metacommunicative relevance. Characterizing relevance requires a theory that allows ontology, grammar, and interaction to be encoded within the interaction conventions. For this purpose I employ Type Theory with Records (TTR) (Cooper, 2005). KoS and TTR are introduced in section 3. After which I sketch an attempt to combine the various notions of relevance so that they can be used to explicate examples of the type (1) above.

2 A Five Step Approach to Analyzing Relevance

2.1 Step 1: answerhood

In constructing our notion of relevance for queries, the first step is the most familiar. Any speaker of a given language can recognize, independently of domain knowledge and of the goals underlying an interaction, that certain propositions are *about* or *directly concern* a given question. This, I suggest, is the answerhood relation needed for characterizing interrogative relevance. It must be sufficiently inclusive to accommodate conditional, weakly modalized, and quantificational answers, all of which are pervasive in actual linguistic use, as in the following BNC examples:

(2) a. Christopher: Can I have some ice-cream then?
Dorothy: you can do if there is any. (BNC)

b. Anon: Are you voting for Tory?
Denise: I might. (BNC, slightly modified)

c. Dorothy: What did grandma have to catch?
Christopher: A bus. (BNC, slightly modified)

d. Elinor: Where are you going to hide it?
Tim: Somewhere you can't have it.

How to formally and empirically characterize aboutness is an interesting topic researched within work on the semantics of interrogatives (see e.g. (Ginzburg and Sag, 2000; Groenendijk, 2006)), though a comprehensive, empirically-based account is still elusive.

2.2 Step 2: q-specificity

The second step we take is somewhat less familiar and already a bit trickier. Any inspection of corpora, nonetheless, reveals the underdiscussed fact that many queries are responded to with a query. A

large proportion of these are clarification requests, to be discussed in section 2.5. But in addition to these, there are query responses whose content directly addresses the question posed, as exemplified in ((3)):

(3) a. A: Who murdered Smith? B: Who was in town?

b. A: Who is going to win the race? B: Who is going to participate?

c. Carol: Right, what do you want for your dinner?
Chris: What do you (pause) suggest? (BNC, KbJ)

d. Chris: Where's mummy?
Emma: Mm?
Chris: Mummy?
Emma: What do you want her for? (BNC, KbJ)

There has been much work on relations among questions within the framework of *Inferential Erotetic Logic* (IEL) (see e.g. (Wiśniewski, 2001; Wiśniewski, 2003)), yielding notions of q(uestion)–*implication*. From this a natural hypothesis can be made about such query responses, as in ((4))a); a related proposal, first articulated by (Carlson, 1983), is that they are constrained by the semantic relations of *dependence*, or its converse *influence*. A straightforward definition of these notions is in ((4)b). Its intuitive rationale is this: discussion of q_2 will necessarily bring about the provision of information about q_1:[4]

(4) a. q_2 can be used to respond to q_1 if $q_1 q-$ implies q_2. (Or q_2 influences q_1; Or q_1 depends on q_2)

b. q_2 influences q_1 iff any proposition p such that p Resolves q_2, also satisfies p entails r such that r is About q_1.

Question implication or dependence seem to constitute a sufficient condition for felicity of a (non-metacommunicative question) response. It does not seem to be a necessary condition. For instance, (3d) is felicitous but does not permit the inference

(5) Where Mummy is depends on what Chris wants for her.

[4]The definition of influence/dependence in ((4)b) makes reference to the answerhood notion of resolvedness, an agent–relative notion of exhaustiveness, as argued in (Ginzburg, 1995). Although for the moment I don't spell this out, this makes influence/dependence agent–relative rather than purely semantic notions, in contrast to aboutness. One could eliminate this asymmetry by using a purely semantic notion of exhaustiveness. This issue is further discussed below.

Consequently, a number of researchers have expressed doubt that it is dependence that underpins the requisite question/question (e.g. (Larsson, 2002; Shaheen, 2009a). Instead, with (e.g. (Asher and Lascarides, 2003)), they suggest that the requisite relation is plan–oriented, as could be articulated in terms of the rhetorical relation **Q(uery)-Elab(oration)** informally summarized in ((6)):

(6) If Q-Elab(α, β) holds between an utterance α uttered by A, where g is a goal associated by convention with utterances of the type α, and the question β uttered by B, then any answer to β must elaborate a plan to achieve g.

This latter proposal, motivated by interaction in cooperative settings, is vulnerable to examples such as ((7)):

(7) a. A: What do you like? B: What do you like?

b. A: What is Brown going to do about it? B: Well, what is Cameron?

I leave the precise characterization of this class of responses as an open issue, which requires more empirical research, both corpora-based and experimental, though for concreteness will assume an account based on q–implication/dependence.

2.3 Step 3: ability to answer

The first departure from a notion determined by questions *per se* is what one might call *metadiscursive* relevance. Irrelevance *implicatures* are an instance of metadiscursive interaction—interaction about what should or should not be discussed at a given point in a conversation:

(8) a. A: What's the problem with the drains?

b. B: I don't know.

c. B: You asked me that already.

d. B: You can't be serious.

e. B: Do we need to talk about this now?

f. B: I don't wish to discuss this now.

g. B: Whatever. Millie called yesterday.

I will mention one possible proposal concerning this aspect of relevance below. The crucial point metadiscursivity emphasizes is that a query introduces the potential for discussion of *other* questions. Specifically in this case the need to address the issue of whether a given question q should be discussed at a particular point by the responder B, an issue we might paraphrase informally as *?WishDiscuss(B, q)*.

2.4 Step 4: Genre specificity

In the case of metadiscursive relevance the issue introduced arises from the query interaction. Another source of relevance is the activity or genre type. Relevance driven by the domain plays an important role, as emphasized by a vast literature in AI, going back at least to (Cohen and Perrault, 1979; Allen and Perrault, 1980). In (9) B's perfectly relevant response is not *about* the query A asked:

(9) A: How can I help you?
B: A second class return ticket to Darlington, leaving this afternoon.

The basic intuition one can pursue is that a move can be made if it *relates to the current activity*. In some cases the activity is very clearly defined and tightly constrains what can be said. In other cases the activity is far less restrictive on what can be said:

(10) a. **Buying a train ticket**: c wants a train ticket: c needs to indicate where to, when leaving, if return, when returning, which class, s needs to indicate how much needs to be paid

b. **Buying in a boulangerie**: c needs to indicate what baked goods are desired, b needs to indicate how much needs to be paid

c. **Chatting among friends**: first: how are conversational participants and their near ones?

d. **Buying in a boulangerie from a long standing acquaintance**: combination of (b) and (d).

Trying to operationalize activity relevance presupposes that we can classify conversations into various *genres*, a term we use following (Bakhtin, 1986) to denote a particular type of interactional domain. There are at present remarkably few such taxonomies (though see (Allwood, 1999) for an informal one.) and we will not attempt to offer one here. However, as we will see below, we can indicate how to classify a conversation into a genre and build a notion of *genre-based* relevance from that.

2.5 Step 5: metacommunicative relevance

The final step for now will involve the most radical moves, ones that are ultimately difficult for many existing logical frameworks. In other words assessing which utterances are relevant as responses to an initial query—or any other type of move for that matter—requires reference to more than the query's content. This is demonstrated most clearly by metacommunicative responses, the two main types being acknowledgements of understanding and clarification requests (CRs). Here I mention a couple of salient facts that any account of metacommunicative relevance needs to address. First, CRs come in four main types, one of which relates to the phonological form of the utterance:

(11) A: Did Jo leave?

a. **intended content queries**:(*Jo?*),

b. **Repetition requests**: (*What?*),

c. **Relevance clarifications**: (*What do you mean?*),

d. **Requests for underlying motivation**: (*Why?*).

Second, there exist syntactic and phonological parallelism conditions on certain CR interpretations:

(12) a. A: Did Bo leave? B: Max? (cannot mean: intended content reading: **Who are you referring to?** or **Who do you mean?**)

b. A: Did he adore the book. B: adore? / #adored?

3 Relevance in KoS

As the underlying logical framework, I use Type Theory with Records (TTR) (Cooper, 2005), a model–theoretic descendant of Martin-Löf Type Theory (Ranta, 1994). What is crucial for current purposes about this formalism, which takes situation semantics as one of its inspirations, is that it provides access to both types and tokens at the object level. Concretely, this enables simultaneous reference to both utterances and utterance types, a key desideratum for modelling metacommunicative interaction. This distinguishes TTR from (standard) Discourse Representation Theory,[5] for instance, where the witnesses are at a model theoretic level, distinct from the level of discourse representations. The provision of entities at both levels of tokens and types allows one to combine aspects of the typed feature structures world and the set theoretic world, enabling its use as a computational grammatical formalism. The formalism can, consequently, be used to build a semantic ontology, and to write conversational interaction and grammar rules.

[5]There are versions of DRT that do allow for the presence of witnesses in the logical representation, e.g. Compositional DRT (Muskens, 1996), employed to underpin the PTT dialogue framework (Poesio and Rieser, 2010).

3.1 Information States

On the view developed in KoS, there is actually no single context, for reasons connected primarily with the integration of metacommunicative and illocutionary interaction, which I will touch on in section 3.5. Instead of a single context, analysis is formulated at a level of information states, one per conversational participant. The type of such information states is given in (13a). I leave the structure of the private part unanalyzed here, for details on this, see (Larsson, 2002). The dialogue gameboard represents information that arises from publicized interactions. Its structure is given in the type specified in (13b)——the *spkr,addr* fields allow one to track turn ownership, *Facts* represents conversationally shared assumptions, *Pending* and *Moves* represent respectively moves that are in the process of/have been grounded, *QUD* tracks the questions currently under discussion:

(13) a. TotalInformationState (TIS):

$$\begin{bmatrix} \text{dialoguegameboard} : \text{DGB} \\ \text{private} : \text{Private} \end{bmatrix}$$

b. DGB =

$$\begin{bmatrix} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{c-utt} : \text{addressing(spkr,addr)} \\ \text{Facts} : \text{set(Proposition)} \\ \text{Pending} : \text{list(locutionary Proposition)} \\ \text{Moves} : \text{list(locutionary Proposition)} \\ \text{QUD} : \text{poset(Question)} \end{bmatrix}$$

Context change is specified in terms of *conversational rules*, rules that specify the *effects* applicable to a DGB that satisfies certain *preconditions*. This allows both illocutionary effects to be modelled (preconditions for and effects of greeting, querying, assertion, parting etc), interleaved with *locutionary effects*. How querying works in this framework I will illustrate in the next section, once we have discussed *q(uestion)–specificity*.

3.2 Questions in context

The basic notion of relevance that has emerged so far can be summarized in term of the notion of q-specificity in ((14)):

(14) q-specific utterance: an utterance whose content is either a proposition p About q or a question q_1 on which q Depends

This can be embedded in 2-person interaction via a protocol as in ((15)):

(15)

querying	assertion
LatestMove = Ask(A,q)	LatestMove = Assert(A,p)
A: push q onto QUD; release turn;	A: push p? onto QUD; release turn
B: push q onto QUD; take turn; make q—specific utterance take turn.	B: push p? onto QUD; take turn; Option 1: Discuss p? Option 2: Accept p
	LatestMove = Accept(B,p)
	B: increment FACTS with p; pop p? from QUD;
	A: increment FACTS with p; pop p? from QUD;

As argued in (Ginzburg, 2011), the only query specific aspect of the query protocol in (15) is the need to increment QUD with q as a consequence of q being posed:

(16) Ask QUD–incrementation:

$$\begin{bmatrix} \text{pre} : \begin{bmatrix} \text{q} : \text{Question} \\ \text{LatestMove} = \text{Ask(spkr,addr,q):IllocProp} \end{bmatrix} \\ \text{effects} : \left[\text{qud} = [\text{q,pre.qud}] : \text{list(Question)}\right] \end{bmatrix}$$

The specification `make q-specific utterance` is an instance of a general constraint that characterizes the contextual background of reactive queries and assertions. This specification can be formulated as in ((17)): the rule states that if q is QUD–maximal, then either participant may make a q–specific move. Whereas the preconditions simply state that q is QUD–maximal, the preconditions underspecify who has the turn and require that the latest move—the first element on the MOVES list—stand in the *Qspecific* relation to q:[6]

(17) QSpec

$$\begin{bmatrix} \text{preconds} : \left[\text{qud} = \langle \text{q, Q} \rangle : \text{poset(Question)}\right] \\ \text{effects} : \begin{bmatrix} \text{spkr} : \text{Ind} \\ \text{c1} : \text{spkr} = \text{preconds.spkr} \vee \text{preconds.addr} \\ \text{addr} : \text{Ind} \\ \text{c2: member(addr,}\{\text{preconds.spkr,preconds.addr}\}\text{)} \\ \wedge\ \text{addr} \neq \text{spkr} \\ \text{r} : \text{AbSemObj} \\ \text{R: IllocRel} \\ \text{Moves} = \langle \text{R(spkr,addr,r)} \rangle \bigoplus m : \text{list(IllocProp)} \\ \text{c1} : \text{Qspecific(r,preconds.qud.q)} \end{bmatrix} \end{bmatrix}$$

The notion of q–specificity still needs some refinements if it is to do its job of regulating responses that address a given question. The most direct refinement concerns *indirect answerhood*:

[6] This underspecification of turn ownership is the basis for a unified account of question posing in monologue, 2-person querying, and multilogue provided in (Ginzburg, 2011).

responses that provide an answer indirectly should clearly be accommodated (Asher and Lascarides, 1998). This means relativizing aboutness by an entailment notion based on common ground information represented in FACTS.

3.3 Metadiscursive Relevance

A natural way to analyze such utterances is along the lines of the conversational rule QSPEC discussed in section 3.2: A introducing q gives B the right to follow up with an utterance about an issue we could paraphrase informally as *?WishDiscuss(q)*. Such a CCUR is sketched in ((18)):

(18) Discussing u?

$$\begin{bmatrix} \text{preconds} & : \text{DGB} \\ \text{effects} & : \begin{bmatrix} \text{spkr = preconds.addr : Ind} \\ \text{addr = preconds.spkr : Ind} \\ \text{r : AbSemObj} \\ \text{R: IllocRel} \\ \text{Moves} = \langle \text{R(spkr,addr,r)} \rangle \\ \bigoplus \text{pre.Moves : list(IllocProp)} \\ \text{c1 : Qspecific(R(spkr,addr,r),} \\ \text{?WishDiscuss(pre.maxqud)} \\ \text{qud} = \left\langle \begin{matrix} \text{?WishDiscuss(pre.maxqud),} \\ \text{pre.qud} \end{matrix} \right\rangle \\ \text{: poset(Question)} \end{bmatrix} \end{bmatrix}$$

3.4 Genre-based Relevance

An account of genre-based relevance presupposes a means of classifying a conversation into a genre.[7] One way of so doing is by providing the description of an information state of a conversational participant who has *successfully* completed such a conversation. Final states of a conversation will then be records of type T for T a subtype of DGB_{fin}, here Questions No (longer) Under Discussion (QNUD) denotes a list of issues characteristic of the genre which will have been resolved in interaction:

(19) $\text{DGB}_{fin} = \begin{bmatrix} \text{Facts : Prop} \\ \text{QNUD = list : list(question)} \\ \text{Moves : list(IllocProp)} \end{bmatrix}$

In ((20)) we exemplify two genres, informally specified in (10):

(20) a. CasualChat:

$$\begin{bmatrix} \text{A, B : Ind} \\ \text{t: TimeInterval} \\ \text{c1 : Speak(A,t)} \vee \text{Speak(B,t)} \\ \text{facts : Set(Prop)} \\ \text{qnud : list(question)} \\ \text{c2: } \{\lambda P.P(A), \lambda P.P(B)\} \subset \text{qnud} \\ \text{moves : list(IllocProp)} \end{bmatrix}$$

b. BakeryChat:

$$\begin{bmatrix} \text{A, B : Ind} \\ \text{t: TimeInterval} \\ \text{c1 : Speak(A,t)} \vee \text{Speak(B,t)} \\ \text{facts : Set(Prop)} \\ \text{qnud : list(question)} \\ \text{c2: } \left\{ \begin{matrix} \lambda P.P(A), \lambda P.P(B), \\ \lambda x.\text{InShopBuy(A,x)}, \\ \lambda x.\text{Pay(A,x)} \end{matrix} \right\} \subset \text{qnud} \\ \text{moves : list(IllocProp)} \end{bmatrix}$$

We can then offer the following definition of *activity relevance*: one can make a move m0 if one believes that that the current conversation updated with m0 is of a certain genre G0. Making move $m0$ given what has happened so far (represented in $dgb0$) can be *anticipated* to conclude as final state $dgb1$ which is a conversation of type G0:

(21) m0 is relevant to G0 in dgb0 for A iff A believes that there exists dgb1 such that $(dgb0 \bigoplus m0) \sqsubset dgb1$, and such that dgb1 : G0

3.5 Metacommunicative Relevance

In the immediate aftermath of a speech event u, **Pending** gets updated with a record of the form $\begin{bmatrix} \text{sit = u} \\ \text{sit-type = T}_u \end{bmatrix}$ (of type *locutionary proposition* (LocProp)). Here T_u is a grammatical type for classifying u that emerges during the process of parsing u. The relationship between u and T_u—describable in terms of the proposition $p_u = \begin{bmatrix} \text{sit = u} \\ \text{sit-type = T}_u \end{bmatrix}$— can be utilized in providing an analysis of grounding/CRification conditions:

(22) a. Grounding: p_u is true: the utterance type fully classifies the utterance token.

b. CRification: p_u is false, either because T_u is weak (e.g. incomplete word recognition) or because u is incompletely specified (e.g. incomplete contextual resolution).

[7] For an application of genre-based relevance to the semantics of *Why*-questions, see (Shaheen, 2009b).

In case p_u is true, p_u becomes the LatestMove and relevance possibilities discussed above come into operation. Otherwise clarification interaction ensues. This involves accommodation of questions into context by means of a particular class of conversational rules—Clarification Context Update Rules (CCURs), whose general substance is paraphrased in ((23)a), with a particular instance given in ((23)b):

(23) a. CCUR$_i$: given u1 a constituent of MaxPending, accommodate as MaxQUD q$_i(u1)$, follow this up with an utterance which is *co-propositional* with q$_i(u1)$.

b. Parameter identification: Input:
$$\begin{bmatrix} \text{Spkr : Ind} \\ \text{MaxPending : LocProp} \\ \text{u1} \in \text{MaxPending.sit.constits} \end{bmatrix}$$
Output:
$$\begin{bmatrix} \text{MaxQUD = What did spkr mean by u1?} \\ \text{LatestMove : LocProp} \\ \text{c1: CoProp(LatestMove.cont,MaxQUD)} \end{bmatrix}$$

CoPropositionality for two questions means that, modulo their domain, the questions involve similar answers. For instance 'Whether Bo left', 'Who left', and 'Which student left' (assuming Bo is a student.) are all co-propositional:

(24) a. Two utterances u_0 and u_1 are *co-propositional* iff the questions q_0 and q_1 they contribute to QUD are co-propositional.

(i) qud-contrib(m0.cont) is m0.cont if m0.cont : Question

(ii) qud-contrib(m0.cont) is ?m0.cont if m0.cont : Prop[8]

b. q_0 and q_1 are co-propositional if there exists a record r such that $q_0(r) = q_1(r)$.

In the current context co-propositionality amounts to: either a CR which differs from MaxQud at most in terms of its domain, or a correction—a proposition that instantiates MaxQud.

4 Combining Relevance

What then does Relevance amount to? Pretheoretically, Relevance relates an utterance u to an information state I just in case there is a way to successfully update I with u. Let us restrict attention for now to the case where the input context is a query. Given a set of conversational rules $\mathcal{C}$, a grammar $\mathcal{G}$ and an information state $I_0 : TIS$, an utterance u is $\mathbf{U}(\mathbf{tterance})_{\mathcal{C},\mathcal{G}}{}^{I_0}$**-relevant** iff there exist $c_1, \ldots, c_{k+1} \in \mathcal{C}, T_u \in \mathcal{G}, k \geq 0$ such that $c_1(I_0) = I_1, \ldots, c_{k+1}(I_k) = I_{k+1}$, where C's information state I_0 satisfies ((25)a); where by means of a sequence of updates the locutionary proposition $p_u = prop(u, T_u)$ becomes the value of LatestMove (condition ((25)b); and the final element of the sequence of updates I_{k+1} is such that one of the conditions in ((25)c-f) is satisfied—u is either q–specific, an appropriate CR, relates to the issue of willingness to discuss q, or is genre–relevant:

(25) a. I_0.DGB.LatestMove = v; v.content = Ask(A,q),

b. $I_{k+1}.DGB.LatestMove = p_u$

c. $p_u.content$ is q–specific relative to I.DGB, Or

d. $p_u.content$ is CoPropositional with some question q0 that satisfies q0 = CCUR1.effects. maxqud($I_0.DGB.MaxPending$) for some Clarification Context Update Rule CCUR1, Or

e. $p_u.content$ is q_0–specific, where q_0 is the question ?WishDiscuss(B,q), Or

f. One of C's beliefs in I_0 is that: for some G0 there exists dgb1 such that $(I_0.DGB \bigoplus p_u) \sqsubset dgb1$, and such that dgb1 : G0

A number of remarks can be made about (25), primarily about the relata of this notion.

- The definition is relative to both the set of conversational rules and to a grammar from which the types T_u from which locutionary propositions originate.
- Relevance is, by and large, DGB oriented. Only ((25)f) explicitly involves reference to the entire information state.

5 Using Relevance

In this section I offer one application of the internalized notion of relevance, formulating a rule underwriting lack of wish to address an utterance.[9] A prototypical example in this respect is given in ((26)a). Two further examples from literary texts convey a similar import:

(26) A: Horrible talk by Rozzo. B: It's very hot in here. **Implicates**: B does not wish to discuss A's utterance.

[8] Recall from the assertion protocol that asserting p introduces p? into QUD.

[9] In seeking to underwrite this inference via conversational rule there is no inconsistency with a Gricean view that such an implicature can be explicated in terms of calculations made by rational agents on the basis of apparent violations of the Cooperative Principle etc. This rule represents a "short circuited" version of the Gricean account.

a. Rumpole: Do you think Prof Clayton killed your husband? Mercy Charles: Do you think you'll get him off? ('Rumpole and the Right to Silence', p. 100)

b. Harry: Is that you James? Stella: What? No, it isn't. Who is it? Harry: Where's James? Stella: He's out. Harry: Out? Oh, well, all right. I'll be straight round. Stella: What are you talking about? Who are you? (Pinter, *The Collection*, p. 133)

In current terms we could formulate the inference as in ((27)):

(27) $\neg$Relevant(u,I) $\mapsto$
A does not wish to address I.dgb.LatestMove.

More formally, we can offer the update rule in ((28))—given that MaxPending is *irrelevant* to the DGB, one can make MaxPending into LatestMove while updating Facts with the fact that the speaker of MaxPending does not wish to discuss MAX-QUD:

(28)
$$\left[\begin{array}{l}\text{preconds: } \left[\begin{array}{l}\text{I : TIS}\\ \text{c: } \neg\text{Relevant(maxpending,I)}\end{array}\right]\\ \text{effects : } \left[\begin{array}{l}\text{LatestMove = pre.pending : LocProp}\\ \text{Facts = pre.Facts} \cup\\ \left\{\neg\ \text{WishDiscuss(pre.spkr,pre.maxqud)}\right\}.\end{array}\right]\end{array}\right]$$

Note that this does not make the *unwillingness to discuss* be the *content* of the offending utterance; it is merely an inference. Still this inference will allow MAX-QUD to be downdated from the DGB via the general mechanisms that regulate QUD downdate in conjunction with FACTS update.

6 Conclusions

Relevance is the most fundamental notion for research on dialogue. Restricting attention here to the case where a query has just taken place, I have elucidated four dimensions of this notion: responses that are question–specific, metadiscursive, genre–specific, and metacommunicative. I have formalized this notion, starting with the intuition that it amounts to a relation between an utterance and an information state where the utterance can successfully update the information state whose most recent move is a query. using the dialogue theory KoS and the formalism of Type Theory with Records. The notion of relevance that emerges is primarily one grounded in publicized contextual information, though it has some important unpublicized components, primarily those relating to genre–dependent knowledge. I have motivated the need for a notion of relevance internalized in the theory of meaning via its application in a class of clarification requests ('What do you mean') and the celebrated Gricean implicatures of lack of desire to address an utterance.

KoS enables us to construct a potentially rich theory of relevance. But as I have made clear there is a slew of issues we are in the dark about. These include:

1. **Empirical coverage**: what aspects does the four cornered characterization offered above intrinsically miss?

2. **The nature of q-responsiveness**: is there a clean way, analogous to answerhood, to characterize the (non-metacommunicative) questions arising from a given query?

3. **Relevancial success**: (Why) are there in practice few relevance CRs?

Acknowledgments

I would like to thank Jon Shaheen for a series of illuminating comments on an extended version of this paper and Robin Cooper and Alex Lascarides for discussion of topics discussed therein.

References

James Allen and Ray Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15:143–178.

Jens Allwood. 1999. The swedish spoken language corpus at göteborg university. In *Proceedings of Fonetik 99*, volume 81 of *Gothenburg Papers in Theoretical Linguistics*.

Nicholas Asher and Alex Lascarides. 1998. Questions in dialogue. *Linguistics and Philosophy*, 21.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

M.M. Bakhtin. 1986. *Speech Genres and Other Late Essays*. University of Texas Press.

Lauri Carlson. 1983. *Dialogue Games*. Synthese Language Library. D. Reidel, Dordrecht.

Philip Cohen and Ray Perrault. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:177–212.

Robin Cooper. 2005. Austinian truth in martin-löf type theory. *Research on Language and Computation*, pages 333–362.

Raquel Fernández. 2006. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King's College, London.

Jonathan Ginzburg and Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*, 27(3):297–366.

Jonathan Ginzburg and Raquel Fernández. 2010. Dialogue. In Alex Clark, Chris Fox, and Shalom Lappin, editors, *Handbook of Computational Linguistics and Natural Language*, Oxford. Blackwell.

Jonathan Ginzburg and Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.

Jonathan Ginzburg. 1994. An update semantics for dialogue. In H. Bunt, editor, *Proceedings of the 1st International Workshop on Computational Semantics*. ITK, Tilburg University, Tilburg.

Jonathan Ginzburg. 1995. Resolving questions, i. *Linguistics and Philosophy*, 18:459–527.

Jonathan Ginzburg. 2011. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford. Draft available from http://www.dcs.kcl.ac.uk/staff/ginzburg/tis1.pdf.

J. Groenendijk and F. Roelofsen. 2009. Inquisitive semantics and pragmatics. In *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric. www. illc. uva. nl/inquisitive-semantics*.

Jeroen Groenendijk. 2006. The logic of interrogation. In Maria Aloni, Alistair Butler, and Paul Dekker, editors, *Questions in Dynamic Semantics*, volume 17 of *Current Research in the Semantics/Pragmatics Interface*, pages 43–62. Elsevier, Amsterdam. An earlier version appeared in 1999 in the Proceedings of SALT 9 under the title 'The Logic of Interrogation. Classical version'.

Staffan Larsson. 2002. *Issue based Dialogue Management*. Ph.D. thesis, Gothenburg University.

R. Muskens. 1996. Combining Montague semantics and discourse representation. *Linguistics and Philosophy*, 19(2):143–186.

Massimo Poesio and Hannes Rieser. 2010. (prolegomena to a theory of) completions, continuations, and coordination in dialogue. *Dialogue and Discourse*, 1.

M. Purver. 2006. Clarie: Handling clarification requests in a dialogue system. *Research on Language & Computation*, 4(2):259–288.

Aarne Ranta. 1994. *Type Theoretical Grammar*. Oxford University Press, Oxford.

Jon Shaheen. 2009a. Comments on Ginzburg's paper, i: Intuitions about query responses and hesitation marking. Presented at the Michigan Fall Symposium on Linguistics and Philosophy, University of Michigan.

Jon Shaheen. 2009b. Genre-based relevance and why-questions. University of Amsterdam unpublished Ms.

A.M. Turing. 1950. Computing Machinery and Intelligence. *Mind*, 59(236):433–460.

Johan van Benthem and Stefan Minica. 2009. Toward a dynamic logic of questions. In Xiangdong He, John Horty, and Eric Pacuit, editors, *Proceedings of Logic, Rationality and Interaction (LORI-II)*.

Andrzej Wiśniewski. 2001. Questions and inferences. *Logique et Analyse*, 173:5–43.

Andrzej Wiśniewski. 2003. Erotetic search scenarios. *Synthese*, 134:389–427.

Radical Inquisitive Semantics

Jeroen Groenendijk & Floris Roelofsen
ILLC/University of Amsterdam

Abstract

In inquisitive semantics the proposition expressed by a sentence is viewed as a proposal to update the common ground. Such a proposal may be inquisitive, in the sense that it may offer the addressee a choice between several alternative updates. Propositions are modeled as sets of possibilities, where possibilities are sets of possible worlds.

Radical inquisitive semantics enriches this notion of meaning by pairing the proposition expressed by a sentence with a counter-proposition. The possibilities in this counter-proposition embody the ways in which the addressee may choose to counter the proposal expressed by the sentence.

Whereas the proposition expressed by a sentence gives directions for positive responses that accepct the given proposal, the counter-proposition for a sentence gives directions for negative responses that reject the proposal.

We will present a radical inquisitive semantics for the language of propositional logic, and illustrate its workings with a number of examples showing how it accounts for positive and negative responses to sentences of natural language.

Much of our attention will be devoted to the treatment of conditional sentences, both indicative conditionals and conditional questions. A remarkable result is that 'denial of the antecedent' responses to a conditional are characterized as negative responses to the 'question behind' that conditional.

References:
Jeroen Groenendijk & Floris Roelofsen, Radical Inquisitive Semantics, *http://sites.google.com/site/inquisitivesemantics/papers-1/in-progress*

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Argumentation in Artificial Intelligence

Henry Prakken
Utrecht University
The University of Groningen

Abstract

Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. This provides a natural mechanism, for example, to handle inconsistent and uncertain information and to resolve conflicts of opinion between intelligent agents. In consequence, argumentation has become a key topic in the logical study of commonsense reasoning and in the dialogical study of inter-agent communication. In this talk an overview will be given of current research in AI on argumentation, with special attention for the dialogical aspects of argumentation.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Clarification Requests as Enthymeme Elicitors

Ellen Breitholtz
University of Gothenburg
Gothenburg, Sweden
ellen@ling.gu.se

Abstract

In this paper, we aim to establish a relation between enthymematic arguments and clarification requests. We illustrate our discussion with examples where the clarification following a clarification request, together with the problematic utterance, make up an enthymeme. We also suggest possible analyses of how conversational participants, in order to work out an enthymeme, draw on topoi - notions or inference patterns that constitute a rhetorical resource for an agent engaging in dialogue.

1 Introduction

Enthymemes, semi-logical arguments drawing on "common knowledge", have evoked interest among scholars within different fields: computer science (Hunter, 2009), and philosophy (Burnyeat, 1996) on the one hand, composition and cultural studies on the other (Rosengren, 2008). Despite this, the enthymeme has not been studied to a great extent as a linguistic phenomenon. However, there is at least one study that elucidates enthymemes as conversational phenomena - Jackson and Jacobs (1980).

Jackson and Jacobs, whose work is in the CA tradition, claim that the enthymeme is linked to disagreements and objections raised in conversation, and therefore is best understood in terms of dialogue rather than monologue. We agree with this, but would like to suggest that the role of the enthymeme is more fundamental. Consider Walker's (1996) example of an interaction between two colleagues on their way to work:

(1)

i A: Let's walk along Walnut Street.
ii A: It's shorter.

Breitholtz and Villing (2008) suggested that the presence of (1)ii despite its informational redundancy (assuming both dialogue participants know that it is shorter to walk along Walnut Street), could be explained in rhetorical terms. The informational content lies in that it refers to an enthymeme according to which, if suitable topoi are employed, Walnut Street being shorter is a good reason for choosing that way to work.

This indicates that enthymemes may play a role in other contexts than just disagreement, for example contexts where an utterance needs to be elaborated, explained, motivated, or in other ways supported in order for grounding to occur. In many dialogue situations, however, reference to an enthymeme is not given spontaneously - attention is called to the need for more information by the posing of a clarification request.

In this paper we will look at the relation between enthymemes and clarification requests, more specifically how problematic utterances and clarifications can be analysed as enthymemes. We will first give some background information about Aristotle's notion of enthymeme, then look at a few examples of dialogues where some sort of communication problem is signalled by a clarification request that elicits reference to an enthymeme.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

2 Enthymemes and topoi

An enthymeme can be described as a rhetorical argument rule similar to an inference rule in logic. In the *Rhetoric* (Kennedy, 2007), Aristotle claims that learned, scientific argumentation differs from argumentation concerning every day matters. In rhetorical discourse, it is inefficient to present chains of logical arguments. Aristotle therefore recommends shortening the arguments, which results in them not being strictly logical. However, Aristotle still emphasises the logos-based, deductive nature of the enthymeme, and calls it a sort of syllogism (Kennedy, 2007).

Some enthymemes can be made into a logical arguments by adding one or more premises, which may be supplied from an agent's knowledge of culture, situation and co-text (what has been said earlier in the discourse), according to argument schemes known as the *topoi* of the enthymeme. These patterns can be very general assumptions based on physical parameters such as volume (if x is smaller than y, x can be contained in y), or more specific assumptions like *the sky is blue*, *dogs bark*, etc.

2.1 Topoi as a resource in dialogue

In his work on *doxology*, a theory of knowledge concerned with what is held to be true rather than what is objectively true, Rosengren (2008) employs rhetorical concepts to describe how common-sense knowledge and reasoning are organised. To know a society, claims Rosengren, is to know its topoi. In a micro-perspective, we could say that an important part of being able to handle a specific dialogue situation is to know relevant topoi. Thus an agent involved in dialogue has at his or her disposal a set of topoi, some of which pertain to the domain, some to the topic discussed and a great number of others that the agent has accumulated through experience. This collection of topoi could be regarded as a rhetorical resource, parallel to the way grammatical and lexical competence may be described as resources available to an agent, as envisaged by Cooper and Ranta (2008), Larsson and Cooper (2009) and Cooper and Larsson (2009).

3 Clarification Requests

Jackson and Jacobs (1980) argue that enthymematic arguments result from disagreement in a system built to prefer agreement. This suggests that the enthymemes we use in conversation are often evoked by some kind of objection, as in Jackson and Jacob's example in (2):

(2)

J: Let's get that one.
A: No. I don't like that one. Let's go somewhere else.
J: Shower curtains are all the same.

Jackson and Jacobs convincingly show that the discourse of disagreement is indeed associated with use of enthymemes. It seems to us, however, that enthymemes are not just used in order to work out disagreements. They should be just as important in situations where a conversational participant does not understand what another conversational participant is saying or why and how his/her utterance is relevant. The type of utterance that would be used in this type of situation is a *clarification request*. Ginzburg (2009) defines the posing of clarification requests (CR:s) as the engaging in "discussion of a partially comprehended utterance". According to a corpus study by Purver (2004), a little less than half of CRs have the function of questioning the semantic contribution of a particular constituent within the entire clausal content (Ginzburg, 2009). This function is referred to by Ginzburg as Clausal confirmation. Ginzburg (2009) gives an example of this type of CR, repeated here in (3). The meaning of the reprise fragment is to clarify if the rendezvous should really be in the drama studio, indicating that it is not an obvious place to meet and that the suggestion of meeting there requires an explanation.

(3)

Unknown: Will you meet me in the drama studio?
Caroline: Drama studio?
Unknown: Yes, I've got an audition.

(Ginzburg, 2009), p 146

The function performed by the clarification in (3), seems to us similar to that of "it's shorter" in (1), namely to validate the proposition made in an earlier utterance in terms of its relevance in the dialogue situation. We would like to argue that the clarification is validating precisely because it gives reference to a specific rhetorical argument, an enthymeme consisting of the utterance that provokes the CR, and the clarification.

3.1 Examples of Enthymematic Clarification

In this section we will consider two examples where references to enthymemes are made explicit by CRs. The examples are extracted from the British National Corpus using SCoRE (Purver, 2001). First, let us consider (4), where a child is being questioned about a character in a narrative:

(4)

i A: Brave
ii B: Brave?
iii B: You thought she was brave?
iv B: Why was she brave?
v A: She went into the woods.

BNC, File D97, Line 518-522

In (4) we have an example of a clausal confirmation CR - (4)ii does not serve to find out why the character was brave in the first place (e. g. because she was born brave) but to elicit a motivation to why A said the character was brave.

A topos that would make sense of the argument would be one concerning danger/courage, for example:

(5)
$$\frac{x \text{ does } A \qquad A \text{ is dangerous}}{\therefore x \text{ is brave}}$$

Our second example works somewhat differently:

(6)

i A: Does the group have an office?
ii B: No.
iii C: We've got our plastic box!
iv A: Plastic?
v C: I know I know everybody will be disappointed but I couldn't get cardboard ones.

BNC, File F72, Line 283-287

First, the clarification (6)v) elicited by the reprise fragment, points to two different arguments. Let us first consider the second half of (6) v, "I couldn't get cardboard ones". The argument is that C could not get cardboard boxes, and therefore got plastic boxes. An important point to make here is that there are many possible topoi that could be used to reach a certain conclusion. Also, it is not the case that one particular topos makes sense in every possible argument - even within a limited domain. Instead, the topoi should be perceived as a resource from which an agent can choose and combine topoi according to the situation. A set of topoi that could be drawn on to resolve this enthymeme is:

(7)
$$\frac{x \text{ is made of } y \qquad y \text{ is bad}}{\therefore x \text{ is bad}}$$

(8)
$$\frac{x \text{ is made of } y \qquad y \text{ is good}}{\therefore x \text{ is good}}$$

(9)
$$\frac{x \text{ is better than } y}{\therefore \text{ choose } x!}$$

(10)
$$\frac{x \text{ is better than } y \qquad x \text{ is unavailable}}{\therefore \text{ choose } y!}$$

The topoi (8), (9) and (10) can be combined to instantiate the enthymeme

(11)
$$\frac{\text{cardboard boxes were unavailable}}{\therefore \text{ I got plastic boxes}}$$

The function of the premise "I couldn't get cardboard ones" is, as in (4) to offer an explana-

tion to, or perhaps more correctly, a justification for, the first, problematic, proposition that plastic boxes had been purchased.

The other enthymeme in (6) is different in that the first half of (6)v, that is elicited by the CR, constitutes the conclusion of the argument rather than a premise, and (6)v does not offer an explanation to (6)iii, but expresses a consequence of (6). The argument could draw on the following topoi:

(12)
$$\frac{\begin{array}{l}x \text{ is made of } y\\ y \text{ is bad}\end{array}}{\therefore x \text{ is bad}}$$

(13)
$$\frac{x \text{ is bad}}{\therefore x \text{ makes people disappointed}}$$

The enthymeme in (6) is an instantiation of the combination of (12) and (13).

(14)
$$\frac{a \text{ is made of plastic}}{\therefore a \text{ makes people disappointed}}$$

4 Conclusions

We have argued that enthymemes may have a function in enabling the interpretation of dialogue contributions in cases where the relevance, adequacy, or suitability, of an utterance proposition in a particular situation is being questioned, and that clarification requests may have the effect of eliciting explicit reference to enthymemes. To support this, we have used examples drawn from the BNC. In the examples discussed, we looked at how a set of possible topoi make up a resource from which an agent could choose and combine different topoi that could be used to work out the enthymeme.

References

Ellen Breitholtz and Jessica Villing. 2008. Can aristotelian enthymemes decrease the cognitive load of a dialogue system user? In *Proceedings of LonDial 2008, the 12th SEMDIAL workshop*, June.

M F Burnyeat, 1996. *Essays on Aristotle's Rhetoric*, chapter Enthymeme: Aristotle on the Rationality of Rhetoric, pages 88–115. University of California Press.

Robin Cooper and Staffan Larsson. 2009. Compositional and ontological semantics in learning from corrective feedback and explicit definition. In Jens Edlund, Joakim Gustafson, Anna Hjalmarsson, and Gabriel Skantze, editors, *Proceedings of DiaHolmia: 2009 Workshop on the Semantics and Pragmatics of Dialogue*, pages 59–66.

Robin Cooper and Aarne Ranta. 2008. Natural languages as collections of resources. In Robin Cooper and Ruth Kempson, editors, *Language in Flux: Dialogue Coordination, Language Variation, Change and Evolution*, Communication, Mind and Language, pages 109–120. College Publications.

Jonathan Ginzburg. 2009. *The Interactive Stance: Meaning for Conversation*.

Philippe Besnard & Anthony Hunter, 2009. *Argumentation in Artificial Intelligence*, chapter Argumentation based on classical logic, pages 133–152. Springer.

Sally Jackson and Scott Jacobs. 1980. Structure of conversational argument: Pragmatic bases for the enthymeme. *Quarterly Journal of Speech*, 66:251–265.

Kennedy. 2007. *Aristotle: On Rhetoric, a theory of civic discourse*. Oxford University Press.

Staffan Larsson and Robin Cooper. 2009. Towards a formal view of corrective feedback. In Afra Alishahi, Thierry Poibeau, and Aline Villavicencio, editors, *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*.

Matthew Purver. 2001. Score: A tool for searching the bnc. Technical Report TR-01-07, Department of Computer Science, King's College London.

Matthew Purver. 2004. *The Theory and Use of Clarification in Dialogue*. Ph.D. thesis, King's College.

Mats Rosengren. 2008. *Doxologi : en essä om kunskap*. Retorikförlaget.

Marilyn A. Walker. 1996. Limited attention and discourse structure. *Computational Linguistics*, 22(2):255–264.

D Walton. 2008. The three bases for the enthymeme: A dialogical theory. *Journal of Applied Logic*, pages 361–379.

A Computational Model for Gossip Initiation

Jenny Brusk
School of Humanities and Informatics
Skövde University, Sweden
jenny.brusk@his.se

Abstract

We are interested in creating non-player characters (NPCs) in games that are capable of engaging in gossip conversations. Gossip could for instance be used to spread news, manipulate, and create tension between characters in the game, so it can have a functional as well as a social purpose. To accomplish this we need a computational model of gossip and such a model does not yet exist. As a first step in that direction we therefore present a model for initiating gossip that calculates whether it is appropriate for the NPC to start a gossip conversation based on the following factors: The (perceived) relationship between the NPC and the player character (PC); the relationship between each of the participants and the potential target; the news value of the gossip story; and how sensitive the story is.

1 Introduction

We are interested in creating non-player characters (NPCs) with the ability to engage in socially oriented interactions. In order for this to happen, the NPCs need (among other things) social awareness and the ability engage in casual conversations, that is, conversations that are motivated by "interpersonal needs" (Eggins and Slade, 1997). One such type of conversation is gossip, broadly defined as *evaluative talk about an absent third person*. Gossip could for instance be used to spread news, manipulate, and create tension between characters in the game, so it can have a functional as well as a social purpose. For this to be possible we need a computational model of gossip and such a model does not yet exist. As a first step to accomplish this, we here propose a model for initiating gossip using Harel statecharts (Harel, 1987). The model calculates whether it is appropriate for the NPC to start a gossip conversation based on the following factors: The (perceived) relationship between the NPC and the player character (PC); the relationship between each of the participants and the potential target; the news value of the gossip story; and how sensitive the story is.

We have combined the theory of politeness (Brown and Levinson, 1987) with research on gossip structure (e.g. Eder and Enke, 1991; Eggins and Slade, 1997) applied on gossip conversations occurring in screenplays. In addition, we have used insights gained from conducting two surveys concerning the identification of gossip.

2 Background

In every social interaction the participants put a great amount of effort in face management actions, i.e., actions that serve to protect one's own and the other participants' public self-image that they want to claim for themselves (Goffman, 1967; Brown and Levinson, 1987). Gossip has been described as containing "morally contaminated information…" which can damage the initiator's reputation (Bergmann, 1993). Because of this, the initiator must make sure that the recipient is willing to gossip (Bergmann, 1993) and that the relationship is sufficiently good to minimize the threat to face.

Brown and Levinson (1987) suggest that the threat to face a certain action has in a particular situation is dependent on three socially determined variables: the social distance (*SD*) between the speaker (*S*) and the hearer (*H*); the hearer's power over the speaker (*P*); and the extent to which the act is rated an imposition in that culture (i.e., the degree to which the act interferes with an agent's wants of self-determination or of approval) (*I*): $Threat = SD(S, H) + P(H, S) + I$. They furthermore propose that the value of *SD* and *P*, respectively, is an integer between 1 and *n*, "where *n* is some small number" (p. 76).

Their description of *I* is too general to be useful for our purposes and does not take into account the participants' relationship to the gossip target, for example; a factor that we mean is essential for determining whether it is appropriate to start gossiping at all. Therefore, we start by exploring the *preconditions* for *S* (the NPC) to even consider a gossip initiation by calculating

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

the interpersonal relationship (abbreviated to ρ) between *S* and *H*: ρ = *SD(S, H) + P(H, S),* where *SD* and *P*, respectively, is an integer between 0 and 3 (thus slightly different from Brown and Levinson's suggestion). A low ρ value means that the relationship is sufficiently good for initiating a gossip conversation. In section 4 we will discuss the additional factors that need to be considered before introducing a specific gossip story.

Previous studies (e.g. Bergmann, 1993; Eder and Enke, 1991; Eggins and Slade, 1997; Hallett et al., 2009) have shown that gossip is built around two key elements: **An absent third person in focus** (henceforth referred to as *F*) and **An evaluation of *F*'s deviant behavior or of *F* as a person.** There are some reservations concerning *F*:

- *F must not be emotionally attached to S or H*, since that would make *F* "virtually" (Bergmann, 1993) or "symbolically" (Goodwin, 1980) present.
- *F is unambiguously the person in focus*. *F* must for example not play a sub-ordinate role as part of a confrontation, self-disclosure, or an insult.

In addition, explanations are commonly (or *always,* according to Eggins and Slade (1997)) used in gossip conversations to motivate the negative evaluations – they *substantiate* the gossip.

3 Harel Statecharts

The model is presented using statechart notation (Harel, 1987), which is a visual formalism for describing reactive behavior. Statecharts are really extended finite state machines that allow us to cluster and refine states by organizing them hierarchically. States can also run in parallel, independently of each other but capable of communicating through broadcast communication. It is also possible to return to a previous configuration by use of a history state. Within a statechart, data can be stored and updated using a datamodel (a.k.a. "extended state variables").

How to read the statechart: The rounded boxes represent states, and states that contain another statechart represent hierarchical states (compound states). The directed arrows denote possible transitions between the states. Labels connected to transitions represent events and/or conditions that trigger the transition. A transition can also be "empty" (ε), such that it will be taken as soon as the state's possible on-entry and on-exit scripts have been executed. An arrow starting from a black dot points to the default start state.

4 Initiating Gossip

Bree: Tisha. Tisha. Oh, I can tell by that look on your face you've got something good. Now, come on, don't be selfish.
Tisha: Well, first off, you're not friends with Maisy Gibbons, are you?
Bree: No.
Tisha: Thank god, because this is too good. Maisy was arrested. While Harold was at work, she was having sex with men in her house for money. Can you imagine?
Bree: No, I can't.
Tisha: And that's not even the best part. Word is, she had a little black book with all her clients' names.
Rex: So, uh...you think that'll get out?
Tisha: Of course. These things always do. Nancy, wait up. I can't wait to tell you this. Wait, wait.

The dialogue above is retrieved from Desperate Housewives[1] and is an example of a typical gossip dialogue. It has a third person focus (Maisy), an evaluation ("this is too good"), and a story in which Maisy's deviant behavior is central (she has been arrested for having sex with men in her house for money while her husband was at work). Notice also that before Tisha initiates the gossip she makes sure that the social distance between the target and the recipients is sufficiently high ("you're not friends with…?").

In the model we propose it is always the NPC that initiates the gossip, assuming that the information may have a gameplay value for the player.

In order to qualify as gossip, the story must have a news value (see for example Bergmann, 1993), which in our model is stored as a parameter, `NewsVal`, with a value ranging between 0 ("common knowledge") and 2 ("recently gained information"). However, if it is indifferent for the subject that the information is revealed or if the behavior is generally acceptable within that culture (e.g. within the group, community, or society) it is unlikely that it will be regarded as gossip. In order to account for this, we have added a `sensitivity` value for the propositional content of the gossip story. `Sensitivity` is here specified to be an integer between 0 and 3, where 0 indicates a generally acceptable

[1] Touchstone Television.

behavior. We assume that the value of `sensitivity` and `NewsVal` decreases over time.

We propose that the social distance (*SD*) can have one of the following values (with approximate correspondences): 0 for intimate relationships; 1 for friends; 2 for acquaintances; and 3 for strangers. The target is then selected on basis of the following factors assuming that there is an NPC (*S*) who is talking to the player character (PC) (*H*):

- *S* perceives that the risk of losing face (ρ) is low in the interaction with *H*, i.e., the social distance between *S* and *H* is (perceived to be) low and there is a (perceived) symmetric power relationship between them (ρ< 3).
- *S* has new, sensitive information about *F*.
- *S* knows *F* and believes that *H* knows, or is acquainted with, *F* too, i.e. *SD(S, F)*<3 and *SD(H, F)*<3.
- *S* does not have an intimate relationship with *F*, and believes that the same holds for *H,* i.e., *SD(S, F)* >0 and *SD(H, F)* > 0.
- *S* believes that *F* cannot hear the conversation.

The model (see figure 1) works as follows: *S* and *H* are engaged in a conversation. If ρ< 3, a transition to the state `InitiateGossip` is triggered (The source state is unspecified, but we can assume that the participants have greeted each other and perhaps small talked for a while before gossip is initiated).

S starts by searching for a potential gossip target (T) in the database (`Get(T,DB)`) according the specification presented previously, which is performed on entry of the state `SelectTarget`. The story must not be about *S* him/herself or about *H* (OP in the graph stands for Other Participants, in this case OP=*H*). If such a target exists in the database (DB), i.e., `T≠void` (and assuming that T=*F*), a transition from `SelectTarget` to `EstablishGossip` is activated. If there is no target that fulfills the initial criteria, the gossip is cancelled (never initiated).

The default start state in `EstablishGossip` is `GetGossipStory`, in which a search for a story about T=*F* is conducted. The search has two possible outcomes: there is a story about *F* that fulfills the criteria (`NewsVal`=2 and `Sensitivity` > 0), or it fails to find such a story. If a story is found, the next step is to establish *H* and *F*'s relationship. If *S* is uncertain of their relationship, a transition is taken to the state `EstablishId`, in which *S* requests a clarification that will help to establish the social distance between *H* and *F*, for instance as a question: "Do you know *F*?" or "Have you heard about *F*". If *H* responds with a request for clarification of who *F* is, then *S* can provide more information about *F,* which is handled in `ExpandId`. If *S* believes that *SD(H, F)*=0, i.e., that they are intimately related, *S* will choose to back away from the gos-

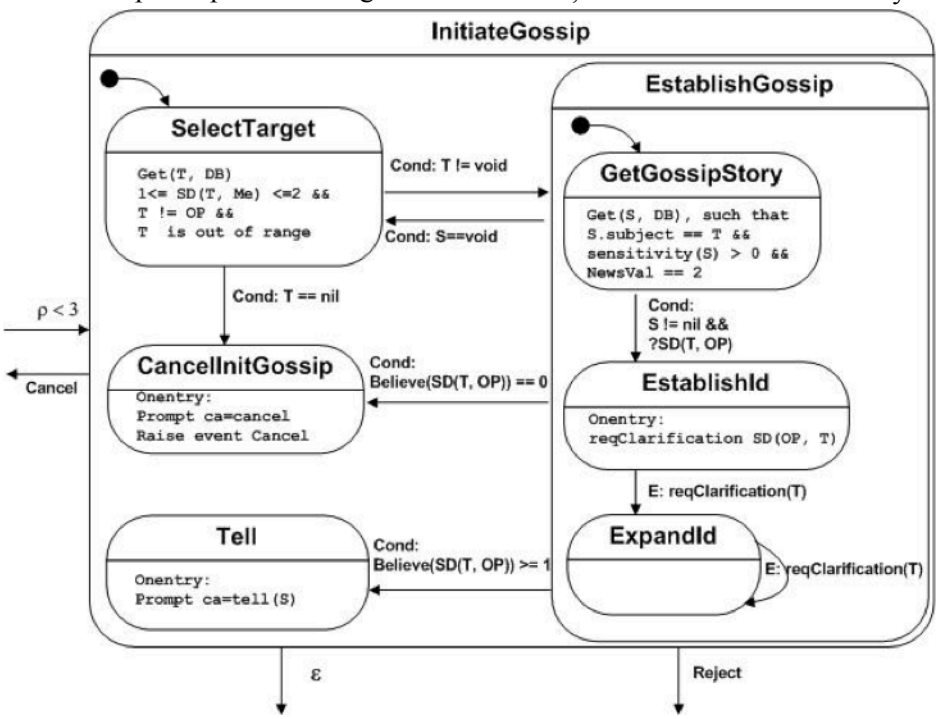


Figure 1. Model for initiating a gossip conversation.

sip and the gossip is cancelled (which corresponds to a transition to `CancelInitGossip`). Otherwise, *S* will spread the gossip (which is performed in the state `Tell`). If no story exists that fulfills the criteria, *S* will attempt to find a new target.

5 Discussion

One of the most important factors of gossip initiation is the status of the relationship between the gossipers and between them and the target. We therefore suggest that the following factors determine whether the NPC can introduce gossip at all: The (perceived) relationship between the NPC and the PC; the relationship between each of the participants and the potential target; the news value of the gossip story; and how sensitive the story is (culturally and personally). More specifically this means that the target must not be intimately related to any of the participants and that the participants must be friends or acquaintances. We have no restrictions concerning gossip between closely related participants, even if it is unclear whether it should to be considered gossip (see e.g. Bergmann, 1993). Such a restriction would be unnecessary since it just means that the risk of losing face is very low.

There are many different forms of gossip (see for example Gilmore (1978)) and many forms in which gossip can be initiated. In the model we propose here we have delimited the gossip to be sensitive news about an absent game character. The target is selected first (either by being mentioned in the previous discourse or by searching the database on entry of `SelectTarget`), but it could equally well be the story that is chosen first. There are a number of reasons why we chose the former alternative: First, even if it is the behavior that is being evaluated, it is always a person that (at least) implicitly is being judged and thereby can be damaged by the gossip. Second, the target may already be in focus or mentioned (for instance in a pre-sequence, see Bergmann (1993)), as in the following example, where the actual gossip is initiated when Jerry[2] expresses his opinion in line 3 (we have removed a sequence in which the participants try to establish the identity of the target):

1. **Jerry:** Hey, by the way, did you ever call that guy from the health club?
2. **Elaine:** Oh yeah! Jimmy.
   […]
3. **Jerry:** Can't believe your going out with him...
4. **Elaine:** Why?
5. **Jerry:** I dunno. He's so strange.
   […]

Third, if the initiator misinterprets the target's relation to the addressee(s), it is the initiator that is considered to behave inappropriately. Hence, by making a mistake in the selection of the target the initiator face the risk that the gossip gets back at him or her.

Acknowledgment
This work was partly conducted at and financed by Gotland University, GSLT, and USC/ICT, respectively. The author would in particular like to thank Dr. David Traum and the rest of the NL group at USC/ICT for all support and suggestions about my work on gossip. Thanks also to the anonymous reviewers for helpful comments on this paper. The author alone is responsible for any errors and omissions.

References

Jörg R. Bergmann. 1993. Discreet Indiscretions: The Social Organization of Gossip. Aldine, New York.

Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Suzanne Eggins and Diana Slade. 1997. *Analysing Casual Conversation*. Equinox Publishing Ltd.

Donna Eder and Janet Lynne Enke. 1991. The Structure of Gossip: Opportunities and Constraints on Collective Expression among Adolescents. *American sociological Review*, Vol. 56, No. 4 (Aug.), pp. 494–508.

David Gilmore. 1978. Varieties of Gossip in a Spanish Rural Community. Ethnology, Vol. 17, No. 1 (Jan.), pp. 89-99.

Erving Goffman. 1967. *Interaction Ritual: Essays on Face-to-face Behavior*. Pantheon Books, New York.

Tim Hallett, Brent Harget and Donna Eder. 2009. Gossip at Work: Unsanctioned Evaluative Talk in Formal School Meetings. *Journal of Contemporary Ethnography*, Vol. 38, No. 5, pp 584–618.

David Harel. 1987. Statecharts: A Visual Formalism for Complex Systems. *Science of Computer Programming*, Vol. 8, pp. 231-274.

Marjorie Harness Goodwin. 1980. He-Said-She-Said: Formal Cultural Procedures for the Construction of a Gossip Dispute Activity. *American Ethnologist*, Vol. 7, No. 4 (Nov.), pp. 674-695.

[2] From Seinfeld, Castle Rock Entertainment.

Evaluating Task Success in a Dialogue System for Indoor Navigation

Nina Dethlefs, Heriberto Cuayáhuitl, Kai-Florian Richter,
Elena Andonova, John Bateman
University of Bremen, Germany
dethlefs@uni-bremen.de

Abstract

In this paper we address the assessment of dialogue systems for indoor wayfinding. Based on the PARADISE evaluation framework we propose and evaluate several task success metrics for such a purpose. According to correlation and multiple linear regression analyses, we found that task success metrics that penalise difficulty in wayfinding are more informative of system performance than a success/failure binary task success metric.

1 Introduction

Wayfinding in (partially) unknown environments poses a considerable challenge for humans. Our work addresses indoor navigation within complex buildings that present significant navigational challenges to new and infrequent visitors. This application scenario is of increasing relevance nowadays, as building complexes become larger and greater attention is paid to making them accessible for a broader range of users. Despite this, automatic systems generating natural language-based in-advance route descriptions have received little attention to date. Previous work consists of either attempts to involve primarily visual support or the quality and effects of the language component are taken for granted (Kray et al., 2005; Callaway, 2007; Kruijff et al., 2007). In particular, there appears to be a lack of agreed-on evaluation metrics for assessing the performance of dialogue systems in the wayfinding domain. This paper addresses this lack and suggests to employ metrics that are sensitive to task difficulty for evaluation of such systems. We present a study that confirms the usefulness of our proposed metrics.

2 System Architecture

We base our study on data collected from a dialogue system that assists users in indoor wayfinding using text-based natural language input and output. The system architecture consists of four different modules for the tasks of Natural Language Understanding, dialogue management, route instruction generation and Natural Language Generation. The following is a sample dialogue translated from German (S=System, U=User):

```
S: Do you need a route description on this
    level? Just write your question into
    the text field.
U: Where can I find room number 3180?
S: I have found the following rooms: b3180,
   a3180. Which are you looking for?
U: b3180
S: Please turn around and go straight until
   the next hallway on the left-hand side.
   Turn left and go until the door b3180
   on the right-hand side.
   Do you have further questions?
```

Since the task success metrics proposed in this paper aim to be independent of specific architectural decisions, we refer the reader to Cuayáhuitl et al. (2010) for details on our indoor navigation dialogue system.

3 Experimental Setting

3.1 Evaluation methodology

Evaluation of the system was performed using objective and subjective metrics mostly derived from the PARADISE framework (Walker et al., 2000). We used the following quantitative metrics. First, the group of *dialogue efficiency* metrics includes 'system turns', 'user turns', and 'elapsed time' (in seconds). The latter includes the time used by both conversants, from the first user utterance until the last system utterance. Second, the group of *dialogue quality* metrics consists of percentages of parsed sentences, sentences with spotted keywords, and unparsed sentences. Third, the

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

group of *task success* metrics includes the well known success/failure Binary Task Success (BTS) defined as

$$\text{BTS} = \begin{cases} 1 & \text{for Finding the Target Location (FTL), with or without problems} \\ 0 & \text{otherwise.} \end{cases}$$

Because this metric does not penalise difficulty in wayfinding, we propose and evaluate the following metrics — referred to as Graded Task Success (GTS) — that penalise with different values:

$$\text{GTS}^a = \begin{cases} 1 & \text{for FTL without problems} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{GTS}^b = \begin{cases} 1 & \text{for FTL with none or small problems} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{GTS}^c = \begin{cases} 1 & \text{for FTL without problems} \\ 1/2 & \text{for FTL with small problems} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{GTS}^d = \begin{cases} 1 & \text{for FTL without problems} \\ 2/3 & \text{for FTL with small problems} \\ 1/3 & \text{for FTL with severe problems} \\ 0 & \text{otherwise.} \end{cases}$$

We coded difficulty in wayfinding, using the categories 'no problems', 'small problems' and 'severe problems' as follows. The value of 1 was given when the user finds the target location without hesitation, the value with 'small problems' was given when the user finds the location with slight confusion(s), and the value with 'severe problems' was given when the user gets lost but eventually finds the target location. The motivation behind using task success metrics that penalise differently the difficulty in wayfinding was to discover a metric that correlates highly with user satisfaction. Such a metric aims to be more informative for assessing task success performance than the traditional binary task success metrics. We tried four different graded metrics, GTS^a - GTS^d, in order to find the metric that best predicted user satisfaction. For the qualitative evaluation we used the subjective metrics described in (Walker et al., 2000).

3.2 Evaluation setup

Twenty-six native speakers of German participated in our study with an average age of 22.5 and a gender distribution of 16 female (62%) and 10 male (38%). Each subject received six dialogue tasks, corresponding to locations to find, which resulted in a total of 156 dialogues. Dialogues consisted of differing numbers of High-Level instructions (HLIs). High-Level Instructions (HLIs) encapsulate a set of low-level instructions (e.g., 'go straight', 'turn left', 'turn around') and are based on major direction changes. Two dialogue tasks used 2 High-Level Instructions (HLIs) such as those shown in the dialogue on page 1. Two other tasks used 3 HLIs, and two used 4 HLIs. The tasks were executed pseudorandomly (from a uniform distribution), so that the order of task execution would not impact on the user ratings. The participants were asked to request a route from the system using natural language, optionally take notes, and then follow the system instructions closely trying to find the locations. They were not allowed to ask anybody for help. Participants could give up when they were unable to find the target location by telling that to the assistant that followed them. It was the task of this assistant as well to judge and take note of the difficulties that subjects encountered in their wayfinding task as described in the previous section. At the end of each dialogue, participants were asked to fill in a questionnaire for obtaining qualitative results using a 5-point Likert scale, where 5 represents the highest score.

4 Experimental Results

Table 1 summarises our results for the quantitative and qualitative metrics. It can be observed from the dialogue efficiency metrics (first group) that user-machine interactions involved short dialogues in terms of turns and interaction time. Once users received instructions from the system, they tended not to ask further. With regard to dialogue quality (second group), we noted that our grammars need to be extended in coverage and that the keyword spotter proved vital in the dialogues. The analysis of task success measures (third group) revealed very high binary task success, and lower scores for the other task success metrics.

4.1 Correlation analysis

In a correlation analysis between task success measures and user satisfaction we obtained the results displayed in Table 2. This can be interpreted as follows: while all metrics correlate moderately with overall user satisfaction, the metrics taking task difficulty into account correlate higher. A more detailed analysis of the corre-

Table 1: Mean values of our evaluation metrics for our wayfinding system based on 156 dialogues, organised in four groups: dialogue efficiency, dialogue quality, task success and user satisfaction.

Measure	Score
System Turns	2.30 ± 0.3
User Turns	1.52 ± 0.5
System Words per Turn	41.30 ± 4.0
User Words per Turn	4.79 ± 2.1
Interaction Time (secs.)	22.14 ± 18.4
Session Duration (secs.)	2014.62 ± 393.2
Parsed Sentences (%)	16.7 ± 16.0
Spotted Keywords (%)	79.9 ± 17.0
Unparsed Sentences (%)	3.4 ± 0.5
Binary Task Success (%)	94.9 ± 8.3
Graded Task Success^a (%)	71.4 ± 15.0
Graded Task Success^b (%)	87.8 ± 15.0
Graded Task Success^c (%)	81.4 ± 13.3
Graded Task Success^d (%)	87.6 ± 8.3
(Q1) Easy to Understand	4.46 ± 0.8
(Q2) System Understood	4.65 ± 0.8
(Q3) Task Easy	4.29 ± 0.9
(Q4) Interaction Pace	4.63 ± 0.5
(Q5) What to Say	4.66 ± 0.7
(Q6) System Response	4.56 ± 0.6
(Q7) Expected Behaviour	4.45 ± 0.8
(Q8) Future Use	4.31 ± 0.9
Overall User Satisfaction (%)	90.0 ± 7.3

lation between task success metrics and individual user satisfaction metrics revealed the following. First, the binary task success showed lower correlations than the other metrics in the subjective metric 'easy to understand' (Q1). Second, while there is no correlation between the subjective metric 'future use' (Q8) and binary task success, the other metrics reveal a moderate correlation. Third, while binary task success shows a moderate correlation for 'task easy' (Q3), the other metrics show a high correlation. Therefore, we can conclude that the task success metrics that penalise difficulty in wayfinding are more informative of user-system interaction performance for indoor wayfinding than the BTS metric. Furthermore, there was no correlation between the number of high-level instructions and overall user satisfaction, i.e. user satisfaction was independent of instruction length (our system performed equally well for short and long routes).

Table 2: Correlation coefficients between task success and user satisfaction measures (significant at $p < 0.05$).

Measure	BTS	GTS^a	GTS^b	GTS^c	GTS^d
Q1	.47	.44	.54	.49	.54
Q2	.20	.17	.19	.19	.20
Q3	.53	.67	.71	.71	.76
Q4	.21	.26	.24	.24	.28
Q5	.20	n.s.	.17	.18	.18
Q6	n.s.	n.s.	n.s.	n.s.	n.s.
Q7	.31	.35	.44	.40	.44
Q8	n.s.	.39	.32	.40	.39
Overall	.43	.52	.55	.55	.60

Note: n.s. - not significant.

4.2 Multiple linear regression analysis

In order to identify the relative contribution that different factors have on the variance found in user satisfaction scores, we performed a standard multiple linear regression analysis on our data. According to the PARADISE framework (Walker et al., 1997), performance can be modeled as a weighted function of task-success measure and dialogue-based cost measures. The latter represent the measures summarised under dialogue efficiency and dialogue quality above. We normalised all task success and cost values to account for the fact that they can be measured on different scales (seconds, percentages, sum, etc.), according to $\mathcal{N}(x) = \frac{x-\bar{x}}{\sigma_x}$, where σ_x corresponds to the standard deviation of x. Then we performed several regression analyses involving these data.

Results revealed that the metrics 'user turns' and 'task success' (for GTS^a, GTS^c and GTS^d) were the only predictors of user satisfaction at $p < 0.05$. The other task success measures were not significant (with BTS at $p = 0.39$ and GTS^b at $p = 0.17$). These results confirm our claim that task success metrics that consider difficulty in wayfinding (specifically GTS^a, GTS^c and GTS^d) are more informative with respect to user satisfaction in the wayfinding domain than a binary success/failure metric. Subjects seem to be sensible to problems they encounter in their wayfinding tasks, which are expressed in their ratings of the system.

4.3 Estimation of a performance function

We use the following equation to obtain a performance function (Walker et al., 1997):

$$\text{Performance} = (\alpha * \mathcal{N}(k)) - \sum_{i=1}^{n} \omega_i * \mathcal{N}(c_i),$$

where, α is a weight on the task success metric k (to be replaced by any of our proposed metrics), and ω_i is a weight on the cost functions c_i. $\mathcal{N}$ represents the normalised value of c_i. Based on the results of our first regression analysis, we ran a second analysis using those variables that were significant predictors in the first regression, i.e. the number of user turns and task success metrics GTS^a, GTS^c and GTS^d. We analysed the correlation between these variables, which resulted in weak negative correlations. We obtained the following performance function for task success metrics GTS^c and GTS^d (because those two accounted for most of the variance in user satisfaction), where UT refers to 'User turns':

$$\text{Performance} = 0.38\mathcal{N}(GTS^{c,d}) - 0.87\mathcal{N}(UT),$$

suggesting that the more successful and efficient the interaction, the better. These results show that GTS^c, GTS^d and UT are significant at $p < 0.01$, and the combination of UT and each of GTS^c and GTS^d account for 62% of variance in user satisfaction. This performance function can be used in future evaluations of the system.

5 Discussion

The idea of taking different degrees of task difficulty into consideration in evaluation is not entirely new (Tullis and Albert, 2008). However, to the best of our knowledge, there have been no previous studies that demonstrated that these metrics do indeed show a higher correlation with user satisfaction scores than the BTS metric, which is typically used to assess task success. This finding was supported by an evaluation in a real environment using an end-to-end dialogue system, and was based on PARADISE, a generic framework for the evaluation of (spoken) dialogue systems. The proposed metrics can therefore be regarded as a useful and important step contributing to the understanding of the performance of situated dialogue systems. Further, our proposed metrics address the lack of standardised evaluation metrics in the wayfinding domain in particular. We presented a concrete performance function that can help future system development in the domain by allowing the estimation of relative contributions of different task success metrics and cost function towards overall user satisfaction.

6 Conclusion

In this paper we addressed the assessment of dialogue systems for indoor navigation using the PARADISE framework and different task success metrics. We found that task success metrics that take difficulty in wayfinding into account correlate higher with overall user satisfaction than a binary task success metric. In addition, a more detailed correlation analysis for subjective metrics of user satisfaction confirmed that our proposed metrics are more informative of system performance for indoor wayfinding than the binary success/failure metric. This result was confirmed by a multiple linear regression analysis that tested for the relative contribution to variance in user satisfaction of different task success metrics and cost measures. Future work can apply these metrics to dialogue systems with different input and output modalities.

Acknowledgements

Part of this work was supported by the DFG Transregional Collaborative Research Center SFB/TR8 "Spatial Cognition".

References

Charles Callaway. 2007. Non-localized, interactive multimodal direction giving. In I. van der Sluis, M. Theune, E. Reiter, and E. Krahmer, editors, *Proceedings of the Workshop on Multimodal Output Generation MOG 2007*, pages 41–50. Centre for Telematics and Information Technology (CTIT), University of Twente.

Heriberto Cuayáhuitl, Nina Dethlefs, Kai-Florian Richter, Thora Tenbrink, and John Bateman. 2010. A dialogue system for indoor wayfinding using text-based natural language. In *International Journal of Computational Linguistics and Applications, ISSN 0976-0962*.

Christian Kray, G. Kortuem, and A. Krüger. 2005. Adaptive navigation support with public displays. In Robert St. Amant, John Riedl, and Anthony Jameson, editors, *Proceedings of IUI 2005. ACM Press, New York*, pages 326–328.

Geert-Jan M. Kruijff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: What, where... and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138, March. Special Issue on Human-Robot Interaction.

Tom Tullis and Bill Albert. 2008. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1997. Paradise: A framework for evaluating spoken dialogue agents. In *ACL*, pages 271–280.

M. Walker, C. Kamm, and D. Litman. 2000. Towards developing general models of usability with PARADISE. *Natural Language Engineering*, 6(3):363–377.

Guiding the User when Searching Information on the Web

Marta Gatius **Meritxell González**
Technical University of Catalonia
Barcelona, Spain
{gatius,mgonzalez}@lsi.upc.edu

Abstract

This paper describes how we approach the problem of guiding the user when accessing informational web services. We developed a mixed-initiative dialogue system that provides access to web services in several languages. In order to facilitate the adaptation of the system to new informational web services dialogue and task management were separated and general descriptions of the several tasks involved in the communication process were incorporated.

1Introduction

This paper describes how we approach the problem of guiding the user when accessing informational web services. We designed a dialogue system (DS) for accessing different types of applications in several languages. The results of the evaluation of the first prototype are described in (Gatius and González, 2009). In order to improve both the functionality and adaptability of the DS we have studied the most appropriate representation of the general and application-specific conceptual knowledge involved when helping the user to access informational services.

When providing access to information-seeking applications DSs use an underspecified set of constraints to restrict the search rather than a defined user's goal (which can be broken down into tasks and subtasks). Hence, the main tasks for DSs providing access to an informational service consist of guiding the user to give the needed constraints as well as presenting in an appropriate way the results. There have been several approaches to face this problem (Rieser and Lemon, 2009; Steedman and Petrick, 2007; Varges et al., 2009). Our approach consists of separating completely dialogue management from task management (following other relevant proposals (Allen et al., 2001)), and defining the general tasks involved when accessing informational services. Besides, general mechanisms using the two main knowledge bases of the system (the dialogue context and the domain conceptual knowledge) are used to relax the query constraints and to state additional constraints.

2 Dialogue and Task Management

The DS we developed consists of five independent modules: the language understanding, the dialogue manager (DM), the task manager, the language generator and the user model, used to adapt automatically the dialogue strategies. Additionally, there are two main data structures accessible for all modules: the information state, representing the dialogue context and the conceptual knowledge, describing the application domain.

The DM follows the information state update model, which provides a complete separation of dialogue and task management. The DM uses communication plans to determine the next system actions that could satisfy user's requirements. These plans are generated (semi)-automatically when a new service is incorporated into the DS by adapting the general communication plan for the service type to the particular service specifications.

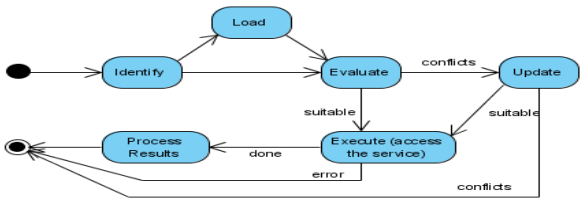


Figure 1: Task Management in the Dialogue System

Figure 1 shows task management in the DS. Main tasks performed by the task manager are the following: identification of the required web service and the specific service task, completion of the data obtained from the user, access to the service and presentation of the results.

Once the communication starts and the first intervention of the user has been interpreted and passed to the task manager, it has to identify the service and the specific service task that has to be accessed. Then, an instantiation of the specific task is generated. There are several general task descriptions for each service type, for the informational services two tasks are considered: find a list of items and describing an item.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Finally, the task manager accesses the web service and decides the most appropriate presentation of the results obtained.

3 Accessing InformationalServices

The tasks involved when the system guides the user to access an informational service are shown in Figure 2. Circled blocks represent the specific information the DM has to obtain from the user: the searched data (requestedData or output parameters) and the data constraining the query (queryConstraints or input parameters). Rectangles represent the three different tasks processing the resulting data: describing a particular item, collecting a list of items and summarizing the results obtained. Colored blocks correspond to the three different processes considered when updating constraints: relaxation, using default values and adding new constraints.

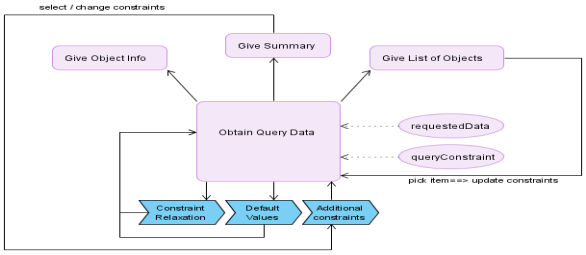


Figure 2: The tasks involved in information-seeking

The process of obtaining the query constraints from the user could be complex, as they are not gathered in a predetermined order. The task manager determines whether a complete query can be generated or if additional information has to be obtained from the user. If the service's definition includes default values, they can be included to complete the query. Parameter values appearing in previous turns can also be used.

The information obtained from the service has to be processed. Four different situations are distinguished: the result is only one item, the number of items obtained belongs to a predefined range, there are too many results and there are no results. In case there is only one item a detailed description of this item is given. In case the number of results is acceptable, a list enumerating all of them is presented to the user, suggesting him to pick up one. In the two latter cases the constraints have to be updated.

In the specific case that there are no results, the task manager can automatically relax the constraints and execute the query again. The constraints can be relaxed at the level of the query and at the level of the parameter's values. In the former, the system removes one or more of the query constraints. In the latter, the system updates the value for one or more of the constraints. The conceptual knowledge base is used to relax the constraints. If taxonomies describing the domain have been incorporated, a class is substituted by the upper class (for example, if the user asks for *drama movies* and there are none, the upper class *movies* would be used). Several strategies for data common to several applications (such as dates and locations) are already considered.

In the specific case that too many items are obtained from the service, the system presents a summary of the results. Information suggesting possible additional constraint values could also be given to the use.

5 Conclusions and Future Work

In order to improve the functionality and adaptability of our DS when guiding the user to accessing informational service we have studied the general and the application specific conceptual knowledge involved in the communication process. In our system this knowledge has been represented as a general scheme from which the communication plans for each informational service are generated and general task that are instantiated for each service. The resulting architecture facilitates the integration of other application types into the system since the task models can be easily extended and adapted.

Future work could include the processing of user's questions which answer involves the processing of data obtained from several web services.

References

J.F. Allen, G. Ferguson and A. Stent 2001. An Architecture for More Realistic Conversational Systems. Intelligent User Interfaces Conference.

M. Gatius and M. González 2009. A Flexible Dialogue System for Enhancing Web Usability. 18th World Wide Web Conference.

M. Steedman and R. Petrick 2007. Planning Dialog Actions. 8th SIGdial Workshop on Discourse and Dialogue.

V. Rieser and O. Lemon 2009. Does this list contain what you were searching for? Learning adaptive dialogue strategies for Interactive Question Answering. Natural Language Engineering, 15(1):57- 375. Cambridge University Press.

S. Varges, F. Weng and H. Ponbarry 2009. Interactive question answering and constraint relaxation in spoken dialogue systems. Natural Language Engineering,15(1).Cambridge University Press.

Semantics and pragmatics of negative polar questions

Maria H. Golka

Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland

Maria.Golka@gmail.com

Abstract

This paper aims to provide a literature review about the meaning and use of negative polar (yes/no) questions and complete it with some Polish data. Semantic and pragmatic factors will be discussed. Attention will be drawn to the fact that most of research concentrate on interrogatives themselves, neglecting their possible answers, whereas the latter may be very informative about the nature of the former.

1 Introduction

From a logical semantic point of view, since a polar question $?\phi$ and its negative counterpart $?\neg\phi$ have the same answers, they are logically equivalent (Groenendijk and Stokhof, 1997). It is obvious, however, that if we consider the natural language use of negative polar interrogatives like for example the one in (2), we can not consider them equivalent to positive ones, as in (1).

(1) Is Jane coming?

(2) Isn't Jane coming?

2 Pragmatic and semantic factors

If negative and positive polar questions are semantically the same, why would we use both of them? Considering some common pragmatic intuitions (captured by numerous concepts like Principle of Economy, Principe of Least Effort, Gricean Maxim of Manner or the minimization of cognitive effort in terms of Relevance Theory) negative interrogatives would not be used if their meaning were not at least pragmatically different from that of positive ones.

These intuitions are confirmed by classic experimental results in psycholinguistics. Syntactic transformations of kernel sentences into other structures like interrogatives or negatives are rather charging for the cognitive system. The syntactic form of a sentence (whether it is an active, passive, interrogative or negative clause) seems to be something distinct and more difficult to recall than its semantic content (Mehler, 1963). Syntactically complex sentences, like questions or negatives, require more capacity of immediate memory. Sentences which are both interrogatives and negatives are the ones that are the most hard to process (Savin and Perchonock, 1965). The usage of negative questions that are semantically equivalent to the positive ones but much more difficult to process can thus be explained by pragmatic factors only.

Nevertheless, some approaches find the nature of the distinction between negative and positive polar questions semantic (e.g. Romero and Han, 2004). They are consistent with Ladd's (1981) observations. As Ladd points out, negative polar questions are systematically ambiguous: in case of the "outside negation" reading the speaker believes that the proposition under question is true, whereas in the "inside negation" one the speaker believes it is false.

In this paper we will discuss some examples which show that in Polish Ladd's ambiguity is much more difficult to capture. We will also take into account the possible answers to questions of this kind. It has not been done by most of authors, but it turns out that if we consider the dialogic factors (which in case of questions seem to be very important), the nature of negative vs. positive polar questions distinction appears to be pragmatic. We will argue that even if the internal ambiguity of negative polar questions is due to semantic factors, it is still likely that the distinction between positive and negative questions is pragmatic.

3 Ladd's ambiguity in Polish

Most of the papers on the subject of Ladd's ambiguity (e.g. Romero and Han, 2004; Reese, 2006) discuss polar questions with preposed negation (English interrogative sentences with a negated auxiliary verb) as the one in (2) and ex-

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

clude from consideration interrogative sentences with non-preposed negation, as the one in (3) which permit a neutral interpretation in an unbiased context.

(3) Is Jane not coming?

Since in Polish polar interrogatives are formed by means of an interrogative particle *czy* or with intonation alone, the distinction like that between (2) and (3) is nonexistent. Instead, we have only one type of structure which is rather similar to the structure of an affirmative clause and can by preceded (4) or not (5) with the interrogative particle. This structure conveys all the three readings discussed in the literature (Ladd's outside and inside negation readings, and the neutral one).

(4) Czy Jane nie przychodzi?
INTERR. PART. Jane NEG come$_{3SG, PRES.}$

(5) Jane nie przychodzi?
Jane NEG come$_{3SG, PRES.}$

In Polish, the word order within a sentence is much less strict than the one in English. Consequently, a Polish equivalent of an ambiguous negative polar interrogative, like (6) (the example of Ladd, 1981) would be more naturally represented by a pair of sentences with different word orders where (7a) expresses the outside negation reading, whereas (8a) the inside negation one.

(6) Isn't there a vegetarian restaurant around here?

(7a) Nie ma w okolicy wegetariańskiej restauracji?
NEG be$_{3SG, PRES}$ in neighborhood$_{LOC, SG}$ vegetarian$_{GEN, SG}$ restaurant$_{GEN, SG}$

(8a) Nie ma wegetariańskiej restauracji w okolicy?
NEG be$_{3SG, PRES}$ vegetarian$_{GEN, SG}$ restaurant$_{GEN, SG}$ in neighborhood$_{LOC, SG}$

This difference in word order seems to corroborate Reese's (2006) intuition that "there is no semantic (…) difference between "outside" and "inside" negation. Rather, what is at issue is whether negation targets the core meaning of an utterance or some secondary meaning".

Further inspection reveals some problems with the inside negation reading of interrogatives constructed with the particle *czy*. Interrogatives like (7b) and (8b) are acceptable but none of them can convey an inside negation reading. It seems that the presence of *czy* can somehow trigger the outside negation or neutral understanding of negative polar questions.

(7b) Czy nie ma w okolicy wegetariańskiej restauracji?

(8b) Czy nie ma wegetariańskiej restauracji w okolicy?

Another very interesting phenomenon is the use of the particle c*zyż*. This form is used to construct rhetorical questions and simultaneously deny the proposition under question. Hence, the negative question preceded with *czyż* conveys an affirmative assertion. This kind of construction seems to be a paradigmatic example of an outside negation interrogative.

4 Conclusions

As we have seen, the origins of negative polar questions are hard to define. There is some evidence suggesting that their nature is pragmatic as well as some other evidence, showing their semantic nature. In this paper we try to bring together these two approaches. We provide some evidence from Polish language, as well as evidence about answers. Thus a mixed, semantic-pragmatic model is needed to describe the meaning and use of negative polar interrogatives..

References

Groenendijk, Jeroen and Martin Stokhof. 1997. Questions. In van Benthem, J. and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier, Amsterdam, pages 1055-1124.

Ladd, Robert D. 1981. A First Look at the Semantics and Pragmatics od Negative Questions and Tag Questions. In *Proceedings of Chicago Linguistic Society*, pages 164-171.

Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of English sentence. *Journal of Verbal Learning and Verbal Behavior,* 2(4): 346-351.

Reese, Brian J. 2006. The Meaning and Use of Negative Polar Interrogatives. In Bonami, O. and P. Cabredo Hofherr, editors, *Empirical Issues in Syntax and Semantics 6*, pages 331-354.

Romero, Maribel & Chung-Hye Han. 2004. On Negative Yes/No Questions. *Linguistics and Philosophy,* 27(5): 609-658.

Savin, Harris B. and Ellen Perchonock. 1965. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 4(5): 348-353.

Context-driven dialogue act generation

Volha Petukhova and **Harry Bunt**
Tilburg Center for Creative Computing
Tilburg University, The Netherlands
{v.petukhova,harry.bunt}@uvt.nl

1 Introduction

Generation of dialogue contributions is a matter of deciding which dialogue act(-s) are licensed by the preceding and current context. This paper presents a context-driven approach to the generation of multiple dialogue acts.

The theoretical framework of Dynamic Interpretation Theory (DIT) opens perspectives for developing dialogue act generators that produce utterances which are multifunctional by design by viewing participation in a dialogue as performing several activities in parallel, such as pursuing the dialogue task, providing feedback, and taking turns (Bunt, 2000).

2 Multidimensional context model

An utterance, when understood as a dialogue act with a certain communicative function and semantic content, evokes certain changes in the participant's context model that includes (1) his beliefs about the dialogue task/domain (*semantic context*); (2) his model of the participants' states of processing (*cognitive context*); (3) assumptions about available perceptual channels (*physical context*); (4) beliefs about communicative obligations and constraints (*social context*); (5) a model of the preceding and planned dialogue contributions (*linguistic context*).

3 Context update mechanisms

As a dialogue evolves, new beliefs are *created*; weak beliefs may become *strengthened* to firm beliefs; and beliefs and goals may be *adopted* or *cancelled* (Bunt, 2005).

Speakers normally expect to be understood and believed (*expected effects*). This is modelled in DIT by the speaker having 'weak belief' that the addressee believes the preconditions to hold (*understanding effects*) and the content of the dialogue act to be true (*adoption effects*). Every dialogue builds up a pressure on the addressee to provide evidence in support of or against these expectations.

A *reactive pressure* (RP) is created when a dialogue act is interpreted successfully, giving rise to the intended update of the addressee's context model. The addressee is assumed to strive to resolve RPs by performing a particular type of reactive act. Table 1 illustrates this for the example of a Question - Answer pair. The created pressures RP_1, RP_2 and RP_3 give rise to multiple reactive acts: a Turn Accepting act, a Feedback act, and a task-related Propositional Answer.

Participants are not always able to resolve pressures, e.g. the addressee may not know the answer. This cancels the relevant pressure created by the previous question immediately. Some pressures cannot be resolved in one turn, e.g. the addressee does not understand the question. In this case the pressure PR_3 cannot be relieved, consequently neither can pressure PR_2 also, because this implies PR_3. These pressures remain present until the addressee resolves the pressure PR_3, e.g. by successful processing of the repeated question.

4 Conclusions

The context-driven approach outlined here enables the construction of genuinely multifunctional dialogue contributions, and allows dialogue systems to apply a variety of dialogue strategies and communication styles, e.g. performing explicit vs implicit dialogue acts making use of different modalities.

References

Harry Bunt. 2000. *Dialogue pragmatics and context specification*. H. Bunt and W. Black (eds.), Abduction, Belief and Context in Dialogue. Amsterdam: Benjamins, 81-150.

Harry Bunt. 2005. *Mechanisms for creating and updating beliefs and goals through dialogue*. Unpublished report, Tilburg University.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Table 1: Example of updated context for Propositional Question-Propositional Answer pair
LC = Linguistic Context; SC = Semantic Context; CC = Cognitive Context; prec = preconditions;
impl = by implication; du = dialogue utterance; da = dialogue act; fs = functional segment;
exp.und = expected understanding; und = understanding; exp.ad = expected adoption; ad = adoption;
bel = believes; mbel = mutually believed; wbel = weakly believes

Context	num	source	S's context	num	source	U's context
SC				u01 u02	prec	$wants(U, knowsif(U,p))$ $believes(U, knowsif(S,p))$
LC				du1	U	Is this a large sample?
LC				fs_1 da_1 da_2 plan	current	is, this, a, large, sample Task; PropositionalQuestion Speaker:U; Addressee:S Turn-M.;Turn-Assign Speaker:U; Addressee:S $Turn\_Allocation(S)$
SC	s1 s2 s3 s4 s01	exp.und:u01 exp.und:u02 und:u1 und:u2 prec	$bel(S, mbel(\{S,U\}, wbel(U,$ $bel(S, wants(U, knowsif(U,p)))$ $bel(S, mbel(\{S,U\}, wbel(U,$ $bel(S, bel(U, knowsif(S,p)))$ $bel(S, wants(U, knowsif(U,p)))$ $bel(S, bel(U, knowif(S,p)))$ $believes(S, \neg p)$	u1 u2	exp.und:u01 exp.und:u02	$bel(U, mbel(\{S,U\}, wbel(U,$ $bel(S, wants(U, knowsif(U,p)))$ $bel(U, mbel(\{S,U\}, wbel(U,$ $bel(S, bel(U, knowsif(S,p)))$
CC	s5	und:u3	$believes(S, {}^{+}Interpreted(S, fs_1))$	u3	exp.und:fs_1	$wbel(U, {}^{+}Interpreted(S, fs_1))$
SocC	RP_1 RP_2 RP_3	prec:u01-u02 impl:da_2 exp.und:u1-u3	Task;PropositionalAnswer Speaker:S; Addressee:U antecedent: da_1 Turn-M.;Turn-Accept Speaker:S; Addressee:U antecedent: da_2 Auto-F.;Interpretation Speaker:S; Addressee:U antecedent: fs_1			
LC	da_3 da_4 da_5	plan	Turn Accept Speaker:S; Addressee:U antecedent: da_2 Auto-F.;Interpretation Speaker:S; Addressee:U antecedent: fs_1 Task; PropositionalAnswer Speaker:S; Addressee:U antecedent: da_1			
LC	du2	S	Well, this is not large sample			
LC	fs_2 da_3 fs_3 da_4 fs_4 da_5	current	well Turn Accept Speaker:S; Addressee:U antecedent: da_2 this,is,large,sample Auto-F.;Pos.Interpretation Speaker:S; Addressee:U antecedent: fs_1 this,is,not,large,sample Task; PropositionalAnswer Speaker:S; Addressee:U antecedent: da_1			
SC	s6 s7	exp.und exp.ad	$bel(S, mbel(\{S,U\}, wbel(S, bel(U, bel(S, \neg p))$ $bel(S, mbel(\{S,U\}, wbel(S, bel(U, \neg p)$	u4 u5 u6 u7	exp.und exp.ad und:s6 ad:s7	$bel(U, mbel(\{S,U\}, wbel(S, bel(U, bel(S, \neg p))$ $bel(U, mbel(\{S,U\}, wbel(S, bel(U, \neg p)$ $bel(U, bel(S, \neg p))$ $bel(U, \neg p)$
SocC	RP_1 RP_2 RP_3		cancelled cancelled cancelled	 RP_4 RP_5	 exp.und:s6 exp.ad:s7	 Auto-F.;Interpretation Speaker:S; Addressee:U antecedent: fs_2, fs_3, fs_4 Auto-F.;Execution Speaker:S; Addressee:U antecedent: da_5

The Communicative Style of the Physically Disabled – a Corpus Study

Joanna Szwabe
Chair of Logic and Cognitive Science
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland
joanna.szwabe@amu.edu.pl

Anna Brzezińska
Institute of Psychology
Adam Mickiewicz University
Poznań, Poland

Abstract

The research presented here focuses on the evaluation of communicative abilities and communicative style of the disabled. The study is a part of a national project devoted to the situation and needs of disabled people in Poland. As a part of the project the Corpus of Dialogs of Disabled Speakers has been compiled and annotated. The corpus analysis focused on differences in language use between disabled speakers and controls as well as within the group of disabled speakers contrasting 6 groups of subjects with various disabilities. Dialogs were analyzed with regard to syntactic, semantic and pragmatic features and their relation to demographic factors. The corpus analysis shows that there are significant differences as far as communicative style of the disabled people is concerned, expressed by excessive use of non-standard forms, specific conceptual metaphors, over-passivisation, and other features. The results will be presented in detail on the poster.

1 Introduction

The present study is a part of a national project devoted to the situation and needs of disabled people in Poland[1]. The goal of the research presented here has been the evaluation of communicative abilities and communicative style of the disabled. The language of the disabled is still poorly understood. The literature dedicated to the subject is scarce not only as far as corpus studies are concerned but also when it comes to research on the adult disabled speech in general with the exception of the analysis of the language of mentally and psychologically disturbed (cf.. Happé 1993, Langdon et al. 2002, Woźniak 2000). At the same time there is a growing social recognition of the problems specific to this group of speakers.

2 Method

In the course of the present project realization the Corpus of Dialogs of Disabled Speakers (CDDS) has been compiled and annotated. The corpus consists of transcribed and annotated group conversations of 113 subjects. The CDDS language is Polish. The video recordings of the conversations have been transcribed and annotated according to a tagset designed for the purpose of this study. The Corpus has 402 146 tokens, including 225 299 words of raw text. The Corpus is fully tagged with nearly 100 types of tags coding various parameters of language structure and both verbal and nonverbal communication, including pragmatic annotation.

The corpus analysis focused on differences in language use between disabled speakers and comparatively controls as well as within the group of disabled speakers contrasting 6 groups of subjects with various disabilities pertaining to: motion, sight, voice, psyche, mind, and other. The utterances were analyzed with regard to syntactic, semantic and pragmatic features and their relation to a number of demographic factors, onset of disability, etc. Over 60 parameters have been examined of which here we mention just a few. Additionally, on the basis of the Corpus the Affective Lexicon Index has been prepared and the question of language negativisation in the speech of the disabled has been examined.

3 Discussion of the results

Typically, disabled speakers live in small and closed linguistic communities around social care centers offering support. This practical solution is undoubtedly advantageous for numerous reasons but may affect their language development. Despite social isolation related to disability linguistic and communicative competence of the

[1] The project has been financially supported by the European Social Fund.

disabled subjects in our study shows no signs of pathology.

Nevertheless, there are significant differences as far as communicative profile of the disabled people is concerned. The corpus analysis shows that this diversity is not limited to few features but affects the whole series of linguistic elements on every level of language structure, so that we can speak of communicative profile of the disabled.

One of the most alarming features is the excessive use of original forms deviating from the socially accepted standard variation such as language errors, slang, regional forms, egocentrically oriented neologisms. The remaining tendencies specific to the the disabled people have a character of a communicative style and are less disturbing as far as language as such is concerned. However, they signal very clearly the problems of psychological nature. The most significant of this group are: firstly, the passive attitude expressed by syntactic and semantic structures, and secondly extensive use of figurative language, a particular preference for non-explicit expressions for the description of plain situations. The pragmatic level of language is not impoverished despite social isolation resulting from disability but even dominates through overwhelming use of non-literal utterances such as implicatures, and especially numerously represented conceptual orientational metaphors. It is worth noting that subjects with motion and sight disability tend to use metaphors related to the experiential basis they are deprived of, namely to walking and seeing.

It was hypothesized that disabled people may be at risk of perceiving themselves not in terms of agents but rather as objects of other people actions. Indeed, the analysis of syntactic and semantic markers of the phenomenon of passiveness such as increased use of passive voice and gerund clause indicates that disabled speakers tend to depict situation from the perspective of an object of action. On the level of utterance semantics, the same topics are differently correlated in the speech of disabled speakers and in the speech of controls. For instance, the topic of problems and obstacles - belonging to the most significantly represented in the speech of disabled speakers from 40 topics tagged and examined in the corpus – is associated in the speech of the disabled with acquiring help, while in the speech of healthy subjects with active problem solving. This clearly reveals problems signaled in language but having roots both in linguistics and psychology. As the communicative competence of the disabled speakers is not deviating from the norm there is a chance for improving the present state. The studies like the one presented here may contribute to a better understanding of the language development under specific conditions of disability and help improve the situation of the disabled by greater integration of this group of speakers within the society as well as supporting their motivation for self-sufficient life, which is a prerequisite for a sense of having control of their lives.

References

Baratta, A. M. (2009). Revealing stance through passive voice. *Journal of Pragmatics*, *41*(7), 1406-1421.

Cameron, L., & Deignan, A. (2003). Combining large and small corpora to investigate tuning devices around metaphor in spoken discourse. *Metaphor and Symbol*, 18(3), 149–160.

Happé, F. (1993) Communicative Competence and Theory of Mind in Autism: a Test of Relevance Theory. Cognition Tom 48; 101-119.

Langdon, R. & Coltheart, M. & Ward, P. & Catts, S. (2002). Disturbed Communication in Schizophrenia: the Role of Poor Pragmatics and Poor Mind-reading w *Psychological Medicine* 32; 1273-1284.

Sanso, A. (2006). Passive and impersonal constructions in some European languages. In W. Abraham & L. Leisio (Eds.), Passivization and typology: Form and function (pp. 232-272). Amsterdam/Philadephia: John Benjamins Publishing Co.

Wikberg, K. (2003). Studying metaphors using a multilingual corpus. In K. M. Jaszczolt & K. Turner J. (Eds.), Meaning through Language Contrast (pp. 109–123). Amsterdam/Philadephia: John Benjamins Publishing Co.

Woźniak, Tomasz. (2000). *Zaburzenia języka w schizofrenii[Disturbed language in schizophrenia]*. Lublin: Wydawnictwo UMCS.

Communicating routes to older and younger addressees

Thora Tenbrink
University of Bremen
Bremen, Germany
tenbrink@uni-bremen.de

Elena Andonova
University of Bremen
Bremen, Germany
andonova@uni-bremen.de

Abstract

Our study addressed the extent to which route descriptions reflect different concepts of addressees as a function of age, with respect to route choice, semantic elaboration, politeness forms and syntactic complexity. 55 native speakers of German wrote route descriptions for imagined addressees supposed to be either 25 or 75 years old. Results reveal that participants' consideration of their addressee is reflected in their differentiated use of politeness forms, degree of syntactic complexity, and in the ways in which routes were selected for younger vs. elderly people. However, route descriptions for elderly addressees did not reflect increased semantic elaboration (here: providing more details about the route).

1 Introduction

Speakers are known to be sensitive to their interaction partners' knowledge and ability. Route descriptions are particularly suitable for investigating the extent to which the abilities presumed on the part of the addressee are taken into account, since they relate to a predefined spatial environment as well as a clear discourse goal: to enable the addressee to reach their destination. Here we address speakers' strategies when asked to write a route description for an addressee about whom they know nothing except age and gender. Our aim is to contribute to research on age-related talk, in particular with respect to the extent to which speakers intuitively adhere to a principle found to be useful for elderly addressees, namely, semantic elaboration – explaining a particular piece of information in more than one way (Kemper et al., 1995). This idea is in the present scenario represented by distinct levels of granularity as defined by Tenbrink and Winter (2009). This framework distinguishes between crucial spatial units (segments of the route) and those that are not always mentioned explicitly in route instructions, and differentiates the types and amount of detail about each spatial unit that is provided by route givers. We hypothesized that speakers may provide different amounts of detail as a function of age of addressee.

Interaction style and language use differ systematically with elderly addressees to such an extent that *elderspeak* has been identified as a special speech register. In elderspeak, speakers appear to adapt to the communicative and cognitive needs of their interlocutors guided by assumptions about their limited language, cognitive, and/or physical ability. In a route drawing task involving dialogues between older and younger speakers, Kemper et al. (1995) found that younger speakers simplified their speech for elderly addressees by talking more slowly, using shorter sentences and fewer subordinate clauses. At the same time, they provided more information about the routes to be drawn by repeating utterances and using more varied vocabulary, as well as providing more location checks per map. Addressing comprehension of route instructions, Kemper & Harden (1999) found that increasing semantic elaborations and reducing use of subordinate and embedded clauses improved performance although reduced length did not.

While earlier studies such as these provide a number of relevant insights about the types of adjustments made for elderly addressees, they do not build on research in spatial cognition that highlights how speakers' concepts of routes are represented in language. Previous research has revealed a range of spatial aspects that speakers typically refer to when describing a route to a wayfinder, such as the route's start and end points, landmarks, directions, paths, actions, regions, and distances (Denis, 1997; Tversky and Lee, 1998). Routes are sometimes described at a finer grained level than that dictated by the decisions to be taken along the way. Landmarks are mentioned not only at decision points but also in between decision points (Herrmann et al., 1998). Also, additional path information may be provided even without a change of direction (Habel,

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

1988). Such information keeps the traveler confident particularly in cases of potential sources of uncertainty (Tversky and Lee 1998).

The perception of what kind of supportive information may be required by a wayfinder can differ widely across individuals and task situations. Since speakers are known to adapt their language to the listener (Clark and Krych, 2004), any aspects known about the addressee could have an impact on the spatial descriptions formulated for them (Herrmann and Grabowski, 1994). For example, the choice of reference frames and perspectives is affected by interactive alignment, adaptation, and interlocutor priming processes as well as by the addressee's perceived abilities (Schober, 1993, 2009; Watson et al., 2004).

For route descriptions, it has been established that levels of granularity or complexity may differ according to the situation, for instance in relation to problematic segments or decision points (Tenbrink and Winter, 2009). However, only little is known about speakers' flexibility in relation to different addressees with respect to the communication of route-related details. While the studies by Kemper and colleagues above point to a positive effect of semantic elaboration on elderly listeners' comprehension that might be used to enhance the efficiency of automatic dialogue systems providing route instructions (Thomas, 2010), it is unknown to date what kinds of spatial concepts should be enhanced semantically. Also, the extent to which a schematic map scenario might transfer to a real-world scenario involving multimodal travel (i.e., public transport in addition to walking) remains unclear. A schematic map offers only a limited amount of information that could be verbalized; in contrast, the real world consists of an almost infinite number of features that might in theory be referred to in a route description. Furthermore, in natural environments there is typically more than one option for traveling. A recent study set in a complex city environment established that routes are chosen differently for one's own future navigation than for somebody who is not familiar with the environment (Hölscher et al., subm.). It stands to reason that route choice might systematically be affected by the age of the intended addressee because of general assumptions of such an addressee's physical, cognitive or communicative constraints.

In our study we set out to investigate how speakers confronted with a route instruction task involving their own natural everyday surroundings react to the requirement of providing route information to either younger or older addressees. We hypothesized that the amount of detail conveyed about a route segment, which is influenced by features of the spatial environment, may be further mediated by the concept of an addressee of a particular age. Furthermore, route givers may select different kinds of routes for their addressees depending on age.

2 Route Description Study

2.1 Method

55 native speakers of German who were familiar with the Bremen university campus were recruited via an email call and participated in the study by email (20 were male and 35 female; 4 between 30–49 and 51 between 18–29 years old). Their task was to describe the route from the train station in Bremen to one of two buildings on campus (the library or the Cartesium building). The intended addressee of the route description was either male or female and either 25 or 75 years old. Participants were assigned to conditions randomly. Thus, the design of the study was 2 (addressee's age: 25 vs. 75 years old) x 2 (addressee's gender: male vs. female) x 2 (destination: the Cartesium or the library).

2.2 Analysis

The route descriptions were annotated by coders blind to the purpose of the study and the design conditions. Since the majority of the participants chose the same routes for the library and the Cartesium destinations, respectively, we focused on these two "standard" routes and identified others as exceptions (alternatives to these routes). We addressed the distribution of spatial details by first identifying the *spatial units* (Tenbrink & Winter, 2009) constituting the two standard routes: segments along the route that were described in a particular order by participants (the temporal order of route travelling). Next, we identified the number of *detail units* (pieces of information given in no particular order within a description) within each spatial unit that were mentioned more than once (i.e., by different participants). Spatial units that were explicitly mentioned by all participants were identified as *crucial*. This analysis yielded a semantically based hierarchical measure of *crucial spatial units* at the highest level of granularity, followed by the number of *spatial units* referred to in a description, and then by the number of (non-idiosyncratic) *detail units*.

We identified those parts of each description that referred to spatial units of the same two standard routes, and normalized the measures by participant by calculating the ratio of occurrence of each category for each participant. Our hypothesis was that, while the *crucial spatial units* and the *mention of a particular spatial unit* should be independent of age, the amount of *detail* should differ if people consider the requirement of semantic elaboration for elderly addressees.

Apart from this semantic analysis of spatial information provided along the route, we also calculated the average number of words per (shared) spatial unit, capturing in this way also the idiosyncratic cases in which further details were mentioned by individuals for a particular spatial unit shared across descriptions, and we looked at the following features of (complete) descriptions:

- mean length of a sentence (the number of words divided by the number of sentences),
- relative frequency of syntactically simple vs. subordinate (complex) sentences (leaving aside co-ordinate sentences, which may be judged as intermediate concerning syntactic complexity),
- form of address: informal "du" vs. formal "Sie"; further alternatives found were neutral infinitives, and third person singular.

2.3 Results

Route choices differed for younger and older addressees, and descriptions differed with respect to politeness forms and syntactic complexity as a function of age of addressee. We ran a series of 2 (age: younger vs. older) x 2 (gender: male vs. female) analyses of variance on the variables of interest. We found robust main effects of addressee age on the use of "Du", $F(1,51)=21.84$, $p<.001$, as well as on the use of the polite "Sie", $F(1,51)=30.56$, $p<.001$, and a two-way interaction between age and gender of addressee on the use of "Sie", $F(1,51)=7.32$, $p<.01$. We also found a marginally significant effect of age, $F(1,51)=3.96$, $p=.052$, on the simple/subordinate sentence ratio. Elderly people (particularly men) were addressed consistently by "Sie" and descriptions tended to be syntactically simpler; whereas young people (particularly men) were addressed informally by "du" (or in a neutral form), and descriptions tended to be syntactically more complex (particularly for young women).

Standard routes were preferred for elderly addressees; younger addressees received alternative, more challenging but possibly shorter routes more often. Altogether, 11 descriptions for 25-year-olds and 5 for 75-year-olds used an alternative route. A 2 (younger vs. older addressee) x 2 (standard vs. alternative route) chi-square analysis showed a marginally significant association between these variables ($\underline{x}^2=3.49$, $p=.062$). Alternative routes for 25-year-olds consistently concerned either walking diagonally across a parking lot towards the Cartesium building, or following the tramline on a narrower path rather than walking directly from the tram into the main university entrance in order to reach the library. Both of these were only suggested in one description each for a 75-year-old; the remaining three alternative routes chosen for elderly addressees concerned variations of public transport that were never offered to younger people.

In contrast to these consistent differences in route descriptions as a function of age of addressee, the analysis of spatial units and density of details did not reveal any systematic differences with respect to semantic elaboration. As expected, the spatial semantics contained in the route descriptions were hierarchically structured, independent of age of addressee. Neither the number of spatial units mentioned for the standard route, nor the mean number of words per shared spatial unit differed as a function of age of addressee. Descriptions for 25-year-olds contained on average 7.74 spatial units with an average of 16.40 words per unit. Descriptions for 75-year-olds contained on average 8.07 spatial units with an average of 15.50 words per unit.

Across all shared spatial units, the mean number of details mentioned per unit was 2.09 for 25-year-olds and 2.11 for 75-year-olds. The number of details varied across spatial units, the highest number was 4.48 details on average towards 25-year-olds and 4.11 towards 75-year olds for one particular spatial unit; the lowest number (for a different spatial unit) was 0.89 details towards 25-year-olds and 0.93 towards 75-year-olds. For each single spatial unit, average numbers were similarly close, i.e., independent of age of addressee, as in these examples. In other words, participants did not provide an enhanced level of detail for any spatial unit for elderly addressees.

2.4 Discussion

In our study, route givers wrote descriptions in different ways as a function of age of addressee. They not only adapted their route descriptions with respect to politeness forms and syntactic complexity, but also carefully considered which route their addressee should take. This yielded systematic differences in route choices in spite of

the fact that the spatial environment apparently supported one feasible "standard" route per destination which was used far more often across all descriptions than any other choice.

However, our analysis of semantic elaboration in terms of spatial granularity revealed no systematic differences as a function of age. This result stands in contrast to earlier findings by Kemper et al. (1995), who found that speakers used semantic elaboration when addressing elderly addressees. One reason for this difference may concern the fact that written descriptions provide a permanent medium of communication to aid the recipient's memory. A further enhancement of already mentioned material, which is easily accessible upon re-reading, would then appear redundant. As a result, semantic elaboration effects may be particularly present in the spoken modality as a way of facilitating memory and (semantic) integration of information. Another reason for the differences found between our analysis and Kemper's work may concern the analytical measures used. While Kemper and colleagues focused on formal measures such as repetition and variability in word forms, our analysis was concerned with the conveyance of facts, i.e., particular details about the environment that may support the traveler in finding the correct route in addition to the communication of essential spatial segments and decision points.

Our analysis revealed that, similar to earlier research (Tenbrink and Winter, 2009), route descriptions exhibited a hierarchical structure in that some of the spatial elements were considered so necessary as to be mentioned by every single route giver and some (more complex ones) were elaborated by many details, while others were left implicit in some descriptions and/or enhanced by fewer details on average. These patterns appeared to reflect solely the features of the spatial environment rather than any specific requirements attributed to the addressee. Thus, while route descriptions systematically varied the levels of granularity in relation to the nature of segments, route givers apparently did not expect elderly wayfinders to require more details about problematic spatial segments than younger ones. Instead, if they judged a particular spatial segment to be too problematic for an elderly addressee, they rather suggested a different route.

Acknowledgements

Funding by the DFG (SFB/TR 8 Spatial Cognition) is gratefully acknowledged. We thank Kavita Thomas and our assistants for their support.

References

Clark, Herbert H. and Meredyth A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language 50*, 62–81.

Denis, Michel. 1997. The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive*, 16(4):409-458.

Habel, Christopher. 1988. Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In: H. Schnelle & G. Rickheit (eds.), *Sprache in Mensch und Computer*. (pp. 107-133). Opladen: WDV.

Herrmann, Theo and Joachim Grabowski. 1994. *Sprechen. Psychologie der Sprachproduktion*. Heidelberg: Spektrum.

Hölscher, Christoph, Thora Tenbrink, and Jan Wiener (subm.) Would you follow your own route description?

Kemper, Susan, and Tamara Harden. 1999. Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology & Aging, 14*(4), 656-670.

Kemper, Susan, Vandeputte, Dixie, Rice, Karla, Cheung, Him, and Gubarchuk, Julia. 1995. Speech adjustments to aging during referential communication task. *Journal of Language and Social Psychology, 14*, 40-59.

Schober, Michael F. 1993. Spatial Perspective-Taking in Conversation. *Cognition 47*: 1-24.

Schober, Michael F. 2009. Spatial Dialogue between Partners with Mismatched Abilities. In Kenny Coventry, Thora Tenbrink, and John Bateman (eds.), *Spatial Language and Dialogue* (pp. 23-39). Oxford University Press.

Tenbrink, Thora and Stephan Winter. 2009. Variable Granularity in Route Directions. *Spatial Cognition and Computation 9*, 64 – 93.

Thomas, Kavita E. 2010. Dialogue systems, spatial tasks and elderly users: A review of research into elderspeak. *SFB/TR 8 Report 021-02/2010*, University of Bremen.

Tversky, Barbara and Paul U. Lee. 1998. How Space Structures Language. In: C. Freksa, C. Habel and K.F. Wender (eds.), *Spatial Cognition.* (pp. 157-175). Berlin: Springer.

Watson, Matt E., Martin Pickering, M. J., and Holly P. Branigan. 2004. Alignment of Reference Frames in Dialogue. In K.D. Forbus, D. Gentner, and T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1434-1440). Hillsdale, NJ: Lawrence Erlbaum.

Entailed feedback: evidence from a ranking experiment

Marcin Włodarczak
Bielefeld University
Bielefeld, Germany

Harry Bunt
Tilburg University
Tilburg, the Netherlands

Volha Petukhova
Tilburg University
Tilburg, the Netherlands

mwlodarczak@uni-bielefeld.de, {harry.bunt,v.petukhova}@uvt.nl

Abstract

In this paper we investigate relationships between entailment relations among communicative functions and dominance judgements in an annotation task in which participants are instructed to rank utterance functions in terms of their importance. It is hypothesised that on average entailed functions should be assigned lower ranks than entailing functions. Preliminary results of an experiment are reported for positive auto-feedback functions which are argued to be entailed by backward-looking functions such as *Confirm*.

1 Introduction

Since communication is a multi-faceted process in which participants must monitor and manage several aspects of their behaviour simultaneously, utterances which they produce are often multifunctional. Moreover, some of the utterance functions might entail or implicate other functions (Bunt, 2009a). In particular, backward-looking functions (Allen and Core, 1997), such as *Confirm*, *Answer* or *Agreement*, are claimed to entail positive feedback about some earlier utterance.

At the same time, it might be argued that some of the utterance functions have a priority over its other functions because achieving some communicative goals is of greater importance to the speaker in a given context.

The paper explores the hypothesis that entailed and implicated functions could be expected to be ranked lower than entailing functions. Specifically, preliminary results for entailed autofeedback are reported.

2 Multidimensional Tagsets

Multifunctionality of utterances suggests that in an annotation task each utterance should be allowed to be labelled with more than one tag. This process is greatly facilitated if tags are organised into clusters such that only one tag for a cluster can be assigned to an utterance (Allen and Core, 1997; Clark and Popescu-Belis, 2004; Popescu-Belis, 2005; Popescu-Belis, 2008). Such clusters are commonly referred to as dimensions.

A conceptually-motivated definition of a a dimension was provided by Bunt (2006). He defines it as an aspect of participating in dialogue which (1) can be addressed by means of dialogue acts that have communicative functions specific for this purpose, (2) can be addressed independently of other aspects. The first criterion requires that the proposed aspects of communication correspond to observable dialogue phenomena. The second requires that the dimensions be orthogonal.

3 Semantic types of multifunctionality

Bunt (2009a) distinguishes the following semantic forms of multifunctionality: *independent*, *entailed*, *implicated* and *indirect*.

3.1 Independent multifunctionality

Two or more communicative functions are independent when each is expressed by some features of a segment. An example could be "thank you" spoken with cheerful intonation and high pitch, which might signal both gratitude and goodbye.

From a point of view of information-state update approaches, such as DIT (Bunt, 1994), in which a dialogue act corresponds to a context update operation, indirect multifunctionality could be interpreted as independent update operations of addressee's information state, one for each function.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

3.2 Entailed multifunctionality

Entailed multifunctionality occurs when preconditions of one communicative function logically imply preconditions of another function.

Such relations usually occur between communicative functions of which one is a specification of the other (e.g *Warning* and *Inform*), and hold between functions in the same dimension. In terms of context update, the update operation of the entailed function is subsumed by the update operation of the entailing function. Such entailed functions are, therefore, semantically vacuous.

There is, however, a less trivial case of entailment between functions, namely between *auto-feedback functions*, which provide information about the speaker's processing of some previous utterance (Bunt, 2009b), and *backward-looking functions*, such as *Answer*, *Confirm* or *Accept Request*, which "indicate how the current utterance relates to the previous discourse" (Allen and Core, 1997). Obviously, responding to some earlier utterance of the communication partner implies positive processing of the utterance being responded to (or at least speaker's belief that this was the case). Importantly, entailed feedback should be seen as a real source of multifunctionality since it involves an update of speaker's assumptions about the processing of a previous utterance by himself and his partner.

3.3 Implicated multifunctionality

Implicated functionality is found when one of the functions of a segment occurs by virtue of a conversational implicature. It is, therefore, context-dependent and intentional, and corresponds to an additional context update operation. An example is positive feedback implicated by shifting to a new but related topic.

3.4 Indirect multifunctionality

Indirect multifunctionality is a result of an indirect speech act. It is argued, however, that in information-state update approaches many of the indirect speech acts can be analysed in terms of conditional dialogue acts. For example, an utterance such as "Do you know what time it is?" is analysed as "Please tell me what time it is if you know." It remains an empirical question whether all indirect acts could be analysed in this way.

4 Ranked Annotation System

Ranked multidimensional dialogue act annotation was proposed by Włodarczak (2009). It assumes that while utterances are multifunctional, in a given context accomplishing some of the speaker's goals is of greater importance than accomplishing some other goals, and, hence, some utterance functions might dominate other functions.

The relative prominence of communicative functions was modelled by means of a *greater or equal prominence* relation, where the term *prominence* denotes the significance of a communicative function relative to other functions of the same utterance. It is assumed that prominences of any two functions of the same utterance are comparable, i.e. it is possible to decide whether one of the functions is more prominent than the other or whether they are equally prominent. Since more than one function is allowed to have the same prominence, the relation in question imposes a *non-strict linear order* on the set of functions of an utterance. As indicated above, the ordering of functions is viewed here from the speaker's point of view, i.e. it reflects the hierarchy of speaker's communicative goals.

Importantly, the above framework offers more flexibility than similar approaches (e.g. *Dominant Function Approximation*, Popescu-Belis (2008)) by allowing more than one highest ranking function, and more than two different ranks.

5 Ranking and types of multifunctionality

Given that entailment relations between communicative functions of utterances are formulated in terms of function preconditions and context update operations, and prominence relations are defined in terms of hierarchy of communicative goals, a connection between the two could be stipulated. Namely, it could be expected that in an annotation task in which participants are instructed to rank communicative functions of each utterance, independent functions, expressed by segment features and possibly corresponding to independent, highest-ranking communicative goals, should be in most cases assigned the same rank. On the other hand, entailed and implicated functions, corresponding to subordinate goals (and, at least for intra-dimensional entailment, semantically vacuous), could be expected to be ranked lower than functions expressed explicitly by utterance features. Similarly, entailed auto-feedback

should in most cases be ranked lower than the entailing backward-looking functions.

6 Experiment

Two experiments were conducted to find out whether the backward-looking *Confirm* function is ranked higher than the entailed positive auto-feedback. In the first experiment, annotators were asked to order functions assigned to segments with respect to their relative prominence. In the second, the annotators were first asked to assign the (possibly multiple) applicable communicative functions to pre-defined segments and then assign a prominance rank to each of them.

6.1 Corpus and tagset

HCRC Map Task Corpus was used in both experiments[1]. Map task dialogues are so-called instructing dialogues in which one participant navigates another participant through a map. The total duration of the data selected for the experiments was equal to 4 minutes and 43 seconds.

The tagset chosen for the experiment was the DIT^{++} dialogue act taxonomy (Bunt, 2009b). It consists of ten dimensions related to managing the task domain (*Task/Activity*), feedback (*Allo-* and *Auto-feedback*), time requirements (*Time Structuring*), problems connected with production of utterances (*Own* and *Partner Communication Management*), attention (*Contact Management*), discourse structure (*Discourse Structuring*) and social conventions (*Social Obligations Management*).

The data was segmented in multiple dimensions according to the approach presented in (Geertzen et al., 2007). 136 functional segments were identified. For the first experiments the data was annotated by two expert annotators. Full agreement was established on segmentation and annotation level beforehand. Specifically, ten segments were labelled with a *Confirm* tag.

6.2 Participants, task and procedure

The experiments were performed by naive annotators. The annotators were four undergraduate students. They had been introduced to the annotation scheme and the underlying theory while participating in a course on pragmatics during which they were exposed to approximately three hours of lecturing and a few small annotation exercises on data other than map task dialogues.

All annotators accomplished both tasks individually, having received the materials (transcriptions and sound files) in electronic form. Time for both tasks was not limited. To encourage high quality of annotations the students were motivated by an award of 10% of the total grade for the pragmatic course.

In both experiments the ordering was done by assigning each function a numerical value from the set of subsequent natural numbers, starting from "1" as the most dominant function. More than one function could be assigned the same numerical value.

Since in the second experiment the same dialogue material was used, a two week break was made between the experiments to avoid the annotators being biased by the pre-annotated data from the first experiment.

As pointed out in section 3, entailed auto-feedback is a source of true multifunctionality. Annotators were, therefore, asked to include it in their annotations.

6.3 Results and discussion

Table 1 presents inter-annotator agreement about functions assigned ranks of one, two and three for each annotator pair and for each experiment calculated using Cohen's kappa (Cohen, 1960). In the first experiment the mean kappa values for the ranks of one, two and three were equal to 0.64, 0.62 and 0.85 respectively, which indicates a high degree of agreement. For the second task the respective values are substantially lower (0.42, 0.27 and 0.58 respectively) but it should be borne in mind that these scores indicate agreement on a joint task of annotation and ranking, and that the annotators only had limited annotation experience.

In the first experiment the annotators reached nearly perfect agreement about ranking the entailed auto-feedback function lower than the entailing *Confirm* function. In the second task the entailed feedback was ranked higher in three cases.[2] Additionally, there were some cases in which the annotators did not annotate the entailed feedback at all, which in itself might indicate that it is treated as less prominent by inexperienced annotators. Table 2 gives inter-annotator agreement regarding the relative ranks of *Confirm* and entailed

[1] Detailed information about the project can be found at `http://www.hcrc.ed.ac.uk/maptask/`

[2] All these cases came from one annotator, and correspond to 27% of all *Confirms* identified by this participant.

Annotators	Ranking of pre-annotated functions			Annotation and ranking (joint task)		
	Rank 1	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
1 & 2	0.63	0.59	0.84	0.53	0.32	0.76
1 & 3	0.76	0.75	0.92	0.38	0.17	0.57
1 & 4	0.62	0.54	0.85	0.43	0.29	0.57
2 & 3	0.53	0.53.	0.82	0.51	0.19	0.57
2 & 4	0.59	0.55	0.83	0.34	0.21	0.61
3 & 4	0.69	0.71	0.86	0.34	0.44	0.42

Table 1: Cohen's kappa scores for two rating experiments per annotator pair

Annotators	Experiment 1	Experiment 2
1 & 2	1.00	0.29
1 & 3	0.89	0.19
1 & 4	0.74	0.37
2 & 3	0.89	0.93
2 & 4	0.74	0.60
3 & 4	0.89	0.48

Table 2: Cohen's kappa scores for relative ranks assigned to auto-feedback and *Confirm* for two experiments per annotator pair

auto-feedback for each experiment and each annotator pair. Mean kappa values were equal to 0.86 and 0.48 for the pre-annotated and not pre-annotatated data respectively[3].

7 Conclusions

A strong tendency was found for entailed positive feedback to be ranked lower than the entailing *Confirm* function by naive annotators. This was true both for dialogues in which functions were pre-annotated by experts and those in which annotators assigned functions to pre-defined segments themselves.

Although the low number of analysed items does not allow to draw definite conclusions, the results suggest that entailment relations might be a major factor influencing relative promincences of communicative functions, with entailed functions being perceived as less prominent than entailing functions.

Additionally, inter-annotator agreement about ranking of pre-annotated functions was found to be very high, with fair to moderate agreement in the joint task of annotation and ranking.

[3]Since three annotators failed to rank functions of some utterances in the first experiment, kappa values are not precisely equal to one.

References

James Allen and Mark Core. 1997. *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1)*. Technical Report, Multiparty Discourse Group, Discourse Resource Initiative

Harry Bunt. 1994. Context and Dialogue Control. *Think Quarterly 3(1)*.

Harry Bunt. 2006. Dimensions in Dialogue Act Annotation. *Proceedings of LREC 2006*, Paris.

Harry Bunt. 2009a. Multifunctionality and multidimensional dialogue semantics. *Proceedings of DiaHolmia 2009*, Stockholm.

Harry Bunt. 2009b. The DIT^{++} Taxonomy for functional dialogue markup. *Proceedings of the EDAML@AAMAS, Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts"*, Budapest.

Alexander Clark and Andrei Popescu-Belis. 2004. Multi-level Dialogue Act Tags. *Proceedings of SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialogue)*, Cambridge, MA.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20: 37–46.

Jeroen Geertzen, Volha Petukhova and Harry Bunt. 2007. A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. In: *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.

Andrei Popescu-Belis. 2005. *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper n. 62, University of Geneva.

Andrei Popescu-Belis. 2008. Dimensionality of dialogue act tagsets: an empirical analysis of large corpora. *Language Resource and Evaluation 42 (1)*.

Marcin Włodarczak. 2009. *Ranked multidimensional dialogue act annotation*. MA thesis, Adam Mickiewicz University, Poznan.

Statistical Evaluation of Intention in Group Decision-making Dialogue

Michael Wunder
Rutgers University
mwunder@cs.rutgers.edu

Matthew Stone
Rutgers University
matthew.stone@rutgers.edu

Abstract

We explore the dialogue implications of a strategic voting game with a communication component and the resulting dialogue from laboratory experiments. The data reveals the role of communication in group decision making with uncertainty and personal biases. The experiment creates a conflict in human players between selfish, biased behavior and beneficial social outcomes. Unstructured chats showcase this conflict as well as issues of trust, deception, and desire to form agreement.

1 Introduction

Communication has a long history within game theory. From the earliest examples of signaling games (Lewis, 1969) to their most recent acceptance as an essential component of interacting computer agents (Shoham and Leyton-Brown, 2009), the exchange of information alters both strategies and incentives. Natural dialogue shows a human response to problem-solving situations.

Previous work has investigated information sharing among networks of individuals with partial information (Choi et al., 2005) or a personal bias towards sending misinformation (Wang et al., 2009). These experiments use the given task to explore how actors reason about the game.

In this work we discuss a task for a group to make a simple binary decision with both of these forces at work. Given noisy private signals about a risky policy, players must decide whether to take the risky route. Players may be biased toward one outcome and while most share their private data, they are not forced to report accurately. Before the vote, there is an opportunity to chat with others using natural language. The data provides insight into the various mechanisms of persuasion and trust people bring to such games. Our insights into subjects' information and decision making offer a complementary perspective to previous studies of collaborative dialogue (DiEugenio et al., 2000).

2 Experiment: Voting Game with Chats

In two series of behavioral economics experiments, five players were told to decide between two collective actions with different payouts.

- Game 1: One Risky Policy, Vote Yes or No. Policy succeeds with probability p. If a group votes yes, it gets average score p. If a majority votes no all get 0.5.
- Game 2: Two Opposing Policies: A or B. If a group votes A, it gets average score p and otherwise $1 - p$.

All players receive a noisy signal s that nobody else sees, drawn from a distribution centered on p. They converse with other players through anonymous chat boxes, and then must vote on the policy. In addition to s, players are given a bias, a personal payoff adjustment ppa that shifts the amount a player receives in the event the policy is passed.

3 Decision Factors

There are several key decisions to be made over the course of a single round. First, each participant has the option to share their private signal. As a result of social pressure, a number is reported almost every time. Since a player controls only this piece of information, there is an opportunity to falsify what they say to push the group's decision towards an outcome that is personally beneficial or beneficial to the score of a subgroup.

The other major decision is the vote. Before the actual vote, typically there is discussion about the merits of each choice, and some agreement may be reached. Two competing factors that affect this decision are reliance on one's known personal signal and the available public knowledge.

Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue

Table 1: Percentage of voter type in experiments

Type of voter	A/B	Y/N
Self-interest vote (s + ppa)	0.12	0.11
Vote based on group interest	0.14	0.19
Both factors align with vote	0.54	0.60
Neither factor aligns with vote	0.20	0.10

By isolating the information ultimately available to each player we identify the goal of most players (see Table 1). In most cases, the interests of the individual and that of the group, measured by the average signal, line up with the chosen option. When indicators conflict, players will choose one condition over the other. For the purpose of modeling communicative strategy, it is useful to know how the voting decision is aligned with the phrases used to propose courses of action.

We have identified five major phrase types that are used when someone would like to indicate voting preference, which are *Declarative*, *Suggestive*, *Question*, *Imperative*, and *Everyone says*. Different forms can have very different connotations even with the same root words. The relative frequency of such phrases answers questions about how people negotiate based on the expected outcomes. The types of utterances do indicate how strong the evidence is, and in turn how committed people are to the vote expressed in the message. In addition, we have found that people use different negotiation tactics based on their interests, such as personal versus group. The poster presents results in detail.

4 Experimental Results

We have posed a number of questions given the corpus attached to this experiment.

- How are conversations organized?

Typically, there are three phases to each conversation. Players first exchange signals, then they discuss the merits of each choice. The reasons discussed include the average signal, biases, and riskiness. Finally they announce decisions by either coming to an agreement or not. Mostly these tasks take much less time than is available and so players also conduct side talk.

- Who lies and how much does it pay off?

We have found that liars do take advantage of gullible partners occasionally. Somewhere between 10% and 20% of the signals passed to others are adjusted in some way. There is also some evidence that too much exaggeration can backfire on the liar as well as the group.

- What phrases do people use to push for one result over another, and how do they affect votes and scores?

Selfish voters and group interest voters differ somewhat in their choice of words used to indicate votes. For instance, when expressing a vote for the public good, players tend to use the words "we" and "everyone" more often.

5 Conclusion

In keeping with previous negotiation models (DiEugenio et al., 2000) we can see our dialogues as playing out of a formal process involving a set of moves and arguments. Our data opens up the possibility to characterize how the specific moves that people choose reflect their interests and expectations about reaching agreement, as well as their interests and strategies for success in the underlying domain task.

6 Acknowledgements

The authors are grateful for their support via HSD-0624191. We would also like to thank R. Lau, M. Littman, B. Sopher, D. Anderson, and J. Birchby for their involvement in this project.

References

Singjoo Choi, Douglas Gale, and Shachar Kariv. 2005. Behavioral aspects of learning in social networks: An experimental study. *Advances in Applied Microeconomics*, 13.

Barbara DiEugenio, Pamela W. Jordan, Richmond H. Thomason, and Johanna D. Moore. 2000. The agreement process: an empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human Computer Studies*.

D. Lewis. 1969. *Convention. A Philosophical Study*.

Y. Shoham and K. Leyton-Brown. 2009. *Multiagent Systems. Algorithmic, Game-Theoretic, and Logical Foundations*.

Joseph T. Wang, Michael Spezio, and Colin F. Camerer. 2009. Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth-telling and deception in sender-receiver games. *American Economic Review*.

ASPECTS OF SEMANTICS AND PRAGMATICS OF DIALOGUE

This book contains papers presented at PozDial, the 14th edition of SemDial, which took place in Poznań, Poland, 16-18 June, 2010.

The SemDial Workshops aim at bringing together researchers working on the semantics and pragmatics of dialogue in fields such as formal semantics and pragmatics, artificial intelligence, computational linguistics, psychology, and neural science.

SemDial Workshop Series homepage: http://www.illc.uva.nl/semdial/

9 788393 091508