

**TWENTE WORKSHOP
on LANGUAGE TECHNOLOGY**

TWLT 13

Formal Semantics and Pragmatics of Dialogue

Joris Hulstijn and Anton Nijholt (eds.)

Twendial'98

**Formal Semantics and Pragmatics
of Dialogue**

Proceedings of the thirteenth
Twente Workshop on Language Technology

Joris Hulstijn and Anton Nijholt (eds.)

CIP GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Hulstijn, J., Nijholt, A.

Formal Semantics and Pragmatics of Dialogue

Proceedings Twente Workshop on Language Technology 13 (Twendial'98)

J. Hulstijn, A. Nijholt

Enschede, Universiteit Twente, Faculteit Informatica

ISSN 0929-0672

trefwoorden: dialogue systems, natural language processing, speech, dialogue management,
language engineering

© Copyright 1998; Universiteit Twente, Enschede

Book orders:

Ms Hoogvliet

University of Twente

Dept. of Computer Science

PO Box 217

NL 7500 AE Enschede

tel: +31 53 4893680

fax: +31 53 4893503

Email: hoogvlie@cs.utwente.nl

Druk- en bindwerk: Reprografie U.T. Service Centrum, Enschede

Preface

TWLT is an acronym of Twente Workshop(s) on Language Technology. These workshops on natural language theory and technology are organised by Project Parlevink (sometimes with the help of others), a language theory and technology project conducted at the Department of Computer Science of the University of Twente, Enschede, The Netherlands. Each workshop has proceedings containing the papers that were presented. For the contents of these proceedings consult the last pages of this volume.

Previous workshops.

- TWLT1, *Tomita's Algorithm: Extensions and Applications*. 22 March, 1991.
- TWLT2, *Linguistic Engineering: Tools and Products*. 20 November, 1991.
- TWLT3, *Connectionism and Natural Language Processing*. 12 and 13 May 1992.
- TWLT4, *Pragmatics in Language Technology*. 23 September, 1992.
- TWLT5, *Natural Language Interfaces*. 3 and 4 June, 1993.
- TWLT6, *Natural Language Parsing*. 16 and 17 December, 1993.
- TWLT7, *Computer Assisted Language Learning*. 16 and 17 June 1994.
- TWLT8, *Speech and Language Engineering*. 1 and 2 December 1994.
- TWLT9, *Corpus-based Approaches to Dialogue Modelling*. 9 June, 1995
- TWLT10, *Algebraic Methods in Language Processing*. 6-8 December, 1995
- TWLT11, *Dialogue Management in Natural Language Systems*. 19-21 June, 1996
- TWLT12, *Automatic Interpretation and Generation of Verbal Humor*. 11-14 Sept. 1996

TWLT 13, the present workshop, has been organized by a committee consisting of R. op den Akker (Twente), N. Asher (Austin), P. Dekker (Amsterdam), G. Jger (Pennsylvania), J. Hulstijn (Twente), G.-J. Kruijff (Prague), J. Kuper (Twente), A. Nijholt (Twente), S. Pulman (Cambridge), G. Veldhuijzen van Zanten (Eindhoven) and H. Zeevat (Amsterdam). The workshop is not only 13th TWLT, but also the follow-up of Mundial'97, a workshop on the same topics, held in Munich, 1997. For that reason the current workshop in Twente has also been given the name Twendial'98. As in Munich, this workshop aims at bringing together researchers from various backgrounds that apply formal methods to the study of dialogue phenomena. Therefore contributions were invited on, among others, DRT, situation theory, dynamic semantics, agent theories, database theory and game theory. One aim of the workshop was to make an attempt to bridge the widely recognised gap between theory and practice in the design of dialogue systems. In addition to four invited speakers

Nicholas Asher (University of Texas at Austin)
Jonathan Ginzburg (Hebrew University, Jerusalem)
Steve Pulman (Cambridge SRI)
Henk Zeevat (University of Amsterdam)

We asked Marc Blasband, project coordinator for European language engineering projects in which the Nederlandse Spoorwegen (Dutch Railways) is involved, to give a talk about his ideas on modelling dialogues. A short paper explaining his ideas is added at the end of this proceedings.

A workshop is the concerted action of many people. We are grateful to the authors and the organizations they represent, for their efforts and contributions. In addition we would like to mention here the people whose efforts have been less visible during the workshop proper, but whose contribution was evidently of crucial importance. Charlotte Byron and Alice Hoogvliet-Haverkate took (as ever) care of the administrative tasks (registration, hotel reservations, etc.). The editors are most grateful to the members of the program committee for their support in the refereeing process.

Joris Hulstijn and Anton Nijholt

May, 1998

Contents

Invited Papers

- Varieties of Discourse Structure in Dialogue* 1.
Nicholas Asher (University of Texas at Austin)
- Clarifying Utterances* 11.
Jonathan Ginzburg (Hebrew University, Jerusalem)
- The TRINDI Project: Some Preliminary Themes* 31.
Steve Pulman (Cambridge SRI)
- Contracts in the Common Ground* 41.
Henk Zeevat (ILLC/University of Amsterdam)

Accepted Papers

- Uncertain Reasoning About Agents' Beliefs and Reasoning, with special attention to Metaphorical Mental State Reports* 49.
John Barnden (University of Birmingham)
- Eye-movement Research and the Investigation of Dialogue Structure* 61.
Thomas Clermont, Marc Pomplun, Elke Prestin and Hannes Rieser (Universität Bielefeld)
- Mixing Situation Theory and Type Theory to Formalize Information States in Dialogue Exchanges* 77.
Robin Cooper (Göteborg University)
- The Interplay of Desire and Necessity in Dialogue* 89.
Jean-louis Dessalles (ENST, Paris)
- Automatic Evaluation of Dialogue Systems* 99.
Wieland Eckert (AT&T)
- Some Remarks on Distributed Knowledge* 111.
Jelle Gerbrandy (ILLC/University of Amsterdam)
- Questions in Update Semantics* 125.
Jeroen Groenendijk (ILLC/University of Amsterdam)
- Theory of Mutuality (Syntactic Skeleton)* 139.
Wolfgang Heydrich (Universität Hamburg / Universität Bielefeld)
- A DRT-style Modelling of Agents' Mental States in Discourse* 149.
Wolfgang Heydrich, Peter Kühnlein and Hannes Rieser (Universität Bielefeld)
- Questions Under Discussion and Dialogue Moves* 163.
Staffan Larsson (Göteborg University)
- Formal Design, Verification and Simulation of Multi-Modal Dialogues* 173.
Ian Lewin (SRI International)
- A Generic framework for Dialogue Game Implementation* 185.
Nicolas Maudet and Fabrice Evrard (IRIT-ENSEEIH, Toulouse)

<i>MALBOT: An Intelligent Dialogue Model using User Modeling</i>	199.
Soo-Jun Park, Keon-Hoe Cha, Won-Kyung Sung, Do Gyu Song, Hyun-A Lee, Jay Duke Park, Dong-In Park (SERI, Korea) and Jörg Höhle (GMD/FIT Birlinghoven)	
<i>Towards an Axiomatization of Dialogue Acts</i>	207.
Massimo Poesio (University of Edinburgh) and David Traum (UMIACS, University of Maryland)	
<i>Making DRT Suitable for the Description of Information Exchange in a Dialogue</i>	223.
Mieke Rats (formerly Delft University)	
<i>Modal subordination in Questions</i>	237.
Robert van Rooy (ILLC/University of Amsterdam)	
<i>A Discourse Theory of Manner and Factive Adverbial Modification</i>	249.
Adam Zachary Wyner (Bar Ilan University)	

Other Contributions

<i>A Simple Semantic Model</i>	269.
Marc Blasband (Nederlandse Spoorwegen)	

Sponsors and Support

We gratefully acknowledge help from:

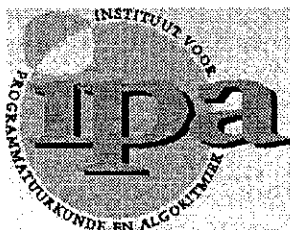
NWO-Geesteswetenschappen



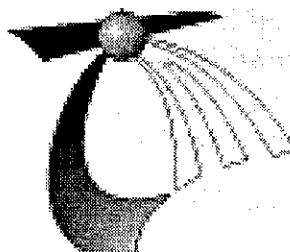
Nederlandse Spoorwegen



Instituut voor Programmatuurkunde en Algorithmeek



Centre for Telematics and Information Technology



University of Twente, Parlevink Research Group



Invited Papers

Varieties of Discourse Structure in Dialogue

Nicholas Asher*

Department of Philosophy,
University of Texas at Austin,
Austin, Texas 78712, USA
nasher@bertie.la.utexas.edu

Abstract

In this paper I examine a variety of discourse structures used in the analysis of dialogue. I propose a way to integrate turn taking structures, question stacks and structures based on rhetorical function, as well as more cognitively based structures like coordination, by distinguishing two tasks, the task of building the discourse structure itself and the task of cognitive modelling.

1 Introduction

In this talk I will examine a variety of discourse structures that we find in dialogue. Structures like conversational turns, question stacks, or those generated by rhetorical function serve as the basis of an analysis of discourse phenomena like the truth conditional effects of discourse structure like temporal and pronominal anaphora, or effects on the pattern of discourse itself, the possible conversational moves. Other theorists of discourse, notably Clark (1992), argue that the underlying explanatory structures for discourse are cognitive and propose a model of coordination of cognitive states as the basic explanatory structure for dialogue phenomena. Each one of these ideas is attractive, but we would like some way of integrating them together.

I propose to examine turns, stacks and coordination from the perspective of a theory of discourse, SDRT, in which discourse structure is founded on the rhetorical functions of constituents. SDRT will enable us to integrate structure based on rhetorical function and function based on conversational turns together. In many dialogues at least we can see the rhetorical structure as refining and giving us a typology of particular types of turns as well as giving us the internal structure of each participant's turn. For explicit questions we can also refine the stack model with rhetorical relations (the implicit questions for discourse segments appealed to by van Kuppelvelt and Roberts I am less sure; but it seems to me that when the task of finding out what the implicit questions are is made more precise, it will look very much like the glue logic used to infer rhetorical relations in SDRT).

In dialogue participants not only construct discourse structures for what is said; typically they construct a perhaps very partial model of participant's cognitive states as well. We can find clues of cognitive modelling in many divergent research programs— in Grice, Sacks, Grosz and Sidner as well as others in the AI tradition, Clark, and many others. As Grice pointed out long ago, part of what makes dialogue so interesting is that there is a subtle and intricate interaction between recovering the intended interpretation of what a dialogue participant says and modelling the beliefs, desires and intentions that underly motivate what the participant said. The bottom up incremental process of SDRS construction allows us to make quite precise several types of interactions between cognitive modelling and discourse structure and explain coordination phenomena.

*Thanks to Myriam Bras, Joan Busquets and Laure Vieu for discussions on dialogue especially with the procope text.

2 More Details on the SDRT Model

In SDRT, we distinguish several levels of interpretation of dialogue. The first has to do with compositional and lexical semantics. This may be drastically underspecified and may yield only fragmentary logical forms when the latter are understood in the usual way (Reyle 1993, Asher and Fernando 1997). The second level attempts to build a coherent and connected SDR discourse structure, a complete information packaging for the content conveyed by a discourse using various pragmatic principles, as well as lexical and compositional semantics.¹ This is the level that SDRT has concerned itself with. We postulate that each participant interprets the discourse in his or her own way constructing an SDRS that can give rise to misunderstandings (explored in Asher and Lascarides, in press). The third level of interpretation involves how one discourse participant reasons about the mental states of another discourse participant given what they have said. At the SDRT level, the interpreter reconstructs the content of the message; at the third, he reconstructs why the speaker said what he did and thus the cognitive-state lying behind the speech act. The third level is one of cognitive modelling. (There's in fact a fourth dynamic process which involves the agent's own beliefs goals and plans and how they evolve as the dialogue progresses).

The second, third and fourth levels of interpretation involve very different resources and inference schemes, and they're designed to do different things. SDRS's are designed to analyze and predict discourse coherence incoherence and related semantic phenomena. Cognitive modelling is an attempt to reconstruct the speaker's mental state and discourse intentions. While SDRSs have to be constructed in order to interpret discourse, cognitive modelling is often not needed; it's a much more complex sort of reasoning than SDRS construction and more uncertain in its conclusions. Information about cognitive states is thus often extremely fragmentary and incomplete.

Nevertheless, cognitive modelling can in several cases be useful to discourse interpretation, and there are lots of interactions between the SDRS level and the cognitive modelling level that I will explore below. The first sort of interaction will show how we can exploit general constraints about cognitive models to determine discourse relations in dialogue and more importantly discourse expectations. Indeed in many forms of conversation cognitive modelling and even belief revision can be crucial to success.

I'll go through each one of these levels in turn.

Discourse Structure for Dialogue

Dialogue is a generalization of monologue from the discourse structure point of view. Because dialogue involves at least two agents, there is a potential for disagreement and conflict in a way that there isn't for monologue (unless the author has split personalities). As argued in Asher and Fernando (1997) at Mundial, this requires a certain intensional interpretation of discourse and it requires us to countenance discourse relations the truth of whose propositional terms is problematic. Examples of such relations are Correction and Counterevidence below:

- (1) a. A: John distributed the copies.
b. B: No, it was Sue who distributed the copies.
- (2) a. A: John went to jail. He was caught embezzling funds from the pension plan.
b. B: Yes, John went to jail, but he did so because he was convicted of tax evasion.
- (3) a. A: Smith shot the guard at the bank.
b. B: He has witnesses that say he was out of town at the time of the robbery.

Nevertheless, the discourse relations present in monologue are also present in dialogue (though as some of the discussions by Groenendijk, Stokhof and Veltman would suggest, there may be additional conditions, say, for Narration in multilogue).

¹There may be also underspecification at this level. See Asher and Fernando (1997).

- (4) a. A: Jones got kicked out of school.
 b. B: He was caught buying liquor.
 c. C: That was really dumb of him.

In (4), B provides an explanation for the event A described, while C offers a Commentary. These are relations that are often observed in monologue (Asher 1993). Other relations like Background, Result are also easily observed in dialogue.

Corrections and Counterevidence are not the only new types of discourse relations in dialogue. We also have relations between questions and answers, questions and questions (one question may elaborate upon another), requests and questions (a question may elaborate upon a request too) and assertions that stand in what I call elsewhere "metatalk" relations explaining why a particular speech act has or is about to be made. Here's an example of a meta-talk relation:

- (5) I'm cold. Please close the window.

Many of these relations are derived from an abstract model of speaker rationality that we build with our cognitive modelling tools (for details see Asher 1998b), and with them are associated speech act related goals or SARGs. For instance, a question implies by default that the questioner wants to know the answer to it. While a lengthy discussion of the axioms and constraints for these rules is beyond the scope of this paper, I list the new relations I will use in dialogue analysis here (the first two are explained in Asher and Lascarides 1998, as well as in the Asher and Lascarides contribution to Mundial97):

- IQAP: Indirect question answer pair. $IQAP(\alpha, \beta)$ holds if β together with other information the questioner has access to allows the questioner to derive a correct answer to his question. Thus indirect question answer pairs subsume direct answers.
- NEI: Not enough Information. $NEI(\alpha, \beta)$ holds if β does not supply enough information to satisfy the associated SARG of α . For instance, if α is a question then by default its SARG is that the questioner wants to know the answer; β may be a response that doesn't yield a correct answer, in which case $NEI(\alpha, \beta)$.
- Q-elab: Question Elaboration. Where α is either a question or request, Q-elab(α, β) holds if β is a question whose answer specifies an essential part of a plan to bring about the SARG of α , which in the case of a question is knowing the answer and in the case of a request is to have the request carried out.
- Explanation* and Result*. Explanation*(α, β) if β explains why the speech act α was performed. Result* holds if the first argument explains the speech act in the second.
- Ack: Acknowledge. Ack(α, β) that the speaker of β accepts α .

I'll assume concerning questions and their answers, that if something is an answer to the question, then it satisfies the SARG of the questioner and this closes off such a segment. Such a closed segment in SDRT is a constituent containing the question and its answer.

Cognitive Modelling

To do cognitive modelling, we need to make some decisions about how to think about agents' cognitive states and their relations to what they say. In earlier papers on belief,² Kamp and I advocated a principle of the unity of thought and information; the same structures that are thought to be useful for characterizing information—be they sets of possible worlds, dynamic information states (sets of world assignment pairs) or more structured objects like DRSs themselves—should also be used to characterize the objects of attitudes like belief and desire. From the SDRT perspective information in discourse is conveyed by an SDRS. So by this principle of unity, we

²See Kamp (1990) or Asher (1989)

might conceive that attitudinal objects involve not just one DRS or some unrelated set of DRSs but rather something more like an SDRS, an interrelated structure of propositions.

The unity of thought and information thus leads us to the conclusion that the logic of the content of attitudinal states is no less complex than the logic of information content itself. But cognitive modelling is uncertain and so reasoning about agents' cognitive states will involve defaults, whose semantics in turn requires consistency tests. If cognitive modelling is to be a task that we can effectively engage in, however, then we will for the same reasons as in building SDRSs have to resort to a shallower, more superficial logic—something like the logic of information packaging. Our cognitive modelling will then take place in a quantifier free language. We'll think of cognitive states as atoms on which we'll write constraints like $\text{Believes}(\phi)(c)$; this means that c contains within it the belief that ϕ . Just as for reasoning about discourse structure, we have defeasible rules, stated with $>$, for reasoning about agents' cognitive states; for instance, Grice's maxim of quality can be used to reason about speakers' cognitive states and is naturally formulated as $\text{Say}(a, \phi) > \text{Believe}(\phi)(c_a)$. The language in which we will reason about beliefs, intentions and desires or goals will be a modal propositional language. We can write update rules for cognitive state representations in such a language, as well as rules about SARGs like the following from Asher and Lascarides (1998): $(A : ?\alpha \wedge \text{QAP}(\alpha, \beta)) > \text{Want}_A \text{Know}(A, \beta)$.

Perhaps we use the same system to update our own beliefs, intentions and desires upon receiving new information. But it may be that we access more of the content to do the updating in our own case. Much more research is needed to deal with this issue in any case.

3 Turn Taking and Rhetorical Function

Because there are two agents in dialogue, they typically have turns. But these turns come in various types, which is what the rhetorical relations of SDRT capture. To see a particular interaction between turn taking and rhetorical structure, let's look at an analysis of the following French dialogue. I have numbered the distinct speech acts that the SDRT analysis will take account of.

- (6) a. A: Je suis tombé en panne (1a). Peux-tu m'aider? (1b) (I have had a breakdown. Can you help?)
- b. B: Où es-tu?(2) (Where are you?)
- c. A: Je suis devant le refuge qui se trouve à environ un km après Couiza(3a). Il y a la une cabine téléphonique (3b). (I am in front of the refuge that is about 1 km after Couiza. There is there a telephone booth.)
- d. B: Il y a plusieurs refuges aux alentours de Couiza (4a). Dans quelle direction es-tu parti de Couiza? (4b) (There are several refuges around Couiza. In which direction did you leave Couiza?)
- e. A: Je suis sorti par la route Paul Sabatier (5a). Puis j'ai roulé vers la montagne (5a). A une clairière j'ai tourne a droite (5c). (I left by the Paul Sabatier road. I then drove toward the mountains. At a clearing I turned left.)
- f. B: Au grand carrefour? (6) (at the big intersection)
- g. A: Non (7a), après, là ou on commence à avoir une belle vue sur la mer (7b). (no afterwards, where you begin to have a nice view of the sea.)
- h. B: Ah, je vois, au Rocher du diable (8). (Oh I see at the Devil's cliff.)
- i. A: C'est possible, il y avait un gros rocher (9). (It's possible; there was a big cliff.)
- j. B: Donc tu es à la Maison de l'aigle (10). J'arrive tout de suite (11). (So, you are at the house of the eagle. I am coming right away.)

The dialogue gives rise to the following SDRS "skeleton". The lines represent elementhood and the discourse relations between constituents are over on the side.

In the incremental construction of this SDRS, there is an interesting transition of the sort discussed in Asher and Lascarides (1998) that occurs at (6bc). A thinks that he has given a direct

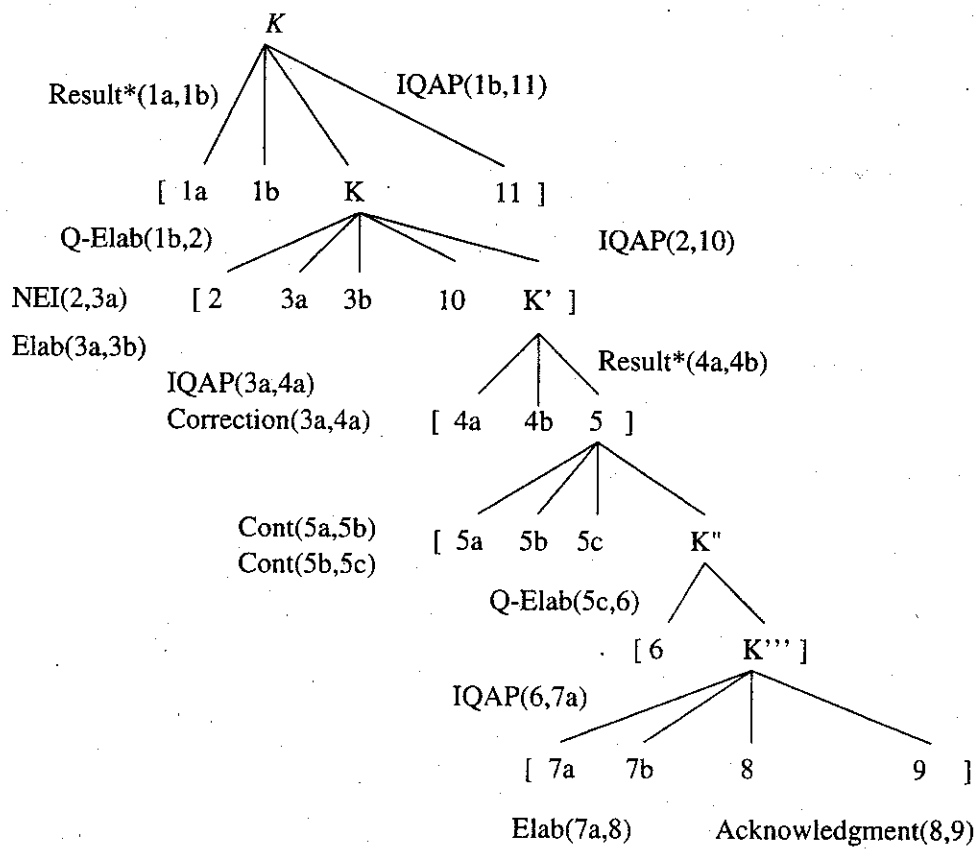


Figure 1: SDRS skeleton of example (6).

answer to the question where he is, but after the Correction he and B both have SDRSs in which this is not so. The relation that holds is the relation NEI (not enough information) saying that the response does not suffice to give a correct answer to the question and thus satisfy the questioner's SARG.

In this SDRS you notice no violation of the SDRS availability constraints. Some speech act constituents are grouped together because they represent closed off constituents containing questions and their answers, which may take several turns to establish. B answers his own question posed in 2 only at speech act 10. In so doing, B pops up past several questions; the assumption is that they are all answered correctly as far as he is concerned. Notice also that these larger constituents are linked to others via one of their elements (kind of like *de re attitudes*). Nevertheless, as the SDRS is built up incrementally, all the constraints on attachment are met. Turn transitions instantiate one or more discourse relations between some constituent and an available attachment site. Notice also that there are sometimes relations within each participant's contribution to the turns. The rhetorical structure tells us what's going on in each turn and the SDRS that results has significant explanatory effects upon several matters. For one, we are able to exploit the fact that B answers his own question to predict that he will now go on to answer A's question that his own elaborated. So we predict given the high attachment of 10 an even higher attachment of 11 using SDRT. The turns themselves don't determine much about SDRS structure at least not the notion of availability here. The question-elaboration relation is crucial however for determining the vertical nature of the SDRS by subordinating one question answer pair turn to another. This subordination is also semantically significant in that the sequence of Q-elabs determines a strategy or plan that B follows in order to come to know where A is.

Is this neat patterning always correct? Consider the following example.

- (7) a. A: There was this guy. He came to the sessions. He never said anything. Then one day he shows up, and he starts talking, interacting.
b. B: Why didn't he say anything before?
c. A: Dunno. Shy maybe. But anyway he's yammerin away and telling these jokes...

It looks as though sometimes availability might be violated. If we analyze this from an SDRT perspective, however, it's actually unclear what the question attaches to. It isn't after all really anaphoric, and so it could in principle attach to A's whole story (with NEI). To test whether you can really attach inside, we have to use something that's sensitive to availability like a pronoun:

- (8) a. A: While we [A and C], we got talking to these three people.
b. B: Uh huh.
c. A: The first guy was a nut. The second was a lady. She was boring. The third guy was a delight, a high school teacher taking kids to Europe.
d. C: ? What are you saying? She was great!

(8) isn't very good. To process the pronoun would require an attachment to an unavailable constituent. So if (8) is bad, then this means that our availability constraint holds in dialogue. In particular, compare (8) to (9) which uses a *convergent* discourse relation to attach.

- (9) a. A: While we [A and C], we got talking to these three people.
b. B: Uh huh.
c. A: The first guy was a nut. The second was a lady. She was OK. The third guy was a delight, a high school teacher taking kids to Europe.
d. C: ?Yeah. She was selling machinery in Hungary.

I find both examples much improved if we use *the lady* and not the pronoun. But one of the predictions of SDRT is that definites to be anaphorically linked do not require an antecedent in an available constituent (because of presuppositions). So it seems as though the right way to think of the turn structure is that its segmentation is typically refined by the structure of rhetorical function. I should note that of course there can be constituents that are made up from several turns. Here is an example from Sacks that would naturally receive such an analysis:

- (10) a. Joe: We were having an automobile discussion
 b. Henry: discussing the psychological motives for
 c. Mel: drag racing in the streets.

Before leaving the subject of turns, one might propose the following constraint (from Sachs's work but adapted to SDRT's use of hierarchical structure) for cooperative discourse.

– **Turn Constraint**

- If the speaker is within a given turn *t*, his contribution should normally attach locally to some previous contribution in *t*. If he is initiating a turn, his contribution may attach to any available constituent.

It would be easy enough to formalize this constraint in the SDRT glue logic DICE as a constraint on attachment, something which is very underconstrained in SDRT at the moment. This is a principle that favors local attachment. We'll see how this fares in the next section.

4 Stacks of questions

Another structure for dialogue is the generalized stack structure that Ginzburg has made seem quite intuitive (but see also Roberts ()). The idea roughly is that questions are given as a partial order in a text. In the stricter stack architectures, you would have to answer the last one in first before you can answer others, unless it's going to be ignored. This doesn't seem right, as the following seems perfectly acceptable.

- (11) a. A: Where were you on the 15th?
 b. B: Why?
 c. A: Do you remember talking to anyone right after the incident?
 d. B: I was at home. I didn't talk to anyone after the incident.
 e. A: What do remember about the incident?

(11) is also a violation of our default turn constraint on attachment. The principle that overrides the default attachment is something like this: when there are several unanswered questions for the speaker, he may use his turn by answering them all in any order.

In SDRT Q-elab does much of the work of the stack of questions; it subordinates one question to another. But with Q-elab, a high attachment will predict in SDRT that the subordinate question is either already answered or moot. Suppose that A and B are both talking on the phone to C who is calling for help.

- (12) a. A: Where are you?
 b. C: Let me see. I'm in a village, uh,
 c. B: What road did you take leaving Toulouse?
 d. C: I'm in Couiza.

When C answers A's question, B's Q-elab question becomes moot.

Now Q-elab doesn't always hold. It holds only if answers to the second question make some steps toward achieving the SARG of the superordinate question. In (11), it's hard to see how the questions could so be related. In any case, if availability is correct, then if Q-elab were the relation between B's two questions, A's responding to the higher question first would mean that the second *couldn't* be answered. We would predict incorrectly that (11) is incoherent.

Questions that aren't in a Q-elaboration relation to each other, then rather form a set of available attachment points, which the respondent may answer all at once in his turn. They are in effect what Ginzburg has called coordinate questions. So if this right, then in (11), we have to coordinate the unanswered questions (13a,c) together into a constituent which is then answered

with a complex constituent, consisting of two simple constituents linked by the SDRS relation of Continuation.

There are limits as to how to take turns with questions, because sometimes we have to use explicit discourse markers like *Let me come back to the question about the fifteenth* to reopen a previously unavailable attachment point. Consider the following deposition like conversation.

- (13) a. A: Where were you on the 15th?
b. B: Why?
c. A: Do you remember talking to anyone after the incident?
d. B: ??I was at home.
e. A: What do remember about the incident?
f. B: ?? I didn't talk to anyone after the incident

In (13c), B doesn't reply to all the open questions at the turn when he replies to one of them. He systematically stays one out of turn. This systematic asynchronization strikes one as bizarre even though it's perfectly understandable. Whether we take the unanswered questions to stand in a Q-elab or to be coordinated, our constraints predict this discourse to be unacceptable.

5 Cognitive Coordination

Many if not most studies of dialogue follow the following intuition: speakers cooperate in various ways to exchange information. This view is at the foundation of modern pragmatics with Grice's work. Grosz's dialogues are examples of cooperative dialogues. SDRT has also subscribed to such a view (see Asher 1998b, Morreau 1995).

Clark and other psychologists like Simon Garrod have put forward what seems to be a competing theory of conversation. Conversation isn't about the cooperative exchange of information; it's about coordinating mental states. What coordination of mental states means seems a little metaphorical, but Garrod's papers give clear examples of one type of coordination. In Garrod's work two participants have to negotiate a maze together. One has to move a person through the maze; the other has information about where to go. To make use of that information, the first has to get the second to understand where his person is in the maze. This, Garrod says, is a matter of coordination, and he shows convincingly how the goal of coordination can drive not only discourse moves but also get speakers to adopt particular cognitive categories and linguistic conventions.

Nevertheless, coordination is still a bit mysterious. What is it? From the present perspective, it is in effect the SARG of a WH-question. The coordination examples drawn from Garrod are driven by the goal of knowing where, though one also might imagine coordination tasks driven by goals of knowing who or knowing what.

To make this more precise, let's look again at (6) again, where we see an example of such coordination.

- (6) Ou es tu? (Where are you?)

is a Q-elab in which we get in (6b). It is B's first step to figure out how to help A, since to help a person with a physical problem like a breakdown, you have to know where he or she is. Knowing where is like knowing who and demands an internal anchor for the location denoted by the NP that replaces the wh-phrase. The SARG behind this question now gives us the goal for B of knowing where A is.

Now what would it mean for B to come to know where A is? It means that he has to identify the location of A with some location with which he is already familiar and could in principle identify. In dynamic semantic accounts of attitudes, this amounts to providing the variable associated with A's location with an *internal anchor* linking it to some other variable with which enough information is given so as to make it familiar and identifiable to B. Now in the 10th information unit in (6j), we see an interesting fact; B supplies the direct answer to his own question that had as its goal knowing where A was. This is the entire coordination task, and B's answer signals

that he now has coordinated A's location within his understanding of space. In other words, he has supplied an internal anchor for the variable representing A's location. This sort of answering of one's own question for a WH question should thus signal by default a particular constraint on cognitive models—namely, the agent has been able to supply an internal anchor for the appropriate variable.

Thus, coordination phenomena of this kind do not constitute an alternative way of thinking about dialogue but fall squarely within the kind of framework elaborated here. More research, however, will be needed to see whether other coordination phenomena are as tractable. It may be that cognitive coordination at other levels may be a matter of ensuring that one's discourse representations or cognitive models are suitably similar.

6 References

References

- Nicholas Asher (1998a). The Logical Foundations of Discourse Structure, *Logic Colloquium 96*, Springer Verlag.
- Nicholas Asher (1998b). Discourse Structure and the Logic of Conversation. Forthcoming in the first volume of *Current Research in the Semantics Pragmatics Interface*.
- Nicholas Asher and Tim Fernando (1997). Representations with Effective Labelling for Disambiguation, *The Second International Workshop on Computational Semantics*, Tilburg, the Netherlands.
- Nicholas Asher and Alex Lascarides (1998). Questions in Dialogue, *Linguistics and Philosophy* 21(2)
- Herbert Clark (1992). *Using Language*, CSLI Publications, University of Chicago Press.
- Simon Garrod and Gwyneth Doherty (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions, *Cognition* 53, 1994, pp. 181-215
- Jonathan Ginzburg (1995) Resolving Questions parts I and II, *Linguistics and Philosophy*, 18(5) and 18(6).
- Paul Grice (1975). Logic and Conversation, in P. Cole and J. Morgan (eds.), *Syntax and Semantics*, 3.
- Hans Kamp (1990) Propositional Attitudes, in C. Anthony Anderson, Joseph Owens (eds.), *The Role of Content in Logic, Language, and Mind*, CSLI Lecture Notes 20, University of Chicago Press.
- Michael Morreau (1995). How to Derive Conveyed Meanings, *Proceedings of the Edinburgh Computational Semantics Workshop*, Edinburgh, UK.
- Craige Roberts (1996). Information Structure, Plans and Implicature, AAAI Spring Symposium on Computational Implicature, Stanford CA.
- Sacks, H. (1992). *Lectures on Conversation* vols I and II. Oxford: Blackwell Publishers.

Clarifying Utterances*

Jonathan Ginzburg
Linguistics Programme
Hebrew University of Jerusalem
Mount Scopus
Jerusalem 91905
Israel
msjihad@mscc.huji.ac.il

Abstract

The paper argues for the importance of utterances as basic units in dialogue, as opposed to “sentences in context”. The main case study is *clarification*: a characterization is provided within KOS, a recently developed synthesis of situation semantics and dialogue games, of the clarifications that can follow up on an utterance. This involves outlining: (a) a theory of understanding for dialogue, (b) a semantics for clarification acts, and (c) an account of how utterances update contexts. I will show that a relational view of meaning, whose original motivation was purely logical, can provide an important component for a theory of utterance understanding and utterance updates.

1 Introduction

“So here’s a question for you. How old did you say you were?” Alice made a short calculation, and said “Seven years and six months.” “Wrong!” Humpty Dumpty exclaimed triumphantly. “You never said a word like it!” “I thought you meant ‘How old *are* you?’ ” Alice explained. “If I’d meant that, I’d have said it,” said Humpty Dumpty.

Although some of Humpty Dumpty’s views on semantics need to be taken with a grain of salt, the distinction he draws between whether something is *true* and whether it is true that it has been *said* is, as I will try to demonstrate, an important one for the semantics of dialogue. In standard logical frameworks this distinction is, for the most part, abstracted away from: this is because the various basic semantic notions (truth or update conditions) are defined for *sentences*. Thus, despite the fact that it is by now universally recognized that it is only *relative to a context* that truth is to be evaluated or an update performed, the distinction between a sentence *S* (in context) and an utterance *u* of *S* remains, more often than not, unheeded—and with good reason, or so have claimed certain philosophers and linguists, most prominently Kaplan (1989).

In domains where perfect communication obtains, the distinction might possibly be a rather finicky one. In this paper I will, however, attempt to offer some grounds why, at least as far as the semantics of dialogue is concerned, the utterance-based perspective is necessary. My main case study will be the issue of clarification in dialogue, more specifically—to what extent can one offer a characterization of the range of possible clarifications that can coherently follow a given utterance act, exemplified in (1):

*I would like to thank Robin Cooper and Massimo Poesio for discussion on issues related to this paper. This paper summarizes a number of chapters from a forthcoming work tentatively titled *Questions and the Semantics of Interaction in Dialogue*. All are available by ftp from ftp.cogsci.ed.ac.uk/pub/ginzburg

- (1) a. A: Did Jill leave?
 b₁. B: Who is this person you're asking about?
 b₂. B: Did Jill leave WHEN?
 b₃. B: Why do you ask?

I will offer one such characterization within KOS, a synthesis of situation semantics and dialogue games which I have recently been developing. The characterization requires three basic components: (a) a theory of meaning and understanding for dialogue, (b) a semantics for clarification acts, and (c) an account of how utterances update contexts—we want to tie in the fact that an utterance with a certain meaning has taken place with its understanding conditions and with a certain class of clarification acts. The theory of meaning I will take on board is essentially a rather traditional one, originally due to Kaplan, subsequently refined in the situation semantics literature—essentially that meanings are functions or n -ary λ -abstracts which arise by abstracting from a content those parameters which “vary with context” (prototypically indexicals). The content of a given utterance u of a sentence S with meaning $\mu(x_1, \dots, x_n)$ arises by fixing values for x_1, \dots, x_n in u . I will show that such a theory of meaning whose original motivation was purely logical—explicating the different modal status of sentences containing indexicals—can provide an important component for a theory of utterance understanding and utterance updates. The latter relies crucially on taking utterances as basic semantic units as a subspecies of events in fact, an emphasis originating in situation semantics and recently articulated in some detail in Poesio (1998). All this combines with a semantics for *reprise* sentences (exemplified by (1b₂) above) developed in Ginzburg and Sag (to appear), to yield a characterization of the clarification space of an utterance.

The paper is structured as follows: section 2 provides some background about KOS. In sections 3 and 4, I sketch accounts of utterance understanding and updates. In section 5, I consider the contextual development that takes place when an utterance is not fully understood. Section 6 provides a bare bones semantics for reprises sentences. Section 7 puts the various pieces together to provide a characterization of the space of possible clarifications, whereas section 8 provides a brief example of a dialogue analyzed.

2 Semantics of Dialogue: Background Notions

In this section I provide a brief introduction to KOS, a version of Situation Semantics enriched to analyze conversation—more detailed descriptions of this and related issues are available in Ginzburg (1995a, 1997, to appear).

The DGB

In Ginzburg (1997), I argued that in dialogue *semantic* mismatches occur: at certain junctures there is no single context—different semantic objects (e.g. for use in anaphora and ellipsis resolution) are available to distinct conversational participants (CP's); this theme will reemerge in section 5. The import of this is that we need a framework in which *inter alia* one and the same semantic input can affect distinct participants differently at a given point. This, and *inter alia* the fact that certain information needed to make sense of contributions is *unpublicized* (cf. discussion in Asher 1993, Ginzburg 1995a) that truth conditions of attitude reports are parametrized by agent-specific parameters such as goals and inferential capabilities) suggests that the “loci of conversational update”, the basic units of semantic description, need to be the mental states of each individual participant (their UNPUB(licized)-MS's). However, it would be undesirable to regress to solipsism; there is much evidence, both semantic and psycholinguistic, that CP's try to maintain a common view of the conversation and its background. This, along with other important insights, is captured by Stalnaker-inspired presupposition theory. Hence, I posit, inspired by work in the tradition of dialogue games (e.g. Hamblin 1970, Carlson 1983, Houghton and Isard 1986), that conversational rules involve updates by each CP of her own *dialogue-gameboard* (DGB), a

quasi-public informational repository (cf. Hamblin's *individual commitment slate*); certain actions are, nonetheless, parametrized by unpublicized information.

The structure of the DGB

My main concern here will be on the nature of the DGB. Moreover, I will treat a DGB as a "semantic entity", that is, as individuated in terms of semantic objects such as facts, questions, and propositions. As a starting point, I assume that the DGB is structured by at least the following attributes:

FACTS: set of commonly agreed upon facts. This component in a number of guises is a traditional one in many approaches to context. For most purposes in the current paper, it will be sufficient to take a simple view of this attribute: a set of SOA's closed (cf. Asher 1993) under meets and joins;¹

QUD ('questions under discussion') : a set that specifies the currently discussable questions, partially ordered by \prec ('takes conversational precedence'). If q is maximal in QUD, it is permissible to provide any information specific to q using (optionally) a short-answer.

L-M ('latest-move'): content of *latest move* made: it is permissible to make whatever moves are available as reactions to the latest move. In this initial section the value of this attribute is rather straightforward: it holds of the information characterizing the content of the most recent dialogue move, for instance 'A asserted that p ', 'A asked q ' etc. A more complex view of this attribute will be offered in section 4.

Querying and Assertion

Viewing querying and assertion as interaction requires us to specify the effect they have on the querier/asserter and the possible (conventionalized) reactions that become available to the other CP. For the moment I specify these effects assuming *perfect communication*; we will later see how to drop this assumption. Given this, KOS offers the following basic account: both querying and assertion involve a question becoming maximal in the querier/asserter's QUD, the posed question q for a query q , $p?$ for an assertion p . As for the responder, she can subsequently either choose to start a discussion (providing information σ that is ABOUT $q/p?$ or posing a question q_1 which INFLUENCES $q/p?$)² or, in the case of assertion, update her FACTS structure with σ . A CP can *downdate* $q/p?$ from QUD when, as far as her (not necessarily public) goals dictate, sufficient information has been accumulated in FACTS. There are two key concepts to isolate from the foregoing:³

¹ In fact, there are a number of considerations that suggest a more refined structure on the DGB, in particular on FACTS. In Ginzburg (to appear), inspired in part by work on fact and propositional anaphora in texts by Webber and Asher, I argue for and develop a ("TOPICAL" v. "STORED") dichotomy between two types of "presupposed informational items": the former enter into FACTS via a "short-term", defeasible repository structured by the elements of QUD. The latter are facts that either (after potentially being discussed) have been stored long-term or arise as side effects to issues that were never explicitly under discussion. Evidence for this includes locality phenomena of fact anaphora and ellipsis and the phenomenon of *hasty accommodation*: the fact that a CP can coherently accommodate something as factual and then later be required to retract. The dichotomy is useful particularly as a conceptual means of distinguishing between "highlighted" and "backgrounded" contextual information.

² 'ABOUT' is a notion of partial answerhood empirically and formally characterized, together with the agent relative notion of exhaustiveness 'RESOLVES', in Ginzburg (1995a). 'INFLUENCES' (or its converse: DEPENDS) is a sort of entailment relation between questions: q_1 INFLUENCES q_2 iff any τ which RESOLVES q_2 also RESOLVES q_1 . Since 'RESOLVES' is agent relative, I will for the most part in this paper resort to a simpler, semantically absolute definition of 'INFLUENCE' in which POTENTIALLY-RESOLVES, the semantic relation underlying 'RESOLVES', substitutes for RESOLVES. Two examples illustrating the INFLUENCES options:

A: Who attacked Marlon? B: Who was in town at the time?

A: Everyone supports the proposal. B: Does Jill support it?

³For simplicity here I'm ignoring an additional option for an adjacency pair, both for querying and assertion, which involves discussion of whether $q/p?$ should be discussed.

Definition 1 Given a question q , an utterance u is q -specific iff content(u) is either ABOUT q or INFLUENCES q .

Definition 2 If L-M = 'A ASSERTS p ', an utterance u is an adjacency pair for L-M iff content(u) is p ?-specific or content(u) is 'B accepts p '.

Definition 3 If L-M = 'A ASKS q ', an utterance u is an adjacency pair for L-M iff content(u) is q -specific.

A Semantic Application: short answers

As a means for observing KOS in action, let us consider a concrete application, the semantics of "short answers", which I assume ties in closely with QUD-maximality, the idea being (backed by some evidence from corpora) that elliptical contributions whose content is resolved to be ABOUT a given question are possible, in principle, arbitrarily far away from the turn in which the question was introduced as long as the question remains under discussion. (2) says that in a DGB configuration DGB_0 , any XP can be expanded as an S whose content is calculated as follows: it is a proposition of the form ($s!σ$). Here s is the event/situation component of the question maximal in QUD; $σ$ arises by predicating of the XP the abstract component of the question maximal in QUD:

- (2) **phrasal utterance construction type** (simplified version)
 $S \rightarrow XP$
 $\text{Content}(S)[DGB_0] = (\text{SIT}(\text{Max-QUD}) ! \lambda\text{-Abstr}(\text{Max-QUD})[\text{Content}(XP)])$

In (4) and (5) is provided a simplified KOS analysis of the dialogue in (3), in particular of the short answer in (3.5):

- (3) 1. A: Who should we invite for tomorrow?
 2. B: Who will agree to come?
 3. A: Helen and Jelle and Fran and maybe Sunil.
 4. B: I see. So,
 5. Jelle.

- (4)

1. A QUD: q_1 2. B QUD: q_1 ; asks q_2 such that q_2 INFLUENCES q_1 ; B QUD: $q_1 \prec q_2$ 3. A QUD:= $q_1 \prec q_2$; asserts p_1 ABOUT q_2 ; A QUD: $q_1 \prec q_2 \prec p_1?$ 4. B QUD:= $q_1 \prec q_2 \prec p_1?$ B FACTS : soa(p_1); downdates $p_1?$, q_2 from QUD: B QUD: q_1 ; 5. asserts p_3 ABOUT q_1 ; B QUD:= $q_1 \prec p_3?$
--

- (5) a. q_1 : ($s_{dinner} ? \lambda x \langle \langle \text{INVITE} - \text{ABLE}, x \rangle \rangle$)
 b. Content of answer phrase: (reference to the individual) j .
 c. Content of the short answer: ($s_{dinner} ! \langle \langle \text{INVITE} - \text{ABLE}, j \rangle \rangle$)

3 Utterance Understanding

One of the key components in a theory of dialogue structure, within a perspective put forward by Herb Clark and his associates (see Clark 1996), is a notion of *understanding an utterance*

sufficiently for current purposes: Clark et al suggest, on the basis of corpus analysis and psycholinguistic experimentation, that in a conversation CP's are constantly monitoring each other for evidence that the latest utterance has been understood. If it has, the utterance is *grounded* and things can proceed unhindered. Otherwise, clarification is called for. Although Clark and his associates have not attempted to formalize such a notion of understanding, the existence of some such formalization is a prerequisite for evaluating the viability of this perspective. Formal semantics has not hitherto paid much explicit attention to (less than perfect) communication. It does, nonetheless, provide some useful tools: my purpose in this section is to sketch a view of communication and understanding which takes as its starting point (what I take to be) the standard but rarely enunciated practice of the communicative process which prevails among formal semanticists. For obvious reasons of space,⁴ the view I will sketch is quite simple minded, most prominently because it involves taking on board an assumption I dub *the unique content assumption*—essentially that understanding involves (one sided) mutual belief that the CP's have resolved the content *identically*. For the problematic nature of this assumption as well as a proposal for how it can be eliminated, see Ginzburg (to appear).

Literal content and understanding

The common sense semantic view of communication takes its initial inspiration in particular from the views of Frege (sense [intension]/reference [extension]), Lewis (on convention), Austin/Searle (on illocutionary meaning), Grice (on implicature). Semanticists⁵ assume that a necessary condition for understanding is (the addressee's) belief that she and the speaker agree about the literal meaning of the sentence in context/utterance. The mechanism for "transmission" of literal meaning is, implicitly, taken to be as described in Lewis (1968) via the notion of *conforming to a regularity as a convention*:

- (6) "whenever S is uttered, the utterer intends to communicate P and the hearer acquires the belief P".

Two points can be made about this assumption, one concerns illocutionary force (in the sense of 'indicator of dialogue move type'), the other more generally about content variability with context. The first point: Lewis' main focus was assertoric uses of declarative sentences. According to Lewis, the 'P' in (6) is to be identified with the proposition asserted by the utterer. However, for other types of speech acts, querying for instance, Lewis would identify 'P' with the illocutionary content of the utterance 'A asks q', 'A orders B to do a' etc. This asymmetry, however, is unjustified and based on a mistaken, non-interactive view of assertion, certainly inappropriate for dialogue. Once we recognize conventionalized move interaction sequences, then the primary content that an addressee recognizes must be viewed as illocutionary: henceforth the notion of 'literal content' for an utterance will, unless explicitly indicated otherwise, always include the illocutionary dimension.⁶

Moving on to the more general problem with (6), this is simply that it is directly applicable only to S's that have no context-dependent elements, perhaps exemplified by 'I, Galileo Galilei, do hereby claim that the earth moves.' However, the number of such S's that surface in conversation is rather low. Consider two very simple counterexamples: in (7a) the content of A's utterance is, roughly, 'A greets B', whereas the content of B's utterance is 'B greets A'; similarly in (7b) the contents are respectively 'A claims that A likes B (at the present time)' and 'B claims that B likes A (at the present time)':

- (7) a. A: hi
B: hi.

⁴Though of course not *only* for that reason.

⁵At least some semanticists have beliefs roughly like these, I think.

⁶Poesio and Traum (1997) make a related assumption: their updates of the common ground include facts about the illocutionary force of utterances.

- b. A: I like you
B: I like you.

The need to accommodate the dependence of content on context lead to the emergence of a distinction between *meaning*, a notion associated with sentences or types of utterances, and content, a notion characterizing utterance tokens. This distinction, the importance of which was first brought home to semanticists and logicians by David Kaplan, building also on insights due to Bar Hillel, Montague and Cresswell, was further developed by work in Situation Semantics (Barwise and Perry 1983, Gawron and Peters 1990, Cooper and Poesio 1994). On this view, a *meaning* is a function or, as I shall assume, an n-ary SOA abstract where the variables abstracted over correspond to the contextual parameters. For instance, in (8) a simplified, tenseless meaning for (7a) is given.

- (8) $\lambda u, a, b \langle \langle GREET, greeter : a, greeted : b \rangle \rangle$
RESTRICTIONS: $(u \models \langle \langle SPEAKER, a \rangle \rangle \wedge \langle \langle ADDRESSEE, b \rangle \rangle)$

The contextual parameters associated with 'hi' concern the greeter and the greatee, and the utterance. Hence, the meaning of 'hi' is a ternary abstract, the three variables allowing the variability of the contents expressed in utterances of 'hi' to be captured. In (8), it is probably clear enough what role *a* and *b* play. What about *u*? Meaning abstracts typically carry *restrictions*, additional conditions on the fillers of their argument roles, that typically concern the utterance.

By the same token a simplified, tenseless meaning for an assertoric use of (9a) is given in (9b). An utterance situation can vary the values for the speaker *a*, the situation described *s*, and the referent *b* for the NP 'Bill':⁷

- (9) a. Bill likes me.
b. $\lambda u, b, a, s \langle \langle ASSERT, a, (s! \langle \langle LIKE, b, a \rangle \rangle) \rangle \rangle$
RESTRICTIONS: $(u \models \langle \langle NAMED, 'Bill', b \rangle \rangle \wedge \langle \langle SPEAKER, a \rangle \rangle \wedge \langle \langle DESCRIBING, a, s \rangle \rangle)$

Now the proponents of the view of meanings as abstracts presumably have the following in mind as a revised version of the context-independent Lewisian regularity (6):⁸

- (10) "Whenever a speaker A makes an utterance *u* of S whose meaning is μ , the utterer communicates that P, where $P = \mu[f]$, where *f* is the assignment *u* provides for the variables of μ , and the hearer acquires the belief that P".

However, (10) is not a complete picture of the communicative process—of course it is a pretty reliable regularity with respect to face to face uses of the examples in (7). The semantic phenomena which motivated Kaplan and Barwise and Perry were primarily ones connected to indexicals 'I', 'you', 'now' etc where there is a strong intuition that the utterance (more traditionally: the context or the discourse situation) *fixes the values of the contextual parameters*. However, transfer (7a) to a phone conversation and immediately the regularity (10) appears to be a less useful description of what is going on:

- (11) [A phones B:]
1. B: Yeah?
2. A: hi.
3. B: Who's speaking?

⁷Arguably, one might also wish to parametrize the illocutionary force so that it ranges over e.g. 'warn', 'guess', 'threaten' etc. An alternative position, consistent with a view that assumes that the number of conventionally recognizable illocutionary forces is *small*, is to view forces such as 'warn', 'guess', 'threaten' as components of intentional meaning. This is not merely an exercise in terminological book-keeping: one can argue that there exists a restricted and conventionalized set of response options to an assertion or query. I am more skeptical about the existence of such "precompiled" response options for warnings, guesses and threats.

⁸For a subtler formulation than this see Perry (1995), Israel and Perry (1991) based on the *pure v. incremental* content an agent can acquire from any regularity.

(11) reminds us that addressees do not automatically recognize the content of an utterance; or, to use Perry and Israel's terminology, the *pure content* is (often) adjudged to be insufficient. Indeed the question B poses signals here that "constituent of the content" of A's utterance the value of which B remains ignorant about. (11) could continue as follows:

- (12) [A phones B:]
4. A: Bill.
 5. B: Which Bill?
 6. A: Bill Smith
 7. B: uh huh
 8. A: Did you phone me?
 9. B: when?

This suggests that a different formulation of the basic semantic regularity is in order. We need to represent the different positions that the speaker and the addressee stand in as a consequence of an utterance. The conception of questions of Ginzburg (1995a) is suggestive in this respect: any utterance u and meaning μ , just like any event/abstract pair, individuate a question, $(u?\mu)$, which I will dub the *conventional content question*. It can be paraphrased by: *what values are assigned to the contextual parameters of μ in u .*

(13) **conventional content question**

- a. u is an utterance of 'hi': the content question is:
 $(u ? \lambda v, a, b \langle \langle GREET, greeter : a, greeted : b \rangle \rangle)$
 RESTRICTIONS: $(v \models \langle \langle SPEAKER, a \rangle \rangle \wedge \langle \langle ADDRESSEE, b \rangle \rangle)$
 ('what speaker greeted what addressee')
- b. u is an utterance of 'Bill likes me': the content question is:
 $(u ? \lambda v, b, a, s \langle \langle ASSERT, a, (s! \langle \langle LIKE, b, a \rangle \rangle) \rangle \rangle)$
 RESTRICTIONS: $(v \models \langle \langle NAMED, 'Bill', b \rangle \rangle \wedge \langle \langle SPEAKER, a \rangle \rangle \wedge \langle \langle DESCRIBING, a, s \rangle \rangle)$
 ('which speaker a was asserting about what situation s and what person b named Bill that b liked a ')

According to the *unique content assumption*, there is a unique answer the addressee needs to find as a necessary condition for communicative success. Thus, the following suggests itself as a revised, interactive, version of (6), which takes context dependence into account:

- (14) "Whenever a speaker A makes an utterance u of S whose meaning is $\mu(u)$, the utterer communicates that P, where $P = \mu[f]$, where f is the assignment u provides for the variables of $\mu(u)$; the addressee B faces the question $(u?\mu(u))$; she needs to find *the exhaustive answer* to $(u?\mu(u))$ as a necessary condition for communicative success. "

Adding in the Intentional Component

Grasping the literal or conventional content might be a necessary condition for understanding, it is of course not sufficient. Occasionally, the primary goal of the speaker is transparently reflected by the literal meaning associated with the sentence she utters, but this is far from the rule. Thus, any model of human communication needs to accommodate the fact that speakers often attempt and succeed to communicate "more" than the literal content they express and, conversely, addressees attempt to find *why* the speaker made his utterance. Failure to understand why the utterance was made is viewed as failure to understand the utterance:

(15) [Context: couple meeting at breakfast table:]

1. A: Hi
2. B: Hi
3. A: How are you this morning?
4. B: Fine. And you?
5. A: Fine. So, the Pope is eating breakfast now.
6. B: huh? What do you mean? why are you saying this?
Why should I care? (etc)

Here even though B can understand what A (can be assumed to be) conveying literally, she finds the utterance quite *unclear* and demands clarification, quite akin to A's having referred to someone B doesn't know or having used a verb expressing a concept she's never previously encountered. This is a first indication of the problematic nature of maintaining a rigid separation of "semantics" (even when its domain of operation is generously defined as 'the study of conventionally conveyed meaning') from "pragmatics" (defined in the same spirit as 'the study of speaker meaning'): as far as grounding goes, i.e. determining when an utterance has been understood, the separation is quite arbitrary—a conversational participant will seek redress, demand clarification, whether the unclear aspect pertains to the conventionally determined aspects or the aspects pertaining to the underlying goals.

Grice in 'Logic and Conversation', and the lion's share of the literature that has built on that paper, does seem to have had a strongly modular, indeed a derivational view of implicature, since he assumes that the addressee has *first* figured out *what was said*,⁹ and then tries to reconcile that with the speaker's having followed the cooperative principle. Grice's "derivational" assumption is clearly too strong, as pointed out by a number of authors including Levinson and Sag (1988), Clark (1996) Thus, to modify an example due to Clark, in (16a) resolving the reference of 'it' to the garage would seem to depend on A's having understood B's implicature that A should fill up the car in the garage; whereas (16b) shows that one can coherently react first about the intentional component and follow that up with a question about the content:

- (16) a. A: My car has run out of petrol.
B: There is a garage round the corner, if it hasn't moved in the meanwhile
- b. A: Is John or Ming going to chair the session?
B: Why do you ask?
A: Just curious
B: You're asking about John Du Pont tomorrow's session etc?

A declarative view of utterance content

How, then, to place literal and intentional understanding on essentially equal footing? Of course the kind of interaction that takes place with *intentional* content, as Austin and Grice originally pointed out, is not peculiar to linguistically-based interaction but to any interaction that requires recognizing an agent's intentions. The utterer has one or more goals she intends to achieve *by* making her utterance. For communicative success the addressee for his part is required to recognize these goals. Or rather, if we take a more detailed view of the interaction, he faces the question *what goals does A intend to achieve by making the utterance u*, which for the moment I notate simply as GOALS(u,A)?. The addressee needs to find an answer to this question. One possible view is that he needs to find *the* exhaustive answer, but other weaker alternatives need also to be considered. Coming to a definite conclusion w/r to GOALS(u,A)? is probably an even more complex affair than with (u ? $\mu(u)$) if only because the evidence is less concrete and because of the hazy border between conscious and unconscious goals. Indeed one might try and simplify things by suggesting, as is common in AI work (see e.g. Grosz and Sidner (1986) notion of 'Discourse Segment Purpose') that the question the addressee faces is more along the lines of *what is the primary goal A intends to achieve by making the utterance u*, in which case finding the exhaustive

⁹'said' here in the technical sense of 'literal content'.

answer seems a more plausible criterion. We can now offer the following picture of communicative interaction:

(17) **Communicative Interaction (basic picture)**

A makes an utterance *u* of *S* whose meaning is $\mu(u)$, attempting to convey content *P*.
A believes that

- a. $\text{lit-content}(u, \mu(u)) = p_0 = \mu[f]$,
 where *f* is the assignment *u* provides for the variables of $\mu(u)$;
- b. $\text{Content}(u) = P = p_0 \wedge p_1$

The addressee *B* faces two questions:

- a. $(u? \mu(u))$: 'what values does *u* provide for the variables of $\mu(u)$ '
- b. $\text{GOALS}(A, u)?$: 'what goals does *A* intend to achieve by making the utterance *u*'.

Unique content assumption:

B grounds *u* iff *B* believes it is mutually believed that he found:

- a. the exhaustive answer to $(u? \mu(u))$ and
- b. the exhaustive answer to $\text{GOALS}(A, u)?$.

Formulating things in this way places literal understanding and intentional understanding on essentially equal footing. It leaves open exactly *how* the addressee comes up with the answers: at times the one answer helps the other answer to be found, at times vice versa and at times the answers can mutually constrain each other. Since on the current view both questions need to be considered essentially simultaneously, or at least not in a clear derivational ordering, it will be convenient to view the understanding process as involving the addressee pondering a single compound question obtained by conjoining $(u? \mu(u))$ and $\text{GOALS}(A, u)?$. I will notate this compound question as follows:¹⁰

Definition 4 $\text{content}(u, A, \mu(u))? =_{\text{def}} \{(u? \mu(u)), \text{GOALS}(A, u)?\}$

4 Utterance Updates

An Initial Formulation of the Protocol

I turn now to sketch an interaction protocol for utterances which makes explicit the view of communication and utterance understanding articulated above. In particular, it allows us to develop accounts of querying and assertion that no longer assume *perfect communication*. I start by formulating quite informally a protocol that applies to perhaps the most mundane case—where full understanding appears to exist. I formulate the protocol as a sequence of instructions to a DP about to update her DGB, from the point where she believes an utterance has taken place, which I will assume simplistically is a TRP,¹¹ and through to the point where she has offered a response.

(18) If an utterance *u* has occurred, try to:

- a. understand *u*, relative to your UNPUB-MS
- b. update LATEST-MOVE (L-M) with *u*
- c. react to L-M according to the protocols for querying/assertion

(18a) and (18c) are components that were elucidated in section 3, and section 2 respectively. The main issue that remains is explicating the updating of L-M, to which I turn shortly. The main point worth noting at this stage about the protocol is a subtle speaker/addressee asymmetry, emphasized by the word 'try': as far as the addressee of an utterance goes, the protocol does not

¹⁰In the semantics for compound questions in Ginzburg and Sag (to appear), a SOA τ is ABOUT (RESOLVES) a question $\{q_1, \dots, q_n\}$ iff τ is ABOUT (RESOLVES) at least one (all) q_i .

¹¹That is, a Transition Relevance Point, a point at which, according to Conversation Analysis the turn could change.

ensure that an update or reaction will be possible—the addressee might fail to understand *u* or updating might result in incoherence.

On the other hand, I am assuming that it is a fundamental requirement in conversations for a speaker, the author of a given utterance, A, to be aware of the content she intends to convey. Her perspective on the utterance then arises as a special case by assuming (18a) to be vacuously satisfied. I explicitly assume that a speaker incorporates her utterance into the DGB as soon as she makes it—having made her utterance A needs to update L-M with the content of her utterance, as it is this update that triggers the update moves associated with querying and assertion (QUD and TOPICAL get updated). These updates must take place immediately following (or simultaneously with) the utterance, if only to allow also for the case where the speaker does not immediately surrender her turn and build on her previous utterance.

What is (in) LATEST-MOVE?

In order to get a working version of the utterance protocol we need to offer an account of how L-M gets updated, which in turn involves us in considering in more detail what information needs to be represented in L-M. I have been assuming that the value of L-M is the (full illocutionary) content of an utterance. This was sufficient for the purpose L-M has played so far, namely as the semantic object relative to which the range of possible updates and reactions are defined. Once we allow for the possibility that perfect communication does not obtain, the nature of L-M needs to be modified rather significantly.

A useful way of thinking about L-M is as a speaker's representation of the utterance where information is held in the medium-term before (certain traces of it are) are potentially stored in more long-term repositories. We need to consider what utterance-related information above and beyond solely the *content* of the utterance, is of "semantic significance" and requiring representation in the DGB. There are in principle at least two types of information: on the one hand there are a wide range of facts about the meaning and form of the utterance, some of them quite ephemeral in nature, which potentially become presupposed. This exemplified in (19): if A makes the utterance in (19.1) a variety of facts about the utterance potentially enter into DGB | FACTS (in italics in (19.2a-d)), as evinced by the possibility of embedding them under a factive-presupposition predicate such as 'interesting'.¹²

- (19) 1. A: Did Mark send you a love letter?
2. B: No, though it's interesting
 a. *that you refer to Mark/my brother.*
 b. *that you bring up the sending of love letters.*
 c. *that you ask about Mark's epistolary habits.*
 d. *that the final two words you just uttered start with 'I'.*

Despite the seeming triviality of some of this information or at least the absence of obvious motivation to preserve such information, one should be cautious about dismissing such information as information that invariably decays and is not preserved, in the spirit of the slogan 'meaning/content is preserved, structure decays', prominent following the famous experimental results of Sachs (1967), which apparently showed just that. Subsequent experimental results have actually suggested that the issue is more complex: thus, studies by Kintsch and Bates (1977) and Keenan et al. (1977) show that detailed verbatim recall can occur, given sufficient reason to do so.

Without taking an overly strong position as to which information about the utterance is *unimportant*, one can venture to suggest which information about the utterance definitely *is* important, important enough to warrant preservation together with the content. Qualitatively put: this is whatever information will enable the speaker to figure out what aspects of the content of her utterance might remain unclear to the addressee. In other words, it is some representation which can enable the speaker to reconstruct what it was she said and what the addressee could fail to

¹²On the analysis developed here, such information will not be TOPICAL (in the sense of footnote 1), and hence cannot be "easily" referred to using anaphora and ellipsis.

grasp. It is of course unrealistic and implausible to assume that subsequent to her utterance, a speaker will be ready to react to each and every possible wrong turning that the addressee could make. Nonetheless, the data concerning clarification we saw in the examples in (1) and more of which we will see below, suggests that the context equips both sides with the ability to access with ease a wide range of potential misunderstandings. Thus, presumably this potential is to be cashed out in terms of some object which encodes the shared linguistic knowledge associated with a given utterance. What could this object be? Given our discussion in section 3, it is no surprise to put forward as a candidate the *meaning* of the sign used in the utterance, where meanings are to be conceived, following the explication of Kaplan and situation semantics, as functions or n-ary abstracts which arise by abstracting from a content those parameters which “vary with context”.

Thus, the discussion above suggests, in line with Poesio (1998), that a plausible view of L-M is that it is a repository that contains two kinds of information:

- Information that characterizes u as a token, namely the (DP relativized) *content* of u
- Information concerning the type which classifies u : its meaning, its syntactic structure, and its phonological structure; a conglomerate of information reified in a number of linguistic frameworks, e.g. HPSG as a *sign*.

In sum, then, the additional information associated with L-M merely means that a more complete record of the utterance event, which a speaker must in any case have available to her at utterance time, is posited as having potential semantic significance for the subsequent interaction.

Current and Latest Move

How does L-M get updated? The update process involves two components: one simply concerning the incorporation of the most recently performed (speaker) or most recently understood (addressee) utterance as L-M, the move to be reacted to. The other component concerns the *old* L-M—is it grounded? In other words, is it to attain the status of a common ground presupposition, viz. get added to FACTS | STORED?^{13,14}

Contending with this issue leads to one concrete amplification in the DGB: in order to formulate L-M update, it will be necessary to maintain simultaneously in the DGB the information concerning the current, about-to-become L-M utterance and the existing L-M, since the former will frequently determine whether the latter is or is not grounded. I dub this attribute ‘CURRENT-MOVE’ (C-M)—its value is of the same type as L-M.

Definition 5 A | L-M is grounded

- a. if $\text{spkr}(L-M) \neq A$
- b. if $\text{spkr}(L-M) = A$ and C-M grounds L-M

Definition 6 C-M grounds L-M

- a. if $\text{content}(C-M)$ is ‘Spkr(C-M) understands L-M’
- b. if $\text{content}(C-M)$ is an adjacency pair to $\text{content}(L-M)$

Given definitions 5, 6, we can formulate L-M update as follows:

Definition 7 L-M update:

- a. if L-M is grounded, then update:
FACTS | STORED := FACTS | STORED + L-M;
- b. L-M: = C-M

¹³ Actually, things are more complex: in certain cases L-M not being grounded will not simply result in failure to store L-M, but rather lead a CP A to discuss $\text{content}(L-M, A, \mu(L-M))$?, e.g. issuing a correction.

¹⁴ With Traum (1994) and contra Clark (1996), I don’t assume that every utterance *requires* grounding, the prime example of such a move are acceptances and acknowledgements.

This means we are now in a position to return to our initial, informal formulation of the utterance incorporation protocol and fill in the specifications with (a certain amount of) more formal substance.

- (20) **Utterance processing protocol (initial formulation)**
 For a CP B: if an utterance u has occurred, try to:
- a. understand u , relative to your DGB/UNPUB-MS:
 $C-M := \langle \text{content: } \tau, \text{ sign-inf: } \sigma \rangle$
 where τ resolves $\text{content}(u, A, \mu(u))$?
 - b. update L-M with u :
 If L-M is grounded,
 then $\text{FACTS} \mid \text{STORED} := \text{FACTS} \mid \text{STORED} + \text{L-M}$;
 $\text{L-M} = \text{C-M}$
 - c. react to L-M according to protocols for querying/assertion.

5 Comprehension problems

Emergence of a Mismatch and its Implications

The protocol, as formulated sofar, could explain successful (or at least apparently successful) cases of comprehension. It is, nonetheless, equally important to be able to explain the smooth conversational development which follows *failure* to fully understand an utterance. Given the theory of understanding available to us, there is a fairly obvious explanatory route: failure for a CP B to fully understand u means that she lacks a resolving fact for $\text{content}(u, A, \mu(u))$?; in a large number of cases, this has two predictable consequences: first, u cannot be reacted to in the expected "adjacency pair" way— u must be set aside at least for the while. Second, clarification about u must ensue—some unresolved aspect of $\text{content}(u, A, \mu(u))$? needs to be raised for discussion, by querying or assertion.

We can, therefore, add an alternative in the protocol:

- (21) **Utterance processing protocol (revised formulation)**
 If an utterance u has occurred:
- a. try to understand u , relative to your DGB/UNPUB-MS
 ...
- Otherwise:
- b. set u aside
 - c. make $\text{content}(u, A, \mu(u))$? QUD-maximal;
 - d. provide a $\text{content}(u, A, \mu(u))$?-specific utterance

In order to formalize (21) we need to make sense of what 'setting u aside' amounts to. Given space constraints the following sketchy remarks will have to suffice: I posit an ordered set PENDING whose elements are utterances—which on the current conception these are events/situations. When a new utterance is encountered it becomes the first element of PENDING. The maximal element u_{max} of PENDING becomes (or more precisely projects information that constitutes) CURRENT-MOVE if the CP understands and can coherently incorporate it in the DGB. When CURRENT-MOVE becomes L-M it is deleted from PENDING. Otherwise u_{max} remains (pending future attention) in PENDING.

Let me now note the central implication of (21), along with a potential problem. The implication is that a possible occurrence following an utterance u is a *semantic mismatch between the CP's*—different semantic objects are available to them at that point: A, the speaker of u , according to the assumptions previously articulated, incorporates the content (and sign information) of u into her DGB | L-M immediately following u . If, for example, u consisted of a query q , the consequence of this is that q becomes maximal in A's QUD, that an address for q exists in A's TOPICAL, and hence, A can accommodate a fact positively resolving q into TOPICAL. At this

point, according to my assumption, $\text{content}(u, A, \mu(u))?$ does *not* figure in A's DGB—A herself has no explicit intention to discuss $\text{content}(u, A, \mu(u))?$, she need not adopt this as an issue in QUD *unless* B subsequently provides indications to the contrary, in other words B indicates that he cannot ground u .

With B the situation is quite different: as soon as she believes u has occurred, she needs to consider whether in her UNPUB-MS there is a fact that resolves $\text{content}(u, A, \mu(u))?$; if there is not, then she is permitted to make $\text{content}(u, A, \mu(u))?$ maximal in QUD, hence providing her with semantic objects unavailable at that point to A. The existence of a semantic mismatch between the CP's, thus, emerges quite naturally within the picture developed here. That such a mismatch occurs in practice I have argued by pointing out *inter alia* the Turn Taking Puzzle (TTP) (see Ginzburg 1997, to appear), exemplified here by example (22): the data at issue here concern the resolution of the bare factive-operator wh-phrase 'why'. (22a), where the original speaker keeps the turn, contrasts minimally with (22b), where two distinct speakers are involved. In (22a), 'why' must pick up on a fact that positively resolves the initial question A poses, whereas when 'why' is uttered by a new speaker the resolution is to a fact characterizing A's initial utterance. Notice that these data cannot be explained merely as a consequence of the differing coherence of an utterance depending on who makes the utterance: the resolution unavailable to A in (22a) is coherent when it arises from a non-elliptical utterance, as in (22c):

(22) The Turn Taking Puzzle (interrogative version)

- a. 1. A: Where was your Grandmother's sister born?
2. Why? ('Why was she born *there*?')
- b. 1. A: Where was your Grandmother's sister born?
2. B: Why? ('Why do you ask where she was born?')
- c. 1. A: Where was your Grandmother's sister born?
2. (and) Why am I asking this question?

Sofar, then, following the need to accommodate the potential for failure to comprehend, I have suggested that this leads to the potential emergence of a mismatch in the DGB's of the CP's, a hypothesis which is exploited in the explanation of a number of phenomena, the TTP among others. Danger lurks, however:¹⁵ the positing of a semantic mismatch in the DGB's of the CP's although useful in explaining asymmetries in production might also suggest, contrary to fact, that asymmetries arise at the level of comprehension. In the TTP data there is no indication that the CP's cannot understand their interlocuter's utterance. Similarly, it is not usually the case that the original speaker of u faces problems in comprehending clarification utterances about u , such as (23) which rely strongly for their interpretation on the semantic and syntactic structure of u , despite the fact that $\text{content}(u, A, \mu(u))?$ is not at that point maximal in QUD:

- (23) 1. A: Who left Bill?
2. B: Bill?

What gives? In fact, there is a fairly straightforward out:¹⁶ consider A who has made an utterance u_0 . B responds with u_1 . *ceteris paribus* A's initial expectation is that u_1 constitutes an adjacency pair with u_0 . What if this is not the case? One option for A is simply to pose a clarification utterance about u_1 . However, before going down this route, A should also consider the possibility that B himself could not comprehend u_0 , made $\text{content}(u_0, A, \mu(u))?$ QUD-maximal in his DGB, and consequently made u_1 with the thus updated DGB. This suggests a revised protocol along the following lines:

¹⁵I would like to thank Mandy Simons for alerting me to the problem, which lead me to reevaluate my strategy on this issue.

¹⁶Another option one might consider developing is that the CP's keep track of the DGB of their interlocuter. Pretheoretically, however, this seems for the most part redundant and to involve too much of an additional burden for each CP; as far as developing protocols for interaction with machines, such a consideration of course does not apply.

- (24) **Utterance processing protocol (final formulation)**
 For a CP B: if an utterance u is PENDING, try to:
- a. understand u , relative to your UNPUB-MS
 - b. update L-M with u
 - c. react to L-M according to protocols for querying/assertion
 - d. If successful, u is removed from PENDING
- Otherwise:
- e. goto (a) with $\text{content}(\text{L-M}, B, \mu(\text{L-M}))?$ QUD-maximal;
- Otherwise:
- f. make $\text{content}(u, A, \mu(u))?$ QUD-maximal;
 - g. provide a $\text{content}(u, A, \mu(u))?$ -specific utterance

This protocol does explicitly assume that adjacency pairs constitute a preferred interpretation in the sense that the first interpretive route down which an addressee goes involves attempting to incorporate an utterance as a follow-up to L-M; if this route is blocked somehow, because $\text{content}(u, A, \mu(u))?$ cannot be resolved or the putative resolution leads to incoherence, only then does the addressee try to “repair” things by making $\text{content}(\text{L-M}, B, \mu(\text{L-M}))?$ QUD-maximal and utilizing that in interpretation. If that too fails, then, as it were, the addressee signals her “distress” about u and makes $\text{content}(u, A, \mu(u))?$ QUD-maximal. In this respect, the proposed protocol is inspired by the observations of Conversation Analysis. There are cases where indeed there is some potential for ambiguity—the utterance can be interpreted either as a clarification or as a direct response. For instance, (25.2) can be interpreted either as a suggestion that perhaps Mary is the only person that admires herself or as an attempt to clarify the reference of ‘Mary’. Although intonation can disambiguate the choices, it need not—it is certainly possible to intone (25.2) in a way that preserves the ambiguity :

- (25) 1. A: Who do you think is the only person that admires Mary?
 2. B: Mary?

The prediction made by the protocol is that A will tend to interpret (25.2) as a response to her question, not as a clarification. Although, as far as I can tell, this seems like a correct prediction in this case clearly this is an area where significant experimental work is required.

6 The Semantics of Reprise Utterance-types

I give here a brief summary of the proposal for the semantics for reprise utterance-types recently proposed in Ginzburg and Sag (to appear), who provide a detailed syntactic and semantic treatment within their HPSG treatment of interrogatives.

Consider for a start cases like the .2 utterances in (26): (26a.2) is not an assertion—B can make the utterance without any evidence about an event involving Merle and Brendan; (26b.2) is not a wh-question, but rather a y/n-question (note that the expected response by B is *not*, say, ‘Bill’); whereas (26c.2) is not a command—B doesn’t necessarily wish for A to go to her home:

- (26) a. 1. A: Merle attacked Brendan yesterday.
 2. B: Merle attacked Brendan yesterday?
- b. 1. A: Who left?
 2. B: Who left?
- c. 1. A: Go home!
 2. B: Go home?

Thus, when used to clarify the content, the reprising utterance gets a content distinct from what it expresses in a nonreprise use—essentially, the repriser attempts to confirm whether the speaker

of the reprised utterance intended to convey a particular content.¹⁷ Call such utterances *literal* reprises. Such utterances involve *use*, not *mention*, as shown most clearly by evidence from indexicals (27), and can be used for a variety of purposes not all of which are strictly “metalinguistic” (e.g. expressing astonishment):

- (27) 1. B: I haven't got a stone circle.
 2. A: You (# I) don't have a stone circle.
 3. B: No.

A rather straightforward paraphrase for such uses is that they are y/n-questions, where the queried SOA is the full illocutionary content of the previous utterance:

- (28) a. Is A claiming that Merie attacked Brendan yesterday?
 b. Is A asking who left?
 c. Is A ordering B to go home?

In addition to literal reprises, there are also *reprise interrogatives*:

- (29) a. Who likes WHAT?
 b. You saw Leda do WHAT?

The only difference between a literal reprise and a reprise use of an interrogative is that in the latter case one or more of the constituents of the previous content is not reprised —or replaced with a coreferential expression — but rather ‘replaced’ as a consequence of associating that argument role with an interrogative phrase. Indeed, subject to certain syntactic constraints, reprise sentences can be paraphrased by INDEPENDENT uses of entirely ‘normal’ fronted interrogative sentences:

- (30) a. You like WHO? (=Who did you say (just now) you like?)
 b. (Go home Billy!) Go WHERE?
 (= Where did you order me (just now) to go?)
 c. Who likes WHAT? (= ??What did you ask me (just now) who likes?)
 Violation of *wh*-island constraint

Similarly, reprise uses can be reported by embedding an interrogative under *ask*, just like other queries:

- (31) a. Jill: You like WHO?
 Jill was amazed at what I told her, so she asked me who it was that I had said I liked.
 b. (Go home Billy!) Go WHERE?
 Bill couldn't believe his ears, and asked, utterly incredulously, where I had ordered him to go to.

So, in the general case, the semantic content of a sentence *u* in its reprise use of an utterance *u*₀ is characterized as follows:

(32) **Reprise Semantics**

- a. The semantic content of *u* is a question whose queried predicate has as its main relation either *assert-rel*, *ask-rel*, *order-rel*; this relation has two argument roles: an utterer and a *message* argument.
 b. The UTTERER filling the utterer role is the individual who uttered the immediately prior sentence in the dialogue.
 c. The *message* argument, which must be of the same type as the semantic content of *u*₀, is the meaning assigned to *u* by the grammar, modulo the interrogative parameters that are abstracted.

¹⁷Such utterances also have an additional type of use in which the repriser indicates that she has not managed to resolve the content of a particular constituent or of the entire utterance. This is briefly discussed further in section 7.

This yields contents such as the following:

- (33) a. Who likes WHAT?
 ASK,
 b. $(u_0? \lambda x \langle \langle \text{utterer:A,}$
 message: $(s ? \lambda y \langle \langle \text{LIKE, liker:y, liked-obj:x} \rangle \rangle) \rangle \rangle)$

Here we have what Karttunen has been called a 'second order' question. On Ginzburg and Sag's account, nothing additional has to be said for such contents beyond the usual issues concerning the embedding of sentential content: since the previous speaker's utterance expressed a question, the SOA of the subsequent reprise question must be of type *question*. The interpretation of reprise questions is thus built from the interpretations that the grammar assigns to various types of clause, with the highest level of the semantics (the outermost question) embedding the content of the clause as the appropriate MESSAGE argument. Finally, it should be noted that this analysis of reprise *wh*-interrogatives immediately allows for LITERAL REPRISSES. These arise as a special case where the set of retrieved parameters is empty.

7 The coherence of reprise utterances

I now want to show how the theory of utterance meaning, the semantics of reprise utterance sketched above, and the view of clarification proposed in section 5 interact. The Kaplan/situation semantics theory of meaning suitably extended assigns (34a) the meaning in (34b). Within such a view, those parameters which "vary with context" are all abstracted away from the content—in this case, these would be taken to be the referent of 'Jill' (*j*), the speaker (*C*), the described situation (*s*), the event time (*t*) and the utterance event/situation (*v*):

- (34) a. Is Jill available?
 b. $\lambda v, C, s, j, t \langle \langle \text{ASK, } C, (s? \langle \langle \text{AVAILABLE, agent: } j, \text{ time: } t \rangle \rangle) \rangle \rangle$
 RESTRICTIONS:
 $v \models \langle \langle \text{NAMED, 'JILL', } j \rangle \rangle \wedge \langle \langle \text{REFER, } C, j \rangle \rangle$
 $\langle \langle =, \text{SPEAKER}(v), C \rangle \rangle \wedge \langle \langle \text{CONCURRENT, } t, \text{time}(v) \rangle \rangle$

I have proposed that if a CP B chooses the clarification option for an utterance *u* made by A, this means that B makes content(*u*, A, $\mu(u)$)? QUD-maximal, and makes an utterance *u*₁ that is specific to that question, either information ABOUT content(*u*, A, $\mu(u)$)? or a question that INFLUENCES content(*u*, A, $\mu(u)$)?

Utterances that provide information ABOUT content(*u*, A, $\mu(u)$)? occur primarily as *acknowledgements* or corrections (for exemplification and discussion see Ginzburg to appear). What of clarification queries? Recall the definition of the INFLUENCE relation between questions:

- (35) q_1 INFLUENCES q_0 if and only if any fact that potentially resolves q_0 also potentially resolves q_1

Now since content(*u*, A, $\mu(u)$)? is a complex question comprised of two questions, the conventional content question, (*u*? $\mu(u)$), and GOALS(*u*, A)?, the characterization of coherent clarification questions can be broken in two: questions that INFLUENCE GOALS(*u*, A)? and questions that INFLUENCE (*u*? $\mu(u)$). Since INFLUENCE is a reflexive relation, it follows in particular that, according to this characterization, it is predicted that GOALS(*u*, A)? is a coherent clarification question. This is a question that is realized in different forms depending on the nature of *u* as:

- (36) a. Why do you ask ... ?
 b. Why do you say ... ?
 c. Any special reason for saying/asking this?
 d. What do you mean (when you say/ask ...)?

Moving on to consider questions that INFLUENCE ($u? \mu(u)$), taking the meaning in (34) as a concrete case. In particular, it means that any of the following predicates can be queried in clarifications (where A_0, s_0, j_0, t_0 are contextually supplied fillers for the relevant roles):

- (37) a. $\lambda A \langle \langle ASK, A, (s_0? \langle \langle AVAILABLE, agent: j_0, time: t_0 \rangle \rangle) \rangle \rangle$
 b. $\lambda j \langle \langle ASK, A_0, (s_0? \langle \langle AVAILABLE, agent: j, time: t_0 \rangle \rangle) \rangle \rangle$
 c. $\lambda t \langle \langle ASK, A_0, (s_0? \langle \langle AVAILABLE, agent: j_0, time: t \rangle \rangle) \rangle \rangle$
 d. $\lambda s \langle \langle ASK, A_0, (s? \langle \langle AVAILABLE, agent: j_0, time: t_0 \rangle \rangle) \rangle \rangle$

Indeed these predicates are respectively the queried predicates of the questions expressed by the interrogatives in (38),¹⁸ all of which do in fact seem coherent follow-ups to (34a):

- (38) a. Who is asking? (cf. Who is speaking?)
 b. Is WHO available?/WHICH Jill (are you asking if she is available)?
 c. Is Jill available WHEN?
 d. Is Jill available for when/under what circumstances?

Similar reasoning shows that questions such as the following are also licensed as coherent follow-ups:

- (39) a. Is WHO available WHEN?
 b. What (kind of question) are you asking?

In this section I have demonstrated the interaction between a theory of meaning, the semantics of reprise utterances, and the range of clarifications that can coherently follow a given utterance. I have given a rough illustration of the range of coherent clarifications recognized by my own account, one which seems to cover at least the most basic cases. To what extent the account *overgenerates* is an open issue. I now move to consider briefly whether it undergenerates.

Anyone who has inspected a corpus will have noticed that one of the commonest means of clarification is by means of reprising a constituent of a previously occurring utterance, as in (40):

- (40) a. A: Did Jack kowtow at the party?
 B: kowtow?
 b. "Clausal" reading: are you asking if Jack KOWTOWED?
 c. "Constituent" reading: what did you mean when you said 'kowtow'?

I dub this "construction type" *clarification ellipsis* (CE). Given that, with certain isolated exceptions, for instance expletives, just about any constituent can be questioned in this way, two questions suggest themselves: (a) Does our characterization of the range of clarifications accommodate the contents expressed in CE? (b) Can we use our account of utterance updating to provide a theory of resolution for CE? Given space considerations, I cannot address these questions in the current paper; I note merely that CE exhibits at least two types of readings, exemplified in (40b,c). Clausal readings are *literal reprises* as discussed in section 6; their coherence follows from the same considerations discussed above. Although, I believe the coherence of constituent readings can be demonstrated fairly straightforwardly, an account of their resolution involves a number of complexities; in particular, since a CP who exploits this content is not committed to be able to resolve the content of the expression he reprises, one needs to appeal to a semantic mechanism which allows any XP to denote a variable of the appropriate type— i.e. behave like a wh-phrase used to reprise.

¹⁸(38d) is the problematic case here: given the vagueness of the notion of event/situation, it is not so clear how one asks about it. If one simply identified events/situations with spatio-temporal chunks, which is a problematic move (see Cooper and Poesio 1994), then either of (38c/d) would roughly correspond. Similarly perhaps the question in (ii):

(i) A: Who annoys Bill? B: Who annoys him WHERE?

8 Some move by move analyses

As a final illustration, I would like to consider an integrated semantic analysis of the dialogue in (41).

- (41) 1. A: Is Jill available?
 2. B: Is WHO available?
 3. A: My adviser
 4. B: Ah.
 5. no.

1. A | L-M: A asks q_1 ; A | QUD q_1 ; (1) becomes PENDING for B; he considers content((1),A, μ (1))?
 B cannot ground (1);
 B updates QUD: B | QUD:= content((1),A, μ (1))?
 poses q_2 that INFLUENCES content((1),A, μ (1))?
 PENDING:(1);B | L-M: B asks q_2 ; B | QUD:= content((1),A, μ (1))? \prec q_2 ;
2. becomes PENDING for A; she considers content((2),B, μ (2))? ; A can ground (2) *only* by making content(L-M,A, μ (L-M))? QUD-maximal: A | QUD:= $q_1 \prec$ content((1),A, μ (1))?
 MOVES UPDATE:C-M: B asks q_2 ; L-M is not STORED since it is not grounded; L-M:=C-M;
 A | QUD:= $q_1 \prec$ content((1),A, μ (1))? \prec q_2 ;
 A asserts p_1 ABOUT q_2 ;
 MOVES UPDATE: C-M:= A asserts p_1 ;
 L-M is stored because B made it: A | FACTS | STORED: B asks q_2 ; L-M:= C-M;
 A | QUD:= $q_1 \prec$ content((1),A, μ (1))? \prec $q_2 \prec$ p_1 ?
3. becomes PENDING for B: B |MOVES | PENDING: (1) \prec (3);
 B considers content((3),A, μ (3))?; B can ground (3): the interpretation of (3) relies on the QUD-maximality of q_2 since it is a short answer ellipse;
 (2) is grounded by (3) as an adj-pair, therefore:
 MOVES UPDATE: C-M:= A asserts p_1 ;
 B | FACTS | STORED: B asks q_2 ; L-M:= C-M;PENDING:=(1)
 B | QUD:= content((1),A, μ (1))? \prec $q_2 \prec$ p_1 ? ;
 B accepts p_1 ;
 B | FACTS | STORED:= FACT-CLOSURE({ B asks q_2 , p_1 }),
 B utters affirmative statement about p_1 ? : MOVES UPDATE: C-M: B accepts p_1 ;
 B | FACTS | STORED:= FACTS-CLOSURE({ B asks q_2 , p_1 , A asserts p_1 }) L-M:= C-M;
 Since p_1 resolves p_1 ?, q_2 , QUD downdate: B | QUD:= content((1),A, μ (1))?
 Assume B can now also resolve content((1),A, μ (1))?, then: QUD:= \emptyset
4. becomes PENDING for A: A considers content((4),A, μ ((4)))?; A can ground (4):
 the interpretation of (3) relies on the QUD-maximality of p_1 ? since it is a short answer ellipse;
 MOVES UPDATE: C-M: B accepts p_1 ;
 A | FACTS | STORED:= FACTS-CLOSURE({ B asks q_2 , A asserts p_1 , }); L-M:=C-M
 A | FACTS | STORED:= FACTS-CLOSURE({ B asks q_2 , p_1 , A asserts p_1 })
 A | QUD:= $q_1 \prec$ content((1),A, μ (1)).
 A can, non-deterministically, assume that content((1),A, μ (1))? is resolved and downdate it from QUD: A | QUD:= q_1
 (1) is PENDING for B at this point. Since she downdated content((1),A, μ (1))? from QUD, this means she can now ground (1):
 MOVES UPDATE: C-M:A asks q_1 ; (4) is not stored (it is unacknowledged as are most acknowledgements);
 L-M:= C-M;PENDING:= \emptyset
 B | QUD:= q_1 ; B makes a q_1 -specific utterance: B asserts p_3 ABOUT q_1 ;
 MOVES UPDATE:C-M: B asserts p_3 ;
 B | FACTS | STORED:= FACTS-CLOSURE({ B asks q_2 , A asserts p_1 , p_1 , A asks q_1 }; L-M:=C-M;
 B | QUD:= $q_1 \prec$ p_3 ?

Perhaps the most open issue concerning the above analysis is the analysis of the utterance 'ah', which given the current setup is most readily interpretable as an acceptance of A's assertion in (3); this involves a subsequent downdating from QUD, which has no patently observable reflexes.

One could suggest that 'ah' is simply a reflex of the grounding of (1), e.g. it accompanies the downdating of content($(1), A, \mu(1)$)? To give some justification for the particular analysis proposed above, I would like to point out an alternative ending for the above dialogue. In this ending the 'ah' is unambiguously an assertion acknowledgement since B continues to seek clarification about (1) and once again there is an unavoidable downdating in QUD of content($(1), A, \mu(1)$)? with no patently observable reflexes. Note that this also illustrates how clarification can recurse and is not confined by locality considerations; this of course means that a significant amount of non-semantic information needs to be maintained long-term in the DGB: the recursion here follows because after the acceptance of p_1 content($(1), A, \mu(1)$)? is maximal in both A and B's QUDs, which licenses updating the DGB with questions that INFLUENCE content- $q(u_1)$:

- (42)
1. A: Is Jill available?
 2. B: Is WHO available?
 3. A: My adviser.
 4. B: Ah. When?
 5. A: Tomorrow
 6. B: No

9 Conclusions

In this paper, I have provided a characterization of the range of possible clarifications that can coherently follow a given utterance act. The characterization is based on:

- An explication of utterance understanding. This is taken here to involve a compound question content($u, A, \mu(u)$)? ("what did A attempt to achieve by uttering u whose meaning is μ ")
- An account of utterance updating in which failure to understand u fully involves making content($u, A, \mu(u)$)? maximal in the contextual repository of questions under discussion (QUD)
- A proposal for characterizing the permitted follow-up questions when q is QUD-maximal to be those questions that INFLUENCE q
- A semantics for reprise sentences (e.g. 'You like WHO')

I have suggested that the characterization offered seems to cover at least the most basic cases; a future project is to use the basic account to construct an account of the ellipsis resolution involved in one of the most pervasive types of clarification utterance, 'Clarification Ellipsis'.

More generally, the paper ties in in a fairly precise way the theory of meaning, utterance understanding, the semantics of reprise utterances, and the range of clarifications that can coherently follow a given utterance; hence, providing an empirical challenge for any prospective theory of meaning.

References

- Asher, N. (1993). *Reference to Abstract Objects in English: a Philosophical Semantics for Natural Language Metaphysics*. Studies in Linguistics and Philosophy. Kluwer, Dordrecht.
- Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Bradford Books. MIT Press, Cambridge.
- Carlson, L. (1983). *Dialogue Games*. Synthese Language Library. D. Reidel, Dordrecht.
- Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- Cooper, R. and Poesio, M. (1994). Situation Theory. In *Fracas Deliverable D8*. The Fracas Consortium, Centre for Cognitive Science, Edinburgh.

- Gawron, M. and Peters, S. (1990). *Anaphora and Quantification in Situation Semantics*. CSLI Lecture Notes. CSLI, Stanford: California.
- Ginzburg, J. (to appear). *A Semantics for Interaction in Dialogue*. CSLI Publications, Stanford: California.
- Ginzburg, J. (1995a). Resolving Questions, I. *Linguistics and Philosophy*, 18:459–527.
- Ginzburg, J. (1997). Structural Mismatch in Dialogue. In Jaeger, G. and Benz, A., editors, *Proceedings of MunDial 97 (Technical Report 97-106)*, pages p. 59–80. Universitaet Muenchen Centrum fuer Informations- und Sprachverarbeitung, Muenchen.
- Ginzburg, J. and Sag, I. (to appear). *English Interrogative Constructions*. Studies in Constraint-based Lexicalism. CSLI Publications, Stanford: California.
- Grosz, B. and Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204.
- Hamblin, C. (1970). *Fallacies*. Methuen, London.
- Houghton, G. and Isard, S. (1986). Why to speak, what to say and how to say it: modelling language production in discourse. In Morris, P., editor, *Modelling Cognition*, pages 249–267. John Wiley, London.
- Israel, D. and Perry, J. (1991). What is Information. *CSLI Report*, 91-145.
- Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In et al, J. A., editor, *Themes from Kaplan*, pages 481–614. Oxford University Press, New York.
- Keenan, E., MacWhinney, B., and Mayhew, D. (1977). Pragmatics in Memory: A study of natural conversation. *Journal of Verbal Learning and Behaviour*, 16:549–560.
- Kintsch, W. and Bates, E. (1977). Recognition Memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 3:150–159.
- Levinson, S. and Sag, I. (1988). Pragmatic Inference and Semantic Interpretation. *CSLI Ms*.
- Lewis, D. (1968). *Convention*. Harvard University Press, Cambridge.
- Perry, J. (1995). Cognitive Significance and New Theories of Reference. In Perry, J., editor, *The Essential Indexical*. Oxford University Press, Oxford.
- Poesio, M. (1998). Underspecified Interpretations and a Theory of Language Processing. *Centre for Cognitive Science Ms*.
- Poesio, M. and Traum, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13:309–347.
- Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *P&P*, 2:437–442.
- Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversations*. PhD thesis, University of Rochester.

The TRINDI Project: Some Preliminary Themes

Stephen G Pulman,
SRI International, Cambridge
and University of Cambridge Computer Laboratory,
New Museums Site,
Cambridge CB2 3GG
sgp@cam.sri.com

Abstract

The TRINDI project (Task oriented INstructional Dialogue) is a recently started EU project involving the Universities of Gothenberg, Edinburgh, and Saarbrücken, as well as SRI Cambridge and Xerox Research Centre Europe (<http://www.ling.gu.se/~cooper/trindi/>). This presentation describes the aims of the project and discusses some preliminary themes that have emerged as central to the achievement of those aims.¹

1 Aims of the Project

By 'task oriented instructional dialogues' we mean dialogues between humans and machines that enable the human to make choices in the performance of a certain task. This task might be a simple one like booking an airline or rail ticket, planning a car journey, or a more complex one like carrying out a repair of some piece of equipment. There are of course many other types of dialogue, but these are of both practical interest, in that many groups are trying to build implementations of systems that will carry out dialogues of this type, and theoretical interest, in that many of the important issues of our understanding of dialogue arise even in these very simple settings.

We also aim to look at some instructional texts, like repair or operating manuals, even though these are not dialogues in the usual sense. The reason for also wanting to look at this type of text is that they display one feature of dialogue that is of particular interest to us: namely, that some implicit assumptions are made by the author about the participant's information state at different points in the text.

The official project aims are to:

- analyze those features of human-human dialogue to do with change of information state of the participants during the dialogue
- examine how such features can be manipulated in order to allow human-machine interaction which is both natural and robust (though more restricted than human-human dialogue)
- build a computational model of information revision in task-oriented and instructional dialogues and instructional texts

To begin with, we are focussing on the 'Autoroute' scenario, a route planning task for which we have a reasonably large corpus² of wizard-of-oz dialogues (Moore and Browning, 1992), and a working (keyboard input) dialogue system for this application based around the Core Language

¹This discussion is very provisional. It should not be taken to commit any of the other project members to anything suggested here.

²We gratefully acknowledge the Speech Research Unit of the United Kingdom Defence and Research Evaluation Agency, Malvern for making the corpus available to the Trindi project.

Engine at SRI Cambridge (Lewin et al., 1993) which will serve as a baseline system for the Trindi project.

The application that actually plans the route ('Autoroute') is a commercially available PC software package distributed free by many suppliers bundled as part of Windows. To obtain the plan of a route from the package a query must be submitted which has two obligatory parameters, the start and end of the journey, and several optional ones. These include: type of car (assumed to affect journey time); quickest or shortest route (not necessarily co-extensive in the UK at least); whether or not to avoid motorways; to choose a scenic route; be advised of places of special interest en route; start and arrival times etc.

The Autoroute scenario is typical of those for which implementations of spoken dialogue systems are commercially attractive: the task involves acquiring certain pieces of information - too many to be displayed as menu items - and then optionally some further details which can then be used to interrogate an application information system. The kind of dialogue that we are aiming at is illustrated below: an edited version of a dialogue between the wizard and a caller from the Autoroute corpus.

- (1) w: Welcome to the Route Planning Service How can I help you
c: A route please
w: Where would you like to start your journey
c: Malvern
w: Where would you like to go
c: Edwinstowe
w: Please wait
Is that Edwinstowe in Nottingham
c: Yes
w: What time would you like to make your journey
c: Six P M
w: Would you like the quickest or the shortest route
c: Quickest
w: Please wait while your route from Malvern to Edwinstowe is calculated
etc.

2 A Dialogue Move Engine?

Virtually all current approaches to the description of dialogue use some notion of dialogue act or conversational move/game (Sadek et al. 1996; Allen et al. 1996; Kowtko et al. 1992; Reithinger and Maier 1996). As a way of approaching the comparison of different versions of this notion, and of exploring their use in dialogue, in Trindi we thought it useful to try to design (and perhaps build) a 'dialogue move engine'.

What would such an engine do? To a first approximation, such an engine would be responsible for:

- classifying an utterance or fragment as an instance of one or perhaps more conversational moves
- choosing the next move, or set of possible next moves

In principle, we could use the dialogue move engine to explore more easily alternative sets of assumptions about the type of conversational game or move available to us, or the relationship assumed between moves and information states. It should also make it possible to experiment with different types of connection between dialogue move and information state, and linguistic processing, both in analysis and generation. For example, one might expect generation to become more difficult, but more natural, as more detail about information state was available to be used by the generator.

Given a provisional characterisation of a dialogue move engine as responsible for move analysis and move choice, this immediately gives us several questions about the architecture of such an engine and its relationship to the other components of a complete dialogue system:

What is the input to the engine?

What is the input to the engine? – or, more generally, what is the relation between a dialogue engine and the other components of linguistic processing?

Earlier in the chain of spoken language processing it is plausible that information from the dialogue move engine is available to restrict the range of hypotheses considered by the speech recogniser. If the last move was a question about a location, then recogniser hypotheses that are interpretable as descriptions of locations should presumably be preferred.

However, sentence processing (i.e. going from a string of words to quasi-logical form(s)) is presumably the responsibility of an independent component (since sentences do not only occur in dialogues), and it seems unlikely that there is any input to this process from the dialogue engine. But utterance processing (i.e. selecting and resolving a quasi-logical form to something that can be evaluated for truth) is not so obviously distinct, and indeed in many systems what is referred to as a dialogue manager is responsible for reference and ellipsis resolution (Lewin and Pulman 1995). Nevertheless, it seems advisable to maintain a clear distinction between reference resolution and dialogue move analysis. But this does not imply that the input to dialogue move recognition is always fully resolved.

Consider the following (made up) exchange:

- (2)
- 1: A. I want to plan a route for tomorrow from London to Cambridge.
 - 2: B. Where are you now?
 - 3: A. I'm there.
 - 4: B. London or Cambridge?

If this is a realistic dialogue, then it is clear that an utterance (3) that is not fully resolved can nevertheless be recognised as realising an appropriate conversational move, in this case a response to a wh-question. The fact that it is not fully resolved leads to a following clarification move, so it is also clear that partially resolved utterances can be used as the basis for deciding on the next move, at least where this move concerns that process of resolution. What is not yet clear, to me at least, is whether a partially resolved utterance can naturally lead to a following move which is something other than a clarification, requesting further resolution.

Notice that we can distinguish the case of under-resolved utterances from the more general phenomenon of 'grounding'. Utterances that would usually count as fully resolved linguistically speaking, like utterance 2 in the following, may nevertheless give rise to subsequent clarifications:

- (3)
1. w: Where would you like to go?
 2. c: Edwinstowe
 3. w: Edwinstowe?
 4. c: Yes
 5. w: Please wait.
 6. Is that Edwinstowe in Nottingham?
 7. c: Yes

(I am assuming that 'Edwinstowe' is a uniquely referring name and that the clarifications arise not through disambiguation but because the wizard is unsure she has heard correctly, even after the first check. There is only one Edwinstowe.) My own tentative conclusion is that the dialogue move engine can accept input at the fullest level of resolution possible. (Our own Autoroute implementation assumes that the input is fully resolved, though not necessarily grounded). If there are cases where dialogue move status can influence contextual reference resolution, that would suggest a more complicated interrelationship.

What is the relation between dialogue move recognition and selection, and information state?

In particular, is it plausible to suppose that the maintenance and update of the information state of a participant, including the participant's model of the information states of the other participants, is the responsibility of a separate component?

The definition of typical conversational moves, just as for the related speech act tradition, makes essential reference to the information state of the participants. For example, something with the syntax of a yes-no interrogative, and content P, will be classified as the conversational move 'yes-no-query' if the speaker is believed not to know whether or not P. If the speaker is believed to either know P or at least if it is believed that P is given information, then the same utterance type would be more likely to be classified as a check or clarification.

The information state component, assuming for a moment that it is separate, is responsible for maintaining and updating the beliefs, desires, and intentions of a participant. It will have to be able to carry out inference, and addition and retraction of information. The general assumption has always been that some kind of theorem proving and belief revision subsystems will be included in it (Pulman 1997). If the dialogue move component is independent of that, then a minimal requirement of it is that the knowledge about recognition and planning of moves is stated in a way that respects the logic of the information state component. It is obvious that it would be asking for trouble to have a condition on a dialogue move like 'not(speaker know P)' where 'not' was interpreted as negation-by-failure when the information state component was doing classical first order reasoning.

It may also be the case that the knowledge about dialogue move conditions which is needed to recognise or plan them, may itself be the subject of explicit reasoning, in the way that some treatments of indirect speech acts use axiomatisations of direct speech act recognition in a kind of meta-level inference (Perrault and Allen 1980). For example, if the only two kinds of conversational moves available for utterances of yes-no interrogative syntax are yes-no-query and clarification, then rhetorical yes-no questions cannot be correctly recognised. At this point we could choose either to add another conversational move type, or try to use the fact that in some sense rhetorical questions are parasitic upon ordinary ones and do some meta-level reasoning to try to make sense of the utterance. If this latter is the correct way to go, then it makes little sense to separate the dialogue move recognition component from the information state component, because everything that is in the former needs to be available to the latter. But if we think of the conditions attached to dialogue move recognition as essentially 'compiled out' pieces of reasoning then there is no cause to make them available for reflective inference.

Are the contents of information states underspecified?

This is really a question - which we do not presume to think we will resolve - about the granularity of the notion of belief for the purposes of dialogue. Consider the following situation: Manfred says 'I want to go there', and I do not know where it is that he wants to go. How should my belief about what Manfred wants be represented? Let us assume that we can represent the meaning of Manfred's utterance as a QLF-like structure:

```
want(manfred,go(manfred,PLACE))
```

where the meta-variable PLACE subsumes all its possible resolvents. Can we represent my belief by, crudely, putting this QLF in the list of my beliefs? Well, certainly not without also assuming that we can do the appropriate kind of reasoning with this underspecified representation, because I can conclude from his utterance that Manfred wants something, that he wants to go somewhere, etc. where these conclusions are, I will assume, represented by fully specified (even if not very definite) expressions that are not subsumed by the original QLF.

Alternatively, we could represent the content of Manfred's utterance as fully specified, but simply indefinite:

```
exists(x,place(x) & want(manfred,go(manfred,x)))
```

This is logically equivalent to

```
want(manfred,go(manfred,cambridge)) \/  
want(manfred,go(manfred,london))    \/ .... etc
```

although of course in an implementation we would not want the latter representation, both for practical reasons and for the fact that we do not want people to be committed to all the logical equivalents of their beliefs. But neither representation now needs any special mechanisms over and above what is needed for any inference within the scope of attitude predicates.

Similar questions arise for all the familiar cases for which linguistic underspecification is required. What can I conclude about Manfred's belief if he utters something like: 'I want to visit two people for three days', and the context does not make it clear to me whether he is making two visits or one. How do I represent that belief?

It is unfortunately clear that we are going to have to confront all the tricky problems concerning attitude reports familiar from the literature (e.g. Asher and Kamp 1989) with the added ingredient of underspecification thrown in for good measure. Our, or at least my, hope is that this extra ingredient will actually make this task a little easier, for being able to represent and reason with underspecified meanings is a large part of what is needed within a representationalist theory of the attitudes for coming to terms with the partiality, sensitivity to form and context, and varying granularity of beliefs and intentions.

A further complexity concerning the appropriate representations for information states arises (although this may not be motivated specifically by dialogue as such) from the need to represent information structure (focus/background, topic/comment, etc.) within our information states, for all the obvious linguistic reasons.

3 What are Conversational Moves and Games?

In most current implementations the framework of conversational moves and games is used as a descriptive device. The repertoire of moves can vary from one application to another (Kowtko et al. 1992; Reithinger and Maier 1996). Unlike most other linguistic units (e.g. noun phrases, generalised quantifiers) we do not yet have independent criteria which allow us to diagnose distinct moves, or to provide a description of linguistically possible or natural moves.

In the Autoroute implementation mentioned earlier (Lewin et al. 1993), we adopted essentially the set of conversational games and moves described in Kowtko et al. (1992). These moves and games are described there in intuitive terms: for example, a 'check' move 'requests the partner to confirm information that the checker has some reason to believe but is not entirely sure about'. These intuitive descriptions are in fact enough for coders to achieve a high level of consistency in annotation, (Carletta 1996), but it would be nice to have some tighter definitions, as well as some assurance that the list of moves that might need to be added was not open ended.

The Maptask set of games is as follows: Instruction, Confirmation, Question-YN, Question-WH, Explanation, and Alignment. Moves are classified into Initiating:

Instruct	(provides instruction)
Check	(confirmation of old info)
Query-yn	(yes-no question for new info)
Query-wh	(wh-question for new info)
Explain	(Gives unelicited description)
Align	(Checks position in task)

and Response or Feedback:

Clarify	(or rephrase old info)
Reply-y	(responds affirmatively)
Reply-n	(negatively)

Reply-wh (respond with requested info)
 Acknowledge (acknowledge and continue)
 Ready (to begin a new game)

In Lewin and Pulman (1997) we³ tried to produce an abstract formulation of moves and games in a way which would cover the way we were using these concepts in our own approach, and also, hopefully, provide a framework within which to eventually define the notion of 'possible conversational game', somewhat in the spirit of one of the earliest formal approaches to the study of dialogue (Hamblin 1971). In fact we took from Hamblin the notion of a conversational 'commitment': a proposition to which a participant is committed by the progress of the conversation, even if it might not, strictly speaking, figure among the beliefs or intentions of the participant.

We also use the notion of the 'propositions under discussion' (similar to the 'questions under discussion' of Ginzburg 1994): propositions which are introduced by utterances or inference, which may or may not correspond to beliefs or goals, and which may or may not eventually become commitments (or beliefs and goals).

The first question which comes to mind looking at the lists above is: what exactly is the difference between moves and games? We took the position that whereas moves, roughly speaking, affect the propositions under discussion in a dialogue, where these propositions are typically those introduced by the opening move of a game, it is at the level of games that commitments are established. If you carry out an assertion (e.g. in an 'Explanation' game) then the 'rules' of the game are such that you are committed to the truth of the assertion. This is, in the limit, independent of whether you actually believe the content of the assertion: if the assertion has been made to all intents and purposes straightforwardly then the other participants in the dialogue are entitled to complain if it turns out later that you were insincere. Likewise, Instructing establishes a public commitment to act in a certain way; Question games imply that the participant is not already in possession of the relevant information and has a genuine interest in it, etc. Commitments are publically established by carrying out conversational games in accordance with implicit rules of which all the participants are aware.

Some of this sounds like a rerun of speech act theory, and in some respects this is the case, because like speech act theory we are ultimately interested in relating utterances to information states. But not all of it is: neither games nor moves correspond exactly to speech acts. Games do not, because they typically stretch over several turns, whereas speech acts are at the level of individual utterances. Moves do not either: there is no specific illocutionary force of replying, aligning or acknowledging, but these are paradigm conversational moves.

We can illustrate the difference between games and moves using the earlier example:

(4)		<i>Move</i>	<i>Game</i>		
	1 w:	Where would you like to go?	query whq	} WH	
	2 c:	Edwinstowe	reply whq		
	3 w:	Edwinstowe?	check		} CHK
	4 c:	Yes	clarify		
	5 w:	Please wait	align/ack		
	6	Is that Edwinstowe in Nottingham?	query-yn/chk?		} CHK
	7 c:	Yes	reply yes		

On the present conception this simple exchange constitutes a wh-question game. It includes at least two sub-games of checking (or a check and a yes-no query, depending on the assumptions we make about the information states of the wizard and the caller). The initial utterance, 1, raises the propositions under discussion. Although the move realised by utterance 2 is a wh-reply, and thus one might think we now had a complete wh-query game, in fact successive moves and sub-games continue the game and refine the set of propositions under discussion until at utterance 7 the commitment 'caller wants to go to Edwinstowe, Notts' (or something like this) is established for both parties. It is not until this utterance that the wh-query game is complete.

³Much of what follows draws on work by Ian Lewin

A remark: in our implementation, participants are modelled as agents planning to achieve goals by playing conversational games. But there seems to be a difference in status between games like wh-query and, say, checking. It seems very likely that agents could plan, consciously or unconsciously, in terms of games like wh-query. It seems very unlikely, however, that anyone would plan in terms of checks or clarifications, which more naturally arise in response to the immediate circumstances, or more generally that anyone would form conversational plans with nested games.

To a first approximation then, we can begin a formalisation of conversational game theory by partitioning information states of participants into 'commitment slates' (to use Hamblin's term) which record commitments (C), and a kind of workspace for 'propositions under discussion' (PD). We assume a finite set of distinct game types (G) and move types (M), and an infinite set of propositions (P). For any participant, CS and PD are in Pow(P). Utterance meanings $U_{1...n}$ are in Pow(P).

A move token consists of a pair $\langle m \in M, U_i \rangle$, and is interpreted as a function from PD to PD. Roughly speaking, M says what operation to perform on the current PD (add, remove, strengthen, weaken, etc), and U_i says what content to use for that operation. (The set of possible and natural operations on information states is then a bound on the set of possible conversational moves.)

A game token consists of a pair $\langle g \in G, p \in \text{Pow}(P) \rangle$ and is interpreted as a function from $\text{PD} \times C$ to C. The game type g says what kind of commitment it is, and the proposition(s) p represent the content of the commitment. The initial set of PD is supplied by the opening move of the game.

A game is a function from $\text{PD} * C$ to C because, as in the example above, a game can be constituted not just by a sequence of moves (which only affect the propositions under discussion) but by sub-games which result in additional commitments.

A dialogue, then, is a sequence of games which take an empty commitment slate to one representing the commitments arising during the dialogue.

Clearly a lot more detail needs to be provided before this formulation is satisfactory: it is merely a starting point.

Is there a grammar of games? Could we describe a game in terms of some canonical sequence of moves? In our implementation we did exactly this (as do many other systems). Each game was associated with a recursive transition network whose arcs were labelled with members of M and G and whose states were labelled with the participant who had the next turn. This enables a characterisation of a well-formed game in terms of a sequence of utterances each of which realises a move in a game or subgame as licenced by the transition network.

I do not know whether this is correct. On the one hand, we want to be able to capture the fact that there are some rules or expectations that govern the sequence of moves that will be encountered in a dialogue. If you are a competent native speaker then you know that questions expect replies, that requests expect acknowledgements, and so on. On the other hand, it seems that a grammar which would accurately permit the wide variety of move sequences that are actually encountered when we move outside the simple Autoroute or similar domains would be so loose as to rule out very little.

4 The Relation between Theory and Implementation

In Pulman (1997) I argued that the current generation of dialogue systems (including our own) based on some version of the traditional belief-desire-intention model and using classical reasoning and planning methods, were unlikely to scale up to large scale domains or to display robust behaviour. I suggested recasting such systems within a probabilistic graphical network (Bayesian network) architecture. Categorising an utterance as realising one or more conversational moves (perhaps simultaneously) is a probabilistic function of various factors: previous move, current game, presumed information state of the other, linguistic form and content, etc., where the probabilities would ideally be empirically acquired from corpora. Choosing the next move is a calculation which takes into account the utilities of the effects that could be achieved by realising any of the currently possible moves, factored by the probability of actually achieving those effects given the

current state of all the other relevant variables. Planning the next game is also a decision theoretic calculation which takes into account the relative utilities of possible games in terms of the information gain that would accrue from playing that game, again factored by the probabilities of the currently relevant variables.

The role of information state management is assumed by the updating of probability values over the relevant components of a belief network which is provided by standard techniques for Bayesian belief revision.

This picture was an attempt both to describe a mechanism which stood a chance of robustness and scalability, and to stress the other game-like aspect of conversational games: they are played according to rules, but they are also played to 'win', in the sense that a 'good' task oriented dialogue is one that transfers the necessary information between the participants with the right compromise between caution, redundancy and speed.

Of course, the Bayesian framework I described, though robust and efficient, is a good deal more restricted than the classical BDI model in terms of the kinds of inferences that can be performed. Taken literally as a theory about the way dialogue works it is almost certainly false. It should rather be thought of as the performance version of the competence theory that is supplied by some version of the BDI or conversational game account. One of the themes that Trindi will try to address is what the relationship between the two conceptions is: one possibility is that it could be possible to 'compile out' much of the performance theory from the competence theory by setting limits to the complexity of inference needed to carry out certain types of dialogue to an agreed level of performance in terms of speed (or number of turns) and accuracy.

References

- Allen, J., Miller, B., Ringger, E., and Sikorski, T. (1996). A robust system for natural spoken dialogue. In *Proceedings of 34th ACL Santa Cruz*, pages 62-70.
- Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22 number 2:249-254.
- Ginzburg, J. (1994). An update semantics for dialogue. In Bunt, H., editor, *Proceedings of the International Workshop on Computational Semantics*, pages 111-120. ITK, Tilburg.
- Hamblin, C. (1971). Mathematical models of discourse. *Theoria*, 37:130-155.
- Kowtko, J., Isard, S., and Doherty, G. (1992). Conversational games within dialogue. HCRC research paper RP-31.
- Lewin, I. and Pulman, S. (1995). Inference in the resolution of ellipsis. In *Proceedings of ESCA Workshop on Spoken Dialogue Systems, Vigso, Denmark*, pages 53-56.
- Lewin, I., Russell, M., Carter, D., Browning, S., Ponting, K., and Pulman, S. (1993). A speech-based route enquiry system built from general-purpose components. In *Eurospeech '93: Proceedings of the 3rd European Conference on speech communication and technology*, volume 3, pages 2047-2050.
- Moore, R. and Browning, S. (1992). Results of an exercise to collect 'genuine' spoken enquiries using woz techniques. In *Proceedings of the Institute of Acoustics 14 6*, pages 613-620.
- Perrault, C. and Allen, J. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3-4):167-182.
- Reithinger, N. and Maier, E. (1996). Utilizing statistical dialogue act processing in verbmobil. In *Proceedings of 33rd ACL, Cambridge Mass.*, pages 116-121.

- Sadek, D., Ferrieux, A., Cozannet, A., Bretier, P., Panaget, F., and Simonin, J. (1996). Effective human computer cooperative spoken dialogue. In *Proceedings ICSLP 96 Conference in Spoken Language Processing, Philadelphia*, pages 546-549.
- Lewin, I. and Pulman, S. G. (1997). Reconfigurable Architectures for Spoken Dialogue Understanding SRI Cambridge final report to Defence and Evaluation Research Agency, Malvern.
- Pulman, S. G. (1997). Conversational games, Belief Revision and Bayesian Networks. In *CLIN VII: Proceedings of 7th Computational Linguistics in the Netherlands meeting, Nov 1996*, Eindhoven, also available at <http://www.cam.sri.com/tr/ABSTRACTS.html>.
- Asher, N. and Kamp, H. (1989). Self-Reference, Attitudes and Paradox. In: G. Chierchia, B. Partee and R. Turner (eds.): *Properties, Types and Meaning*. Kluwer.

Contracts in the Common Ground

Henk Zeevat

Computational Linguistics, University of Amsterdam
Spuistraat 134, 1012VB Amsterdam
henk@ai.let.uva.nl

Abstract

The paper explores the introduction of contracts in the common ground by means of requests, promises and questions. Various modal operators are introduced to achieve the modelling of these new additions to the conversational record. The paper extends the treatment of Zeevat (1997) and introduces the beginnings of a theory of action.

1 Introduction

In my Mundial paper (Zeevat, 1997), I gave the outlines of a framework for formal pragmatics. I take it to be the case that an important parameter in the explanation of linguistic behaviour is the estimate of the language user of the common ground that obtains between himself and his audience. This parameter controls the use of anaphoric and presuppositional devices, the choice of words in general and the communication plan of the speaker in general. It also controls the interpretation process, when the hearer tries to make sense of the utterances and the communicative behaviour of the speaker. The framework proposed is a special version of update semantics: update semantics in which the information state that gets changed is (the user's pictures of) the common ground. The model is intended as a further elaboration of Stalnaker's model for pragmatics as developed in Assertion.

Unlike others, I take it to be essential that the common ground is not taken to be something that really exists—a byproduct of the speaker and his audience having certain beliefs—but an object that is constructed independently of the actual beliefs of the communication partners. Common grounds also typically contain the beliefs of the communication partners, thus allowing the modelling of conflict. This leads to a treatment of speech acts that is more fine-grained than the traditional treatments.

I take it that update semantics combines traditional logic with a set of pragmatic operations. Logical operations are distinguished by the following characteristics. They can be characterised modeltheoretically by a Tarskian truth-definition, they do not allow truthvalue gaps and they are fully recursive. In contrast, pragmatic operators are not normally fully recursive, they fall outside the scope of model theoretic semantics and they tend to be defined only in certain circumstances. Typical pragmatic operators are assertion and other speech acts, presupposition and the operation that turns an information state into a common ground.

A shortcoming of the earlier work was that the whole dimension of the common ground was epistemic and that therefore there was no way of expressing very common speech acts such as questions, promises and requests into the framework. The current paper tries to give a first attempt at coming to grips with incorporating these notions into the framework and giving some tentative analyses of the speech acts involved. My hope is that these analyses contribute to the analysis of the relation that holds between a question and an information state and that expresses that the information state answers the question. This relation is crucial for the understanding of topic/focus articulation and for understanding communicative behaviour. Even on such a sophisticated theory of questions as the provided by Stokhof & Groenendijk, the relation is problematic for *why*- and *how*-questions (and for certain normal *wh*-questions such as *who is John?*). I have elsewhere

tried to show that the problems fully extend to *wh*-questions in an epistemic framework, where the partition approach does not work anymore. Getting more grip on the concepts involved in a pragmatic analysis of these questions therefore seems crucial.

2 Goals, Schedules and What Must Be

A goal is something that one wants to achieve. Having a goal is not necessarily the same as having decided to act on that goal. It may be the case that the goal is not immediately achievable and that certain subgoals have to be reached first. It is also not given that goals are mutually compatible: think of various ways of getting dinner tonight.

Goals lend themselves for a preference semantics. We can leave outside consideration those goals of which the subject recognises that they are unattainable. Together that then gives a preference preorder over the belief alternatives of a subject. $W_x p$ will be true iff x prefers those of his belief alternatives in which p is true over the ones in which p is false. We will assume that the relation is irreflexive, transitive and asymmetric. The state of no goals is the empty relation. The occurrence of symmetry makes the set of goals inconsistent.

Goals can be updated by refining the order, as long as asymmetry can be maintained. We will make the assumption that goals are false, mainly because it is a clean way to get rid of goals that are satisfied.

On the common ground as a whole, we can have agreed goals. Consider the the common ground of a speaker s and a hearer h . If it is the case that $W_s p$ and $W_h p$ both hold in the common ground, then it is a common goal. The common goals inherit their consistency from the goals in the participants' belief states represented in the common ground. $W\varphi$ abbreviates $W_s \varphi \wedge W_h \varphi$, if s and h are the two participants of the common ground.

Certain goals have been decided to act upon and change their status. In the first place, if a goal has been decided upon, it will be carried out. Not that we have infallibility in carrying out our actions, we sometimes fail. But scheduling an action is believing in its success. So it seems a reasonable approximation to assume that an actor believes that the scheduling of an action leads to its realisation. The principle is not entirely accurate since we may decide for actions whose failure brings only little trouble to schedule them even if we assume a fairly high risk of failure. But we will adopt it since we do not plan to deal here with sophisticated decision making.

The principle can be stated as:

$$B_x(P_x p \rightarrow \Box p)$$

Here $P_x p$ expresses that x has decided to carry out p . This requires that p is an action of which x is the agent.

The second principle that we adopt about courses of action that have been decided upon is that they continue to be goals:

$$P_x p \rightarrow W_x p$$

We further assume about P_x and W_x that they are transparant to belief:

$$\begin{aligned} P_x \varphi &\leftrightarrow B_x P_x \varphi \\ W_x \varphi &\leftrightarrow B_x W_x \varphi \\ B_x \varphi &\leftrightarrow B_x B_x \varphi \end{aligned}$$

This leads to the principle that on a common ground with s and h we have that

$$\begin{aligned} P_s \varphi \wedge B_h \Box \varphi &\rightarrow \Box \varphi \\ P_h \varphi \wedge B_s \Box \varphi &\rightarrow \Box \varphi \end{aligned}$$

This allows an abbreviation: we can say that the common ground schedules the action φ iff one of the participants has made it common ground that he has scheduled it and the others trust him. Trust of x in y with respect to a scheduled action p is the principle:

$$B_x(P_y p \rightarrow \Box p)$$

The last modal operator that we have to consider is that of causal necessity: in our model we consider a unique set of times forming a linear order and a set of possibilities such that each pair (t, i) , with t a time and i a possibility is an atemporal possibility. The future of a possibility (t, i) is defined by a function from (t, i) to a set of possibilities that coincide up to t with i .

Updates can so eliminate possible futures as well as possibilities.

3 Speech Acts

Request

A request by s for h to carry out A can be analysed as involving the proposition p , that is the predication of A to h . The speech act of the request can be treated as follows.

The preconditions are:

$$\begin{array}{ll} CG \not\models p & (p \text{ has not happened yet}) \\ CG \not\models P_h p & (\text{the hearer is not yet planning } p) \\ CG \not\models \neg \Box p & (p \text{ is not impossible}) \\ CG \not\models \neg W_s p & (\text{the speaker does not want } p \text{ to happen}) \end{array}$$

and perhaps:

$$CG \not\models \neg W_h p \quad (\text{the hearer is not disposed against } p)$$

All these conditions are part of the action presuppositions of the speech act. If p has already been done the speech act of requesting for p is void. We could strengthen this condition to

$$CG \models \neg p$$

But then we do not allow any more for the situation that p is the case but s does not know so. Here we want to part of the common ground. The fact that the minimal contribution is $W_s p$, entails that s believes p is not the case. Also if the plan is already to carry out p there is no point in carrying out the request: it cannot lead to a change. It is less than necessary to ask for p iff it is already known that h is going to do A , in fact one would expect it to be already CG that s prefers p to happen. It is important that it is not CG that p is impossible: in that case the request for p cannot be granted. If the speaker is not self-correcting, it cannot be CG that he does not want p to happen. If it has already been established that h does not want p , h 's acceptance of the the request would be a self-correction. Perhaps proper commands are a case where the preferences of the hearer are not supposed to be important.

The minimal contribution of the request is that s wants p to be carried out. So the request always changes the common ground to:

$$CG1 \models W_s p$$

In addition, the speaker also makes it clear that he trusts the hearer. This can be modelled as his public belief:

$$CG1 \models B_s(P_h p \rightarrow \Box p)$$

The intended reaction to a request is the acceptance of the request. This can be equated to a promise to carry out the action or the immediate execution of the action. We can then represent the first reaction by $P_h p$ which leads to the new $CG2$ such that

$$CG2 \models P_p$$

The second reaction can be formalised as

$$CG2 \models p$$

and it will no longer hold that $W_s p$.

Negative reactions to the request are first of all the refusal: $\neg W_h p$ or $\neg P_h p$.

Promise

The second speech act is the promise: a promise has the preconditions:

- $CG \not\models p$ (p has not happened yet)
- $CG \not\models P_p$ (p has not been planned yet)
- $CG \not\models P_s p$ (the speaker is not yet planning p)
- $CG \not\models \neg \Box p$ (p is not impossible)
- $CG \not\models \neg W_s p$ (the speaker does not want p to happen)
- $CG \not\models \neg W_h p$ (the hearer does not want p to happen)

The minimal contribution is here not: the speaker wants to carry out p but the stronger: the speaker plans to carry out p . The assent leads to $CG1$ in which it is planned to carry out p . In addition, the speaker must believe in the hearer's trust:

$$B_s B_h (P_s p \rightarrow \Box p)$$

By making the promise, the speaker also expresses his belief in the hearer's trust.

There is an ethical dimension to promises and requests, which has to do with the goal of these speech acts. The speech acts have to do with making our actions transparent to others and to allow the construction of actions that depend on the actions of others. Reliability is helpful here and a prerequisite for collaborative action. A more adventurous connection can be made with some of the formulations of the *Categorical Imperative* in Kant's *Prolegomena*. In our current setting, this would entail considering a common ground of all thinking beings and asking that a principle of action can be a principle that can be shared in that common ground. But a full exploration of this connection is not in the scope of this papersave.

Questions

A question in this setting is then both the expression of ignorance and the expression of a plan: the plan to get the conversational partner to supply information that meets the question. The following preconditions can be stated for *yes - no*-questions "whether q "

- $CG \not\models q$ (asking the question makes sense)
- $CG \not\models \neg q$ (asking the question makes sense)
- $CG \not\models B_h q$ (asking the question makes sense)
- $CG \not\models B_h \neg q$ (asking the question makes sense)
- $CG \not\models B_s q$ (asking the question makes sense)
- $CG \not\models B_s \neg q$ (asking the question makes sense)
- $CG \not\models \neg W_h (B_s q \vee B_s \neg q)$ (no correction of hearer)
- $CG \not\models \neg W_s (B_s q \vee B_s \neg q)$ (no correction of hearer)

The minimal contribution of the question is to communicate to the hearer that $\neg B_s q \wedge \neg B_s \neg q$. But it seems equally essential to inform the hearer that the speaker wants to know the answer:

$$W_s(B_s q \vee B_s \neg q)$$

Again there is trust implied in asking the question. Here the speaker implies that he will the hearer's word: $B_h q \rightarrow B_s q \wedge B_h \neg q \rightarrow B_s \neg q$. The cooperative hearer by choosing to answer the question by saying q or $\neg q$ meets the request and destroys the ignorance. In fact, destroying the ignorance destroys the goal.

The analysis can be extended to any question that can be represented by a disjunction of answers, such that any two disjuncts are incompatible.

$$\begin{aligned} CG &\models p_1 \vee \dots \vee p_n \\ CG &\models (p_i \wedge p_j) \text{ for any } i \text{ and } j \text{ such that } j \neq i \\ CG &\not\models \neg p_i \text{ for any } i \\ CG &\not\models B_h p_i \text{ for any } i \\ CG &\not\models \text{models } B_s p_i \text{ for any } i \end{aligned}$$

And the minimal contribution can be stated as:

$$CG \models W_s(\bigvee_i B_s p_i)$$

The ignorance of the speaker follows from the definition as above.

It is however not at all a trivial matter to come from an arbitrary question to a disjunctive representation as above. In fact, there are in general a number of ways in which a *wh-*, a *why-* or *how-*question can be represented as a disjunction and the exact way of doing so is highly influenced by the context. Of course, the partition view of *wh-*questions suggests that there is such a way: starting from a fixed domain, the partition is generated by considering each of the subsets as the answer. Here, we have to look at the question as standing in need of further resolution to a disjunctive question.

If the question expresses a goal of the speaker, it is related to other goals. One of the processes that guides the resolution is to infer that the speaker can achieve a further goal once he has achieved this one. In the following table I give a short overview.

- (1) Where can I buy cigarettes?
Inferred goal: speaker wants to buy cigarettes.
Directions to the nearest shop.
- (2) Who is John?
Inferred goal: The speaker wants to know how to get hold of John's paper.
A guy from Saarbruecken.
- (3) Who is Louis XIV?
Inferred goal: The speaker wants to know whether the candidate has properly studied his history book.
Lengthy descriptions as in the book.
- (4) Who ate the cake?
The speaker wants to get the culprit to buy a new one.
Identification knowledge.
- (5) Who attended the workshop?
Inferred goal: The speaker wants to know what to write in his report for the funding agency.
Which stars? Which disciplines were represented?
Either the list of stars or the list of represented disciplines.

(6) What does he look like?

Inferred goal: Speaker wants to have the means to recognise him at the drink.

Descriptions of sufficient detail to distinguish him from the others.

The above list of examples suggests that the disjunctive representation is often generated not so much from the domain of objects but from the domain of possible answers.

4 Planning

The resolution of questions to a disjunctive representation makes use of the process of getting what one wants by means of planning. A goal that cannot yet be acted upon or scheduled for action, is defective because the actor cannot yet have sufficient confidence in the course of action. What is still missing is the satisfaction of certain prerequisites of the action. To post a letter one needs to be able to move to a mailbox and in order to do so one must have the capacity of movement and knowledge as to the location of the mailbox or have a strategy (like asking or random movement) that guarantees success.

If one of these factors is missing, one lacks the confidence necessary for carrying out the action. The solution is obvious: make it a goal to supply the missing factor. We can capture the principle as the interpretive principle governing the resolution of questions:

If $CG \models W_s p$ and
 $CG \models \diamond q \rightarrow \diamond p$ and
 $CG \models W_s X$ and
 X can be interpreted as q then interpret X as q .

5 Towards a Model

What we are dealing with in this model are three future oriented operators: *want*, *scheduled* and *must*. The last one is a fairly standard modal operator.

We assume a linear order T and a set of possibilities meeting the following demands for $t \in T$, p a propositional letter and x a subject.

1. $i(t)(p) \in 2$.
2. $i(t)(\square) \subset POSS$
3. $i(t)(B_x) \subset POSS$
4. $i(t)(W_x) \subset i(t)(B_x) \times i(t)(B_x)$
5. $i(t)(P_x) \subset i(t)(B_x)$

In addition it must hold that if $j \in i(t)(\square)$ then $\forall t_1 < t \forall Z \in range(i(t)) i(t_1)(Z) = j(t_1)(Z)$.

B_x must be euclidean, also with respect to the extra structure supplied by W_x and P_x .

$i(t)(W_x)$ must be an irreflexive, asymmetric and transitive preorder.

$j \in i(t)(P_x)$ must be less than $k \notin i(t)(P_x)$ according to $i(t)(W_x)$.

An information state is pair consisting of an element of t and a set of possibilities. A common ground is an information state meeting the common ground conditions: $CG \models B_h \varphi \wedge B_s \varphi \rightarrow \varphi$ and $CG \models \varphi \Rightarrow CG \models B_s \varphi \wedge B_h \varphi$.

We define:

1. $i, t \models p$ iff $i(t)(p) = 1$
2. $i, t \models \neg \varphi$ iff $i, t \not\models \varphi$

3. $i, t \models \varphi \wedge \psi$ iff $i, t \models \varphi$ and $i, t \models \psi$
4. $i, t \models \text{Box}_x \varphi$ iff $\forall j \in i(t)(B_x) \exists t_1 > t, t_1 \models \varphi$
5. $i, t \models B_x \varphi$ iff $\forall j \in i(t)(B_x) j, t \models \varphi$
6. $i, t \models W_x \varphi$ iff $\forall j, k \in i(t)(B_x) (\exists t_1 (t \leq t_1 \wedge j, t_1 \models \varphi \wedge \neg \exists t_1 (t \leq t_1 \wedge k, t_1 \models \varphi) \rightarrow \langle j, k \rangle \in i(t)(W_x))$
7. $i, t \models P_x \varphi$ iff $\forall j \in i(t)(P_x) \exists t_1 (t \leq t_1 \wedge j, t_1 \models \varphi$

All of these are normal pointwise operators and they give only rise to a slightly different interpretation when transported to a specific information state such a common ground.

References

- Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Dissertation, University of Amsterdam.
- Stalnaker, R. (1979). Assertion. In: P. Cole (ed.). *Syntax and Semantics, Vol. 9, Pragmatics*. New York: Academic Press, pp. 315-332.
- Zeevat, H. (1997). The Common Ground as a Dialogue Parameter. In: A. Benz & G. Jaeger (eds.) *Proceedings of Mundial 1997*. CIS, University of Munich. München, p. 195-214.

Accepted Papers

Uncertain Reasoning About Agents' Beliefs and Reasoning, with special attention to Metaphorical Mental State Reports*

John A. Barnden
School of Computer Science
University of Birmingham
Birmingham B15 2TY, U.K.
J.A.Barnden@cs.bham.ac.uk

Abstract

In handling dialogue, or natural language discourse more generally, it is important for a system to reason uncertainly about the beliefs and the reasoning of agents. An implemented system is described that accomplishes this, thoroughly integrating an uncertain form of simulative reasoning into a general uncertain-reasoning framework. The system can also reason (uncertainly) on the basis of metaphorical utterances about mental states. Such utterances are common in mundane discourse.

1 Introduction

The paper takes it for granted that, for the pragmatics of dialogue, it is important for participants to reason about each other's beliefs and reasoning acts supporting those beliefs, and about the beliefs and reasoning of agents mentioned in the dialogue. This requirement is widely recognized, but less attention has been paid to the complications introduced by the dire need to reason *uncertainly* about agent's mental states and processes. Although some past research in some way brings together uncertainty and reasoning about agents (e.g., Asher & Lascarides 1994, Chalupsky 1996, Parsons *et al.* 1998), much remains to be done to achieve a full integration.

This paper is centred on an implemented AI reasoning system, ATT-Meta, that goes a long way towards such an integration. It can reason about agent's beliefs and reasoning to any depth of nesting of agents, and this reasoning is fully integrated with a (qualitative) handling of uncertainty. It can thus reason uncertainly about other agents' uncertain reasoning other agents' ... Its central technique for reasoning about beliefs and reasoning is the well-known method of *simulative reasoning*. The system's uncertain simulative reasoning is fully and richly integrated with uncertain non-simulative reasoning.

A further aim of ATT-Meta is to handle mundane metaphorical utterances, especially ones involving metaphors of mind. Metaphors generally, including metaphors of mind, are recognized within Cognitive Linguistics as being important in mundane discourse. ATT-Meta's metaphor-based reasoning is sketched towards the end of the paper, and the way ATT-Meta can handle uncertainty in that reasoning is emphasized. However, most of the paper is about non-metaphorical matters. There is more information on the metaphorical aspects in Barnden *et al.* (1994, 1996) and Barnden (in press).

*This work was supported in part by grant number IRI-9101354 from the National Science Foundation of the U.S.A.

To give a precise and full account of the system would be too lengthy, so the paper concentrates on the underlying principles and on various intricate problems that arise once uncertainty is given due attention. In any case, the issues are rather independent of the particular style of underlying representation and reasoning used in the system. A caveat: ATT-Meta is merely a reasoning system, and does not itself deal with natural language input directly. Rather, a user supplies hand-coded logic formulae that are intended to couch the (literal) meaning of small discourse chunks (two or three sentences). Another caveat: the purpose of the paper is merely to provide reasoning "infrastructure," and does not show in any broad way how the infrastructure might be used in discourse processing.

2 Uncertainty in Belief Reasoning

Suppose an agent X is reasoning about the beliefs and attendant reasoning of an agent Y . X could be an AI system itself, or it could in turn be the target of reasoning being done by a further agent W . Note that X 's reasoning about Y 's beliefs can rely both on reasoning about Y 's reasoning and on matters other than Y 's own reasoning. In the latter case, X does uncertain "Y-external" reasoning about Y : for instance, X may think Y possesses or lacks a belief because of the sort of person Y is, or a reliable informant Z may have told X that Y possesses or lacks a belief.

Altogether, there are various ways in which X 's reasoning about Y usually needs to be uncertain.

- **Uncertainty within Ascribed Reasoning Itself:** Y 's reasoning (alleged by X) may itself be uncertain. For instance, if Y is allegedly reasoning about Peter, where Y allegedly believes Peter to be a bird, then X should allow that Y may only uncertainly conclude that Peter can fly, because it is only a default generalization that birds fly. Equally, X might want to ascribe more complex forms of uncertain reasoning to Y , such as abduction, induction or analogy-based reasoning.
- **Uncertainty about Performance of Reasoning Steps:** Even if it were certain that an agent Y believed some propositions P_i , where some Q follows from the P_i by some argument (certain or uncertain), it is not certain that Y would believe Q , because Y might not undertake all the reasoning steps needed to go from the P_i to Q . Usually, X can at most take it as a working assumption that Y performs particular inference steps, such as inferring that something can fly on the basis that it is a bird.
- **Possibility of Y-Internal Counter-Arguments:** Even if X does allege that Y undertakes an argument to Q like that in the previous point, X must still allow that there is some other argument that Y might pursue and that would defeat the former argument. The counter-argument may not even be one that X is equipped to realize that Y is capable of undertaking, either because of the style of the reasoning steps, or because X is unaware that Y believes some of the premises of the argument. In the bird example, for all X knows, Y might believe that Peter is a penguin, and might therefore believe that Peter cannot fly.
- **Possibility of Y-External Counter-Arguments:** Whether or not X thinks Y has pursued an argument that culminates in a proposition Q , X might do Y-external reasoning to the effect that it is not the case that Y believes Q . X might then need to resolve the conflict between the evidence accruing from the alleged Y-internal reasoning and the evidence from the Y-external reasoning. For instance, in the bird example, X might be told by Z that Y fails to believe that Peter can fly. Z 's statement may be expressly uncertain, and in any case X must usually be uncertain about whether Z is telling the truth.
- **Uncertainty in Premise Ascription:** Whether Y believes the premises P_i mentioned above is itself generally uncertain, and can depend partly on Y-external reasoning. Much as in the previous point, an informant Z may have told X that Y believes that Peter is a bird.

3 Uncertain Belief Reasoning: Complications

A Violent Combinatorial Explosion

Suppose one believes with certainty level λ that agent Y believes with certainty level λ' that something is the case. Then λ and λ' are logically independent from each other. For instance, one can be fairly certain that Y believes P with low certainty, or one can believe with low certainty that Y is fairly certain that P . Any combination of levels of certainty is possible here. And the believed proposition P may itself be about the belief of some agent: the level λ'' involved in *that* belief is independent of the other two levels λ and λ' .

Actually, we can couch our discussion in terms of propositions of the form *Y believes to AT LEAST level λ'* . This is the tack taken in ATT-Meta, so I will mainly stick to it here, but nothing fundamental hangs on the move. The independence of the uncertainty measures appearing at different layers of belief still holds.

The real complication that this independence of certainty levels leads to is as follows. Consider, say, the two propositions, *Y believes P with at least low certainty* and *Y believes P with at least a high level of certainty*.

The levels of certainty one might have about these two propositions are to some extent independent. Of course, to be consistent, one should have at least as much certainty in the former as in the latter. But there is no reverse influence, and one can consistently believe the former with high certainty while believing the latter with low certainty. Also, the negations of such propositions may need to be considered during reasoning.

To generalize, an agent X may need to entertain several coexisting propositions of the form *Y believes P with at least such-and-such a level of certainty*. Similarly, an agent X may need to entertain several different propositions about Y's level of belief in $\neg P$ (where \neg indicates negation). All these propositions can have unrelated lines of evidence for them. For example, X might have strong evidence from considering Y's reasoning that Y believes P with at least low certainty, weak evidence from Y-external reasoning of some sort that Y believes P with at least reasonably high certainty, and strong evidence from natural language input that Y believes $\neg P$ with at least low certainty (e.g., a reliable informant may have told X that Y has a suspicion that $\neg P$ is the case).

We have been discussing propositions P' that X might hold about Y's beliefs with regard to P . But similar considerations hold one level up about the propositions another agent W might entertain about X's belief with regard to *any one* such P' . That is, W may need to entertain several different propositions of the form *X believes P' to at least such-and-such a level of certainty*. This holds for each different P' . Clearly, as one increases the number of levels there is a severe combinatorial explosion of the set of propositions that may be of interest somewhere within the overall system.

The explosion greatly expands the task of reasoning about beliefs beyond what is normally assumed, and/or the design task of optimizing the system in order to limit the explosion. I stress that it is uncertainty of belief that is to blame for the severity of the problems, even though in the case of all-or-nothing belief we do get the danger of a smaller explosion, with a maximum multiplying factor of only 2 (though that's bad enough). This smaller explosion is generated as follows: if we have three agents, say, then we have propositions of the form *W believes that X believes that Y believes P* , *W believes that X believes that Y believes $\neg P$* , *W believes that it's not the case that X believes that Y believes P* , and so on. However, in an all-or-nothing system there is no need to consider the negation of a hypothesis H that has been proven, unless the system is trying to check its own consistency, or is considering the possibility of another agent being inconsistent. On the other hand, in general, in order to finally conclude that some *non-certain* hypothesis H is strongly supported, one must look at the negation of H in order to check that the evidence for that negation is not stronger than the evidence for H itself.

Conflict Within and Across Belief Layers

When there is strong evidence for a proposition P as well as its negation $\neg P$, there is a conflict resolution problem: how to decide, if it is possible to do so, which of P and $\neg P$ should “win.” This is a classical problem arising in uncertain reasoning systems in AI (see, e.g. Pollock 1994), irrespective of the issue of reasoning about beliefs. Whatever method is used to address the general problem, special complications arise when we turn to uncertain belief reasoning.

Consider again the question of X 's reasoning about Y 's beliefs in P and in $\neg P$. According to X , agent Y may have strong lines of reasoning supporting P and $\neg P$. Therefore, X has strong evidence that Y strongly believes P , and X has strong evidence that Y strongly believes $\neg P$. Moreover, from the latter proposition, X can strongly infer that Y lacks a strong belief in P . But then X 's propositions are themselves in conflict. The conflicting propositions are:

- (a) Y has a strong belief in P
- (b) Y lacks a strong belief in P .

If X assumes that Y does not believe both P and $\neg P$ strongly, X could seek to reason about Y 's resolution of that conflict. However, it may not be legitimate to “descend into Y ” in this way in order to resolve the conflict, because X may have Y -external arguments for or against propositions (a) and/or (b) above. So, the general situation is that the various propositions at the X level, about Y 's beliefs, are supported by a combination of X 's Y -external evidence and X 's evidence about Y 's own reasoning. Perhaps part of the former evidence is strong enough to defeat the evidence that Y takes one of the lines of reasoning that Y might otherwise be plausibly thought by X to take. Conflict resolution within Y should only be reasoned about in certain circumstances, made clear in the discussion below of ATT-Meta's approach.

As an example, suppose X thinks that Y has arguments for and against the hypothesis that Roger is the culprit in some murder scenario. Suppose that other people have made statements to X about whether or not Y possesses a belief that Roger is the culprit, and/or about whether or not Y possesses a belief that Roger is not the culprit. X must somehow resolve the conflicting evidence about Y 's possession of these beliefs, where the evidence comes both from outside Y and from considering Y 's own arguments.

4 ATT-Meta's Basic Reasoning

ATT-Meta is a rule-based reasoning system. Reasoning goals and provided facts are both called *hypotheses* for convenience. Hypotheses are actually terms in a first-order, episode-based logic akin to that of, say, Hobbs (1990), but for simplicity of presentation they will be portrayed here as formulas in a first-order modal logic, and in fact the basic style of representation is not important for the purposes of the present paper. (See Barnden *et al.* 1994 for some representational detail.) ATT-Meta has no facilities for reasoning about time and change, even though in examples hypotheses expressed in the past tense may be used in the interests of naturalness, and even though change of belief is a very important consideration for dealing with natural language discourse.

Hypotheses may contain variables, but this possibility is not considered in this paper (for the sake of brevity). A fact or goal can involve negation, conjunction and disjunction, but in this paper we will only touch on negation. The reader may assume that for each goal entertained, the *complement* is also entertained, even though this is only roughly true. Two hypotheses are complementary when one is the negation of the other.

ATT-Meta deals only in qualitative measures of uncertainty, as opposed to, say, probabilistic measures. This is in part a simplification imposed to make the project more manageable, and in part reflects a claim that qualitative uncertainty is more appropriate for some purposes, notably some aspects of natural language understanding. Arguing this matter is beyond the scope of the current paper (but see Barnden 1998). At any time any particular hypothesis is tagged with a *qualitative certainty level*, one of certain, presumed, suggested, possible or certainly-not. The intuitive meanings are as follows. Certain: The hypothesis is true without reservation.

Certainly-not: The complement is certain. **Presumed:** the hypothesis is a default: i.e., it is taken as a working assumption, pending further evidence. **Suggested:** there is evidence for the hypothesis, but it is not (yet) strong enough to enable the hypothesis to be presumed. **Possible:** No evidence at all has yet been found for the hypothesis, but the negation is not certain.

A rule in ATT-Meta is of familiar form, having an *antecedent* and a *consequent*, but it also has a *qualitative certainty level*. We can view rules as being in the following format:

```
IF <antecedent component 1> AND ... AND <antecedent component N>
THEN [<certainty level>] <consequent>
```

The consequent is of the same form as a goal, and may contain variables. Each component of the antecedent is an expression in goal form. Implicitly, all variables in the rule are universally quantified, outside the whole rule. The certainty level of a rule is one of *suggested*, *presumed*, or *certain*.

ATT-Meta applies its rules in a backchaining style. It is given a reasoning goal, and uses rules to generate subgoals in the normal way. Goals can of course also be satisfied by provided facts. When a rule application successfully supports a goal, it supplies a level of certainty to it, calculated as the minimum of the rule's own certainty level and the levels picked up from the hypotheses satisfying the rule's antecedent. (The rule only supplies support to the goal in this way if all those certainty levels are at least *suggested*, because every hypothesis starts off as *possible* anyway.) When several rules support a hypothesis, the maximum of their certainty contributions is taken. Minimum and maximum are with respect to the obvious ordering of the certainty levels. Currently, ATT-Meta tries to apply all rules to all goals. Also, all rules are currently regarded as believed by all agents. This simplification will be corrected in future system versions.

Note that if a rule contributes a certainty level of *presumed* to a goal, the rule is trying to establish the goal as a working hypothesis, i.e. a default conclusion. Therefore, the rule can be said to be a type of *default rule*. However, unlike default rules in some other schemes, e.g. default logic (Reiter 1980), there is no justification part that acts as a gate. (Nevertheless, there is a roundabout way in which the effect of justification parts can be obtained. Rules in future versions of the system may be given justification parts.)

If both a hypothesis and its complement are supported to level at least *presumed*, conflict-resolution takes place. The most interesting case is when both hypotheses are exactly *presumed*. In this case, the system attempts to see whether one hypothesis has *more specific* evidence than the other. The former is then the "winner." If ATT-Meta determines a winner, it downgrades the certainty level of the loser to *suggested*, keeping the winner's at *presumed*. If it cannot determine a winner, it downgrades both to *suggested*. It is therefore conservative in its conflict resolution, avoiding arbitrary decisions between hypotheses. Specificity comparison is a commonly used heuristic for conflict-resolution in AI (e.g., Loui *et al.* 1993)—although serious problems remain in coming up with adequate and practical heuristics—and has as a special case the ancestor-closeness heuristic central to most semantic networks. The details of ATT-Meta's specificity-comparison are not important for understanding the rest of the paper. (An early version is detailed in Barnden *et al.* 1994.)

5 ATT-Meta's Belief Reasoning

ATT-Meta has facilities for reasoning about the beliefs and reasoning acts of agents, including where those beliefs and acts are about the beliefs and reasoning for further agents, and so forth. The implemented ATT-Meta system allows for the types of uncertainty listed in section 2.

Although ATT-Meta can reason about beliefs in an ordinary rule-based way, its main tool is *simulative reasoning* (widely used in AI — see, e.g., Creary 1979, Haas 1986, Chalupsky 1996; see also related work in philosophy and psychology, e.g. in Davies & Stone 1995). In attempting to show that agent Y believes P from the fact that Y believes Q, the system puts P as a goal and Q as an established hypothesis in a *simulation cocoon* for Y, which is a special environment which is meant to reflect Y's own reasoning processes. Reasoning from Q to P in the cocoon is alleged (by

default) to be reasoning by Y. The reasoning within the cocoon can involve ordinary rule-based reasoning and/or simulation of other agents. In particular, the reasoning can be uncertain. Also, the result of the simulation of Y is itself uncertain: even if the simulation supports the hypothesis that Y believes P, ordinary rule-based reasoning may support the negation more strongly.

An *atomic belief hypothesis* is a formula of the form $B(a, \phi, \lambda)$, where B is a modal operator, a is an individual constant denoting the agent, ϕ is a hypothesis formula, and λ is a certainty level greater than or equal to possible. Intuitively the hypothesis means that the agent believes ϕ with certainty AT LEAST λ . Because of the λ argument, the system expresses uncertainty in every layer of belief. Hypotheses involving belief take part in rule-based reasoning just as other hypotheses do (with an implementational optimization concerning the λ arguments in belief hypotheses).

Rule-based reasoning involving belief hypotheses is not intended to cater for reasoning that is alleged to be performed by other agents. Simulative reasoning is implemented for this purpose. Whenever a hypothesis that is an atomic belief hypothesis is created, the hypothesis formed by stripping off the application of the belief operator B is also formed (unless it already exists), and placed inside the "simulation cocoon," or "reasoning space," for the agent in question (similar to the belief spaces in many other authors' work). This happens recursively in the case of nested belief hypotheses, so that we get nested reasoning spaces (nested cocoons). Conversely, whenever a goal is created inside an agent's reasoning space, an application of B is wrapped around it and placed in the immediately surrounding space (which may be the system's own).

Although discussions of simulative reasoning often mention that the output of a simulation (i.e., a conclusion to the effect that the simulated agent Y believes some Q) must be uncertain and subject to possible defeat by other evidence, researchers have not said much in detail about how this might happen, and do not appear to have observed that there are much more far-reaching implications for the realization of simulative reasoning. In particular, discussions of simulation hardly ever go into the question of complex forms of uncertain reasoning within agents. But this is a key point in the present paper.

For the purposes of later sections, two technicalities need to be explained. First, suppose that ordinary rule-based reasoning supports an atomic belief hypothesis $B(a, \phi, \lambda)$ to level at least presumed, or that this hypothesis is a fact with level at least presumed. Then, within the reasoning space for a , hypothesis ϕ is recorded as a *fact*, although, as with any fact, it can also have rule-based support. It is also important to realize that any two facts are incomparable as regards specificity.

Secondly, each rule-application within a non-system reasoning space is "lifted" to the next reasoning-space out. That is, consider a hypothesis $B(a, \phi, \lambda)$ within a reasoning space R , and hypothesis ϕ within R_a , the reasoning space for a nested within R . For each rule-application α supporting ϕ within R_a , a "lifted rule application" ($\Lambda(\alpha)$) is created as a support for $B(a, \phi, \lambda)$ within R . Let $\Delta(\rho)$ be the set of hypotheses directly used by a rule application ρ . Then, $\Delta(\Lambda(\alpha))$ is the set of hypotheses formed by wrapping $B(a, \dots)$ around each member of $\Delta(\alpha)$, plus an extra "agent-inference hypothesis" that will not be mentioned further in this paper (but see Barnden 1998). The system always takes the strength of the imaginary rule of which $\Lambda(\alpha)$ is an application to be presumed. This value is important for coping with the "Possibility of Y-Internal Counter-Arguments" in section 2.

Such lifted rule applications account for almost all the direct effect that the simulation of a (i.e., the reasoning within R_a) has on the reasoning about $B(a, \phi, \lambda)$ within space R . Importantly, *lifted rule applications for a hypothesis are treated in the same way as ordinary rule applications for the hypothesis in respect of certainty-value accumulation and conflict resolution involving the hypothesis*. This principle supplies a valuable amount of modularity in the design of the system.

Finally, lifted rule applications are themselves subject to lifting just as any ordinary application is. Thus, if X is being simulated by another agent W, the lifted applications in the X layer resulting from rule-applications within the Y-layer are themselves lifted to W's own layer.

6 Coping with the Complications

Coping with the Hypothesis Explosion

In ATT-Meta, there is a danger of multiplying the number of propositions by eight for each layer of belief. Suppose the system is considering John's belief state about Mary's belief state concerning the proposition that Peter can fly. There are two propositions of interest "within Mary" (namely the proposition that Peter can fly and its negation), possibly 16 of interest within John, and possibly 128 at the system's own level. The factor of 8 arises because four different levels of certainty λ can be used in a belief hypothesis $B(a, \phi, \lambda)$, and the negations of these hypotheses are also of interest.

Fortunately, by means of some optimizations (discussed in Barnden 1997) that economize on the number of propositions that are explicitly represented in the system, the multiplying factor in ATT-Meta is reduced in practice to about 1.7 or less in cases of typical complexity (and as low as 1.17 in especially favourable cases). One of the mentioned optimizations in ATT-Meta is to exploit the possibility of ignoring the negations of hypotheses whenever it is apparent that doing so would not contribute to the overall reasoning goal. Another is, under suitable circumstances, to avoid maintaining separate hypotheses $B(a, \phi, \lambda)$ for a given a and ϕ but different values of λ , by attaching the body of evidence for each of these hypotheses to just one such hypothesis (and treating the different bodies differently during reasoning).

Coping with Conflict Within and Across Belief Layers

The considerations of the second part of section 3 suggest that conflict-resolution can be needed within simulations as well as at the top layer of reasoning. This raises the difficult question of how to coordinate conflict-resolution on different layers. Suppose X is (allegedly) simulating Y , and is investigating

$B(Y, P, \text{presumed})$
 $\neg B(Y, P, \text{presumed})$
 $B(Y, \neg P, \text{presumed})$
 $\neg B(Y, \neg P, \text{presumed})$.

X could be the system itself, or some agent at an intermediate layer in a tower of simulation layers. Within the simulation of Y , the relevant hypotheses are P and $\neg P$. There can be rule applications supporting hypotheses both within in the Y layer and within the X layer. Below I will assume that the rule applications for each of the hypotheses taken individually would give that hypothesis a rating of presumed were it not for conflict with negations. Henceforth, for brevity, I will suppress explicit mention of uncertainty levels within Y , and accordingly omit the uncertainty argument from applications of the B operator.

Technically, a battle between $B(Y, P)$ and $\neg B(Y, P)$ is separate from a battle between $B(Y, \neg P)$ and $\neg B(Y, \neg P)$. However, ATT-Meta has a rule that says that if Y believes something Q then, presumably, Y lacks the belief that $\neg Q$. This rule is relevant here, with Q being P or $\neg P$. So the two battles actually share whatever evidence there is for $B(Y, P)$ and $B(Y, \neg P)$. In the battle between $B(Y, P)$ and its negation, the evidence for the latter includes that for $B(Y, \neg P)$, and similarly for the other battle.

The simplest type of case is where, actually, there is no evidence within the Y layer supporting either P or $\neg P$. Then, conflict resolution can proceed in the normal way between $B(Y, P)$ and $\neg B(Y, P)$, and also between $B(Y, \neg P)$ and $\neg B(Y, \neg P)$.

The opposite type of case is when in the X layer there is no Y -external rule-based argument for or against the hypotheses that $B(Y, P)$ and $B(Y, \neg P)$. Unfortunately, this does not mean that conflict-resolution can necessarily be done in the Y layer. Consider any fact F that is in the support for P or $\neg P$, within the simulation. The corresponding X -layer hypothesis $B(Y, F)$ is either a fact or has rule-based support within that layer. Therefore this non-simulative evidence is, via the simulated reasoning connecting F to P , indirect evidence also for one or more of $B(Y,$

P) and so on. As a result, the situation is as complex as when there is *direct* non-simulative evidence for one or more of $B(Y, P)$ and so on.

We now proceed to describe the general approach that ATT-Meta takes, and that copes with all the types of situations we have considered. Basically, ATT-Meta uses the rule-lifting technique described in section 5 to lift the evidence in the Y layer up into the X layer, in preparation for *possibly* doing all the conflict-resolution in the X layer. The lifted evidence is used in an algorithm that decides whether to do conflict-resolution in the X layer, or whether it should be done in the Y layer. (Of course, different decisions will in general be made for different hypotheses P .)

Suppose ATT-Meta is considering the competing hypotheses $B(Y, P)$ and $\neg B(Y, P)$. Then ATT-Meta does conflict-resolution in the X layer between these hypotheses if and only if at least one of these hypotheses has "unlowerable support." A hypothesis H has unlowerable support if and only if H itself or some hypothesis directly or indirectly supporting H via lifted rule applications is non-"finalized" and has a non-lifted rule-application supporting it (or is a fact). A hypothesis is finalized if all decisions, including ones emanating from conflict-resolution, have been done for it. The (complex) process by which finalization is determined will not be detailed here.

If at least one of $B(Y, P)$ and $\neg B(Y, P)$ has unlowerable support, ATT-Meta seeks to resolve the conflict between these directly, rather than going down into the Y simulation. ATT-Meta applies the ordinary specificity comparison algorithm alluded to in section 4 to these contending hypotheses – *it involves no special treatment of lifted applications or hypotheses about belief*. As normal, if a winner is found, the other contender is downgraded; if not, both contenders are downgraded. A similar process occurs with $B(Y, \neg P)$ and $\neg B(Y, \neg P)$. Downgrade of, say, $B(Y, P)$ causes the rule applications supporting P within the Y simulation to be suppressed, causing P itself to go down in certainty, so that there is no need for conflict resolution between P and $\neg P$. (Special action is taken in the rare cases when $B(Y, P)$ and $B(Y, \neg P)$ both win their battles.)

If, on the other hand, neither $B(Y, P)$ nor $\neg B(Y, P)$ has unlowerable support, those hypotheses are *not* downgraded, and instead ATT-Meta will descend into the Y simulation and consider the conflict between P and $\neg P$. The resolution of this conflict will then indirectly cause the conflicts in the X layer to be resolved.

We have only been considering two layers, the X layer and the Y layer. But these could be intermediate in a tower of layers: for instance, P could be about a belief of some agent, and/or the X layer may be a simulation layer within another layer. ATT-Meta proceeds by considering conflicts in the top layer, resolving them when appropriate and possible, then moving down to the next layer, and so on.

Concerning a Specific Example

An extended example of ATT-Meta's belief reasoning is treated in detail in Barnden (1998). It is based on a real-life mugging event and illustrates various possibilities for how conflict resolution can proceed across belief layers. In the mugging event, the victim, whom I will call Vic, says to the police that he believes the mugger seemed to be trying *not* to hurt him. This is in contrast to the rule that muggers normally try to hurt their victims. This rule is (as always, currently) taken to be common knowledge. Taking the police and Vic to be X and Y respectively, ATT-Meta is able to sort out the situation by doing conflict-resolution outside Y, using Y's statement about the non-hurting to indirectly cause the defeat of the belief that X would otherwise have that Y believes that the mugger was trying to hurt him, and allowing X to conclude that Y believes that the mugger was not trying to hurt him. When the example is varied so as to have Y stating that the mugger is a non-violent person, so that the idea that the mugger does not try to hurt Y is merely an inference Y can make from the non-violent property, it turns out that ATT-Meta does the conflict-resolution *within* Y. However, the overall effect is the same as before.

7 Metaphor in ATT-Meta

We turn now to the remaining main theme of this paper, namely metaphor. ATT-Meta is partly aimed at the reasoning needed to extract useful information from metaphorical utterances in mundane natural language discourse. A metaphorical utterance is one that *manifests* (instantiates) a metaphor, where a *metaphor* is a conceptual view of one topic as another. Here I broadly follow Lakoff (e.g., Lakoff 1993). An example of a metaphor is the view of the mind as a three-dimensional physical region (MIND AS PHYSICAL SPACE). A metaphor is the view itself, as opposed to some piece of natural language that manifests the view. Such a manifestation might be “John believed in the recesses of his mind that ...,” in the case of MIND AS PHYSICAL SPACE. In a manifestation, the topic actually being discussed (John’s mind, in the example) is the *tenor*, and the topic it is metaphorically cast as (physical space, in the example) is the *vehicle*.

ATT-Meta does not currently deal with novel metaphors — rather, it has pre-given knowledge of a specific set of metaphors. But it is specifically designed to handle novel *manifestations* of those metaphors. Its knowledge of a metaphor consists mostly of a relatively small set of very general “conversion rules” that map between the vehicle and tenors. The degree of novelty the system can handle in a manifestation of a metaphor is limited only by the amount of knowledge it has about the vehicle and by the generality of the conversion rules.

The ATT-Meta research has concentrated on a specific type of metaphor, namely metaphors for mental states (and processes), such as MIND AS PHYSICAL SPACE, although the principles and algorithms implemented are not restricted to metaphors of mind. Mundane discourses, such as ordinary conversations and newspaper articles, often use metaphor in talking about mental states/processes of agents (mainly people). Indeed, as with many abstract topics, as soon as anything subtle or complex needs to be said, metaphor is practically essential. There are many mental-state metaphors apart from MIND AS PHYSICAL SPACE. Some are as follows: COGNITION AS VISION, as when understanding, realization, knowledge, etc. is cast as vision; IDEAS AS INTERNAL UTTERANCES, which is manifested when a person’s thoughts are described as internal speech or writing (internal speech is not *literally* speech); and MIND PARTS AS PERSONS, under which a person’s mind is cast as containing several sub-agents with their own thoughts, emotions, etc. Many real-discourse examples of manifestations of metaphors for mental states and processes can be found in databanks on the web (notably in the author’s own databank, <http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank>).

8 Metaphor-Based Reasoning: Principles

Notoriously, metaphorical utterances can be difficult if not impossible to paraphrase in non-metaphorical terms. Equally, it can be difficult if not impossible to give them internal, non-metaphorical, meaning representations. Consider, for instance. “One part of John was insisting that Sally was right.” This manifests the metaphor of MIND PARTS AS PERSONS, the mentioned part engages in natural language utterance—the insistence—so that we also have IDEAS AS INTERNAL UTTERANCES being applied to John.) I claim that we simply do not know enough about how the mind works to give a full, definite, detailed account of what was going on in John’s mind according to the sentence. After all, what non-metaphorical account can be given of some “part” of John “insisting” something? Rather, the utterance connotes things such as that John had reasons both to believe that Sally was right and to believe the opposite.

This particular connotation arises from the observation that someone usually insists something only when someone else has stated the opposite. So, the sentence suggests that some other “part” of John stated, and therefore probably believed, that Sally was *not* right. Then, because of the thoughts of the two sub-agents with John (the two parts), we can infer that John had reasons to believe the mentioned things about Sally.

Some investigators may wish to call such an inference (part of) the underlying *meaning* of the utterance. The ATT-Meta research project has refrained from this step, and only explicitly countenances *literal* meanings for metaphorical utterances. (The literal meaning of the above

utterance is the ridiculous claim that John literally had a part that literally insisted that Sally was right.) However, the project presents no objection to the step. Thus, we can say that ATT-Meta is "semantically agnostic" as regards metaphor. (The approach is akin to but less extreme than that of Davidson (1979), which can be regarded as semantically "atheist.")

ATT-Meta's approach is one of *literal pretence*. A literal-meaning representation for the metaphorical input utterance is constructed. The system then *pretends* that this representation, however ridiculous, is true. Within the context of this pretence, the system can do any reasoning that arises from its knowledge of the vehicles (or tenors) of the metaphors involved. In our example, it can use knowledge about interaction within groups of people, and knowledge about communicative acts such as insistence. As a result of this knowledge, the system can infer that the explicitly mentioned part of John believed (as well as insisted) that Sally was right, and some other, unmentioned, part of John believed (as well as stated) that Sally was not right. Suppose now that, as part of the system's knowledge of the MIND PARTS AS PERSONS metaphor, there is the knowledge that if a "part" of someone believes something P, then the person has reasons to believe P. The system can now infer both that John had reasons to believe that Sally was right and that John had reasons to believe that Sally was not right.

Note here that the key point is that the reasoning from the *literal* meaning of the utterance, conducted within the pretence, link up with the just-mentioned knowledge. That knowledge is itself of a very fundamental, general nature, and does not, for instance, rely on the notion of insistence or any other sort of communicative act. *Any* line of within-pretence inference that linked up with that knowledge could lead to conclusions that John had reasons to believe certain things. This is the way in which ATT-Meta can deal with *novel manifestations* of metaphors. There is no need for it at all to have any knowledge of how *insistence* by a "part" of a person maps to some non-metaphorically describable feature of the person.

9 Metaphor: Implementation and Uncertainty

To implement a "literal pretence," ATT-Meta constructs a computational environment called a *metaphorical pretence cocoon*. The representation of the literal meaning of the utterance, namely, in our example, the hypothesis that a part PJ of John insisted that Sally was right, is placed as a fact L inside this cocoon. Corresponding to this, outside the cocoon, the system has a hypothesis (a fact) SL that it itself (the system) is pretending that L holds. Also, the system has the fact SPJ, outside the cocoon, that it is pretending that PJ is a person.

As usual, the system has a goal, such as the hypothesis that John believes that Sally is right. Assume the system has a rule that if someone Y has reasons to believe P then, presumably, Y believes P. Thus, one subgoal that arises is that John had reasons to believe that Sally was right.

Now, in the earlier section we referred to the system's knowledge about the MIND PARTS AS PERSONS metaphor. The mentioned knowledge is couched in a rule that can be paraphrased as follows:

IF I (the system) am pretending that part Z of agent Y is a person AND I am pretending that Z believes Q THEN (presumably) Y has reasons to believe Q.

We call this a *conversion rule*, as it maps between pretence and reality. Because of the subgoal that John had reasons to believe that Sally was right, the conversion leads to the setting up of the subgoal that the system is pretending that PJ (the mentioned part of John) believes that Sally is right. This subgoal is itself outside the cocoon, but it automatically leads to the subgoal that PJ believes that Sally is right, *within* the cocoon. This subgoal can then be inferred (as a default) from the hypothesis that PJ stated that Sally was right, which itself can be inferred (as a default) from the existing within-cocoon fact that PJ insisted that Sally was right. Notice carefully that these last two steps are *entirely within the cocoon and merely use commonsense knowledge about real-life communication*.

As well as the original goal (John believed that Sally was right) the system also looks at the negation of this, and hence indirectly at the hypothesis that John has reasons to believe that Sally

was *not* right. This subgoal gets support in a rather similar way to the above process, but it involves richer reasoning within the cocoon.

ATT-Meta incorporates a handling, at least partial, of various types of uncertainty in metaphor-based reasoning. First, the system can be unsure whether a metaphor holds, by having merely presumed as the level of certainty for a fact like SPJ above to the effect that the system pretends that part PJ of John is a person. This fact is then potentially subject to defeat in the ordinary way. Secondly, notice the "presumably" in the above conversion rule, indicating that its certainty level is presumed. It is possible for there to be evidence that is strong enough to defeat a conclusion made by the rule. Conversely, although there may be evidence against the conclusion of the rule, it may be weak enough to get defeated by the evidence for that conclusion. Thus, whether a piece of metaphorical reasoning overrides or fails to override other lines of reasoning about the tenor depends on the case at hand. Finally, the reasoning within the cocoon is itself usually uncertain, since commonsense knowledge rules are usually uncertain.

Because there is uncertain reasoning both within and outside the cocoon, complications arise for conflict resolution much as in the belief case. An extra one is that the pretence cocoon is taken to contain as a fact any fact sitting outside. This importation of facts is needed because arbitrary information about, say, physical objects may be needed in a pretence cocoon used for a metaphor like MIND AS PHYSICAL SPACE. Within the cocoon, the imported facts may support something that conflicts with conclusions drawn from the special metaphorical facts inserted into the cocoon at the start (e.g., the fact (PJ) that part PJ of John is a person). However, the system adopts the heuristic that such metaphorical facts supply added specificity. Therefore, ATT-Meta proceeds as follows: within a metaphorical pretence cocoon, specificity-comparison is first attempted in a mode where all reasoning lines partially dependent on imported facts are thrown away. Only if this does not yield a winner are those lines restored, and specificity-comparison attempted again.

10 Conclusion

The paper has explained some considerations addressed in the design of an implemented system for uncertain belief reasoning, ATT-Meta, based on simulative reasoning. Belief reasoning has to be much more complicated than is usually recognized. This comes from the necessary involvement of uncertainty handling, including conflict-resolution. Some possible types of uncertainty arise peculiarly from natural language discourse, whereas others may arise in any multiagent situation where agents are reasoning about each other, but particularly when they discourse with or about each other. Specialized needs for uncertain reasoning are raised by metaphorical utterances about mental states. These are common in ordinary conversation and writing, and ATT-Meta handles various types of metaphor-based reasoning.

The pretence-based metaphorical reasoning operates largely identically to simulative reasoning about belief. Consequently, the system readily allows nesting of metaphorical reasoning within other metaphorical reasoning, and nesting of metaphorical reasoning and belief-reasoning either way round. As a result, it can provide an approach to aspects of some important discourse phenomena, addressed briefly in Barnden (in press): chained metaphor, personification, and reasoning about other agents' own metaphorical thoughts.

References

- Asher, N. & Lascarides, A. (1994). Intentions and information in discourse. In *Procs. 32nd Annual Meeting of the Association for Computational Linguistics*, pp.34-41. Association for Computational Linguistics.
- Barnden, J.A. (1997). Simulation and uncertainty in reasoning about agents' beliefs. *Memoranda in Computer and Cognitive Science*, No. MCCS-97-310, Computing Research Laboratory, New Mexico State University, Las Cruces, NM 88003, U.S.A.

- Barnden, J.A. (1998). Uncertain reasoning about agents' beliefs and reasoning. Technical Report CSRP-98-11, School of Computer Science, The University of Birmingham, U.K. Revised version of invited submission to a special issue of *Artificial Intelligence and Law*, ed. E. Nissan, forthcoming.
- Barnden, J.A. (in press). An AI system for metaphorical reasoning about mental states in discourse. In Koenig, J-P. (Ed.), *Conceptual Structure, Discourse, and Language II*. Stanford, CA: CSLI/Cambridge University Press.
- Barnden, J.A., Helmreich, S., Iverson, E. & Stein, G.C. (1994). An integrated implementation of simulative, uncertain and metaphorical reasoning about mental states. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference*, pp.27-38. San Mateo, CA: Morgan Kaufmann.
- Barnden, J.A., Helmreich, S., Iverson, E. & Stein, G.C. (1996). Artificial intelligence and metaphors of mind: within-vehicle reasoning and its benefits. *Metaphor and Symbolic Activity*, 11(2), pp.101-123.
- Chalupsky, H. (1996). Belief ascription by way of simulative reasoning. Ph.D. Dissertation, Department of Computer Science, State University of New York at Buffalo.
- Creary, L. G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Proc. 6th. Int. Joint Conf. on Artificial Intelligence*, pp.176-181. Los Altos, CA: Morgan Kaufmann.
- Davidson, D. (1979). What metaphors mean. In S. Sacks (Ed.), *On Metaphor*, pp.29-45. U. Chicago Press.
- Davies, M & Stone, T. (Eds) (1995). *Mental simulation: evaluations and applications*. Oxford, U.K.: Blackwell.
- Haas, A.R. (1986). A syntactic theory of belief and action. *Artificial Intelligence*, 28, 245-292.
- Hobbs, J.R. (1990). *Literature and cognition*. CSLI Lecture Notes, No. 21, Center for the Study of Language and Information, Stanford University.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, 2nd edition, pp.202-251. New York and Cambridge, U.K.: Cambridge University Press.
- Loui, R.P., Norman, J., Olson, J. & Merrill, A. (1993). A design for reasoning with policies, precedents, and rationales. In *Fourth International Conference on Artificial Intelligence and Law: Proceedings of the Conference*, pp.202-211. New York: Association for Computing Machinery.
- Parsons, S., Sierra, C. & Jennings, N. (1998). Multi-context argumentative agents. In *Working Papers of the Fourth Symp. on Logical Formalizations of Commonsense Reasoning (COMMON SENSE '98)*, Queen Mary and Westfield College, London, 7-9 January 1998.
- Pollock, J.L. (1994). Justification and defeat. *Artificial Intelligence*, 67 (2), pp.377-407.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13, pp.81-132.

Eye-Movement Research and the Investigation of Dialogue Structure

Thomas Clermont and Marc Pomplun and Elke Prestin and Hannes Rieser
SFB 360 University of Bielefeld
Universitätsstr. 25
D-33615 Bielefeld
rieser@lili.uni-bielefeld.de

Abstract

The paper reports about eye-tracking research as it has been going on in different institutions at Bielefeld University. Initially, a short introduction into eye-tracker technology is given and the two scenarios used in the investigation of agents' behaviour in task-oriented-dialogue are presented: A 2D-setting working with one eye-tracker recording the visual attention of the instructor, and a 3D-setting using eye-tracking for the instructor as well as for the constructor. The 3D-setting is to our knowledge the first one that has been developed for the investigation of agents' actions in dialogue. Subsequently, it is shown that the investigation of situated language behaviour can profit a lot from the application of eye-tracker-technology: We present the results of three empirical studies, two 2D-studies and one using the 3D-setting. They reveal interesting interdependencies between focussing, planning, language production and the organization of larger units in discourse such as sequences of turns.

1 Why Eye-Tracking?

Eye-tracking is based on the idea that the eyes can be used as "windows to the brain". When we are looking at static scenes or pictures, our eye-movements are never performed continuously. In fact, the gaze position "jumps" between inspected locations. These quick jumps are called *saccades*, and the motionless phases in between are called *fixations*. Visual information can only be perceived during fixations, which is beneficial to the brain's job of constructing a stable image of the surroundings.

Referring to the idea of the eyes being indicators of the brain's performance ("windows to the brain"), this means that *fixation duration* can be considered as a measure of the effort of information processing. The longer our attention rests on a certain location, the longer it presumably takes us to deal with the visual information presented there. This relationship has been coined the "eye-mind" hypothesis (Just and Carpenter, 1987). It is strongly supported by results of reading research. The fixation duration of a subject reading written text clearly depends on the length of the currently fixated word and its frequency in the language. Since fixation duration is *not* influenced by the characteristics of the previously read word, the syntactic and semantic analysis of a word is evidently performed during its fixation. It is plausible to assume that the eye-mind hypothesis holds for other kinds of stimuli, e.g. pictures and real-world scenes, as well.

Furthermore, *saccade length* reveals how thoroughly a certain region of the visual field is scanned. Short saccades indicate that the fovea is directed to positions close to each other, signifying a "high resolution" scanning process. Long saccades entail that the local scene is only roughly perceived or that its information content is low. Fixation duration and saccade length are the basic eye-movement variables. Additionally, eye-movements yield data about where and in which temporal order a subject acquires visual information, i.e. eye-movements reveal a subject's distribution and dynamics of visual *attention*.

In many situations, eye-movements are controlled by factors operating on a high semantic level. For instance, we all know a wide variety of faces, buildings, streets etc. and thus have certain concepts of how to perceive them. This knowledge influences our eye-movement patterns. For example experiments with ambiguous pictures Pomplun et al. (1996) showed that the distribution of attention depends not only on the geometrical structure of the inspected stimulus, but also on its semantic interpretation by the subjects. Ambiguous pictures – like the famous “Necker Cube” – can be interpreted in two different ways. A subject’s perception of such a picture is likely to continuously switch between these alternatives. Pomplun et al. (1996) demonstrated that there are distinct patterns of eye movements during each of the two interpretations, showing that eye-movements are not only controlled by the geometrical structure of a stimulus, but also by the subject’s semantic interpretation of it.

Obviously, eye-movements give considerably more insight into mental processes than sheer manual responses, which are obtained in psychophysical standard experiments. As a consequence, eye-movements have been studied in many different fields of research:

Reading research: While reading written text, a subject’s eye-movements tell us the duration needed for processing a particular word. These data enable scientists to draw conclusions about the structure of language information stored in our brain.

Medical research: Eye-movement measurement can help physicians to diagnose certain diseases of the nervous system, e.g. schizophrenia or Parkinson’s disease, because these diseases lead to characteristic distortions of eye-movement parameters. Moreover, eye-movement analysis can provide information about the state of a patient’s healing process during his/her therapy.

Traffic research: A car driver’s eye-movements tell scientists which factors distract the driver’s attention and are thus likely to cause traffic accidents. The arrangement of instruments, for example, can be optimized with the help of these investigations.

Consumer research: It is important for advertising agencies to test the eye appeal of their commercial spots or brochures before launching a publicity campaign. Subjects’ eye-movements can indicate which parts of the spot or brochure attract most of the viewers’ attention. In particular, it can be investigated whether the name of the boosted product is shown in an adequate position to be recognized.

2 The Technical Basis of Eye-Tracking

The idea of using the eyes as “windows to the brain”, i.e. conducting eye-tracking experiments, is far from being novel. It was more than thirty years ago that first experiments of this kind took place. The pioneers of eye-tracking, however, imposed a severe restriction on the subject’s freedom of action: The subject’s head was being fixed. In most cases this was accomplished by an individually made bite bar on which the subject had to bite throughout the trial. Evidently, this experimental setting did not provide completely natural conditions. In particular, investigations of the relationship between eye-movements and speech production were impossible.

The 2D-Scenario

Fortunately, the Neuroinformatics Group at the University of Bielefeld owns two modern eye-trackers that do not require the subject’s head to be fixed. There are the systems *OMNITRACK1* and its successor *SMI EyeLink*, which use basically the same technique (for further details see Stampe 1993). As shown in Figure (1), the *OMNITRACK1* system utilizes a head-set to be worn by the subject. Two small digital infra-red cameras are fastened to this head-set, an *eye camera* pointing at the subject’s right eye, and a *head camera* which “looks” roughly in the same direction as the subject. The subject is placed in front of a computer monitor for stimulus presentation. Four infra-red LEDs (*light emitting diodes*) near the corners of the monitor screen are essential

for the “trick” of head movement compensation. The head camera yields an infra-red image of the subject’s field of view, and of course the four LEDs cause the only four light spots in the otherwise black image. The locations of these spots in the image and thus the relative position of the subject’s head with respect to the screen can effortlessly be calculated.

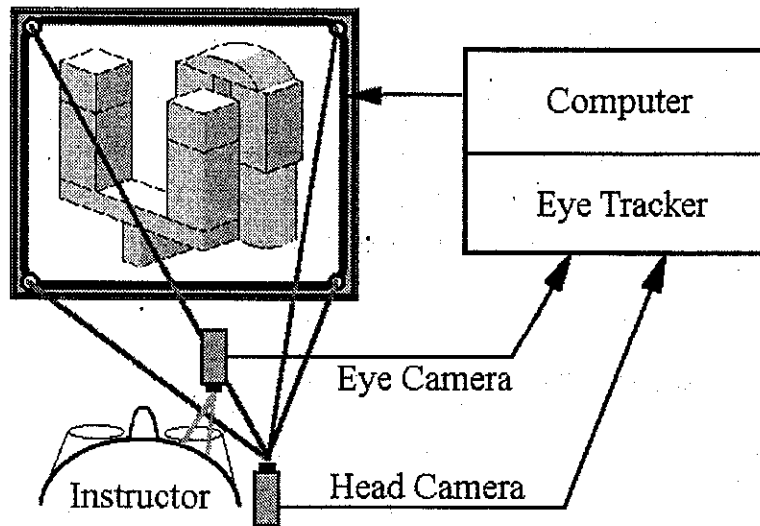


Figure 1: Measurement of the Instructor’s eye-movements on a 2D-image using the OMNITRACK1 device

The pupil position is derived from the eye camera image with the help of specialized hardware for fast image processing. Proceeding from the head and pupil positions, a non-linear projection leads to the gaze position coordinates on the screen. The parameters of this projection are gained from a calibration procedure which has to be performed before the start of any experiment. In this procedure, a dot is shown sequentially at nine different positions on the screen, and the subject has to track it visually. This comparably quick calibration leads to a spatial accuracy which ranges from 0.8 to 1.0 degrees of visual angle. With the help of a neural network interface introducing another calibration procedure, the average error can be reduced to 0.4 degrees (see Pomplun et al. 1994). Neither small head movements (maximal 10 degrees in the horizontal or vertical direction) nor speaking impairs the data accuracy. The temporal resolution of the OMNITRACK1 device is 16.7 ms due to the cameras’ frame rate of 60 Hz.

The *SMI EyeLink* device is based on the same principle as OMNITRACK1, but it presents several improvements. For instance, now there are *two* eye cameras, one for each eye, facilitating binocular eye-tracking. Furthermore, the frequency of measurement was increased to 250 Hz, which means that the actual gaze positions of both eyes can be measured every 4 ms. The spatial accuracy is the same as for OMNITRACK1.

The 3D-Scenario

The fact that there are two modern eye-trackers available in the same laboratory now enables us to investigate the foci of *both partners* during task completion. Thus, the *interaction* of the instructor’s and the constructor’s focus can be analysed. In the novel 3D-scenario, both partners can manipulate real objects while their eye-movements are recorded. Hence, the dynamics of foci can be studied within situations of higher ecological validity.

The 3D-setting is arranged in the following way: As in the 2D-“blocks world” study (see section 3.1), two subjects (instructor and constructor) take part at a time. They can neither see each other nor each other’s workspace; only verbal communication is allowed. Again, the instructor

is wearing the OMNITRACK1 head-set and is sitting in front of the monitor screen with the four infra-red markers for head alignment. The calibration procedure is still performed using nine calibration dots that are displayed on the screen. Instead of viewing a 2D-image of the object to be constructed, however, the instructor is holding a *real* 3D-object between his eyes and the monitor screen this time. In this scenario, the monitor is only needed for the calibration of the system and the compensation of head movements.

The instructor may rotate the object, for example a "Baufix" toy-plane, and view it from any perspective, as long as his visual axes hit the screen. This restriction is due to the fact that the OMNITRACK1 device does not calculate gaze positions that are located outside the screen area. As a matter of fact, this eye-tracking scenario brings about an important problem: Other than in the previous 2D-study (see fig. 1 above), the stimulus is *dynamic*. So how can the eye-movement data be analysed if the locations of elementary items are variable and not even "known" by the eye-tracker system?

In order to solve this problem, an additional camera is introduced. This static camera is placed behind the instructor and is "looking over his shoulder". It records the scene viewed by the instructor during an experiment. By using a neural network similar to the one applied in (Pomplun et al., 1994), the fixation data measured by OMNITRACK1 can be converted into the coordinate system of the static camera. This means that the converted fixations can be superimposed on the recorded video tape, indicating the subject's actual viewing direction in the three-dimensional scene. It is even possible to calculate a valid cluster analysis on these fixations if the duration of single clusters is kept short to limit the movements of items during this time.

At the same time, it is possible to obtain the constructor's eye movements with the help of the second eye-tracker. Since the calibration of the EyeLink system is not necessarily restricted to the monitor screen, we are able to observe the constructor's focus within a wide workspace on the desk in front of him. This is important, because the constructor has to survey a distribution of elementary items which he has to interconnect according to the instructions he receives from his partner. Hence, the constructor needs a considerably larger workspace—and thus eye-tracking space—than the instructor.

But how can the system be calibrated on a desk instead of the screen? For this purpose, we installed a 3×3 matrix of LEDs on the constructor's desk. These LEDs emit visible red light. During the calibration procedure, LEDs are successively switched on (one at a time), indicating the calibration point the constructor has to look at. Furthermore, the four infra-red LED markers for head alignment are located at the corners of the constructor's rectangular workspace. This method enables eye-tracking on a wide range on the desk without the compensation of head movements being impaired.

As with the instructor, the constructor's visual scene is "observed" by another static camera. A video showing eye-movements and fixation clusters can be derived in an analogous way. Figure (2) shows a scheme of the 3D-setting for co-operative experiments.

Since the EyeLink system provides *binocular* eye-movement data, it would be possible to calculate the constructor's *gaze depth*, i.e. the distance of the currently inspected item. Recent experiments in our group showed that the calculation of gaze depth on the basis of the eyes' vergence angle can be achieved with an accuracy of about 2 to 3 cm Essig (1998). In many cases, this would be sufficient to decide which of two objects located in almost the same visual angle is viewed by the constructor. The implementation of this technique is subject to current research.

The fixation and cluster data of the instructor and the constructor can be combined to a *single* video tape showing both scenes side by side. In addition, the transcript data can be shown as subtitles in the videos. This arrangement allows us to investigate interactions between the two visual foci, speech production and speech perception.

3 Linguistic Background

In this paper we report about eye-movement research and the investigation of dialogue structure as it has been going on at several institutions at Bielefeld University (cf. the reports listed in

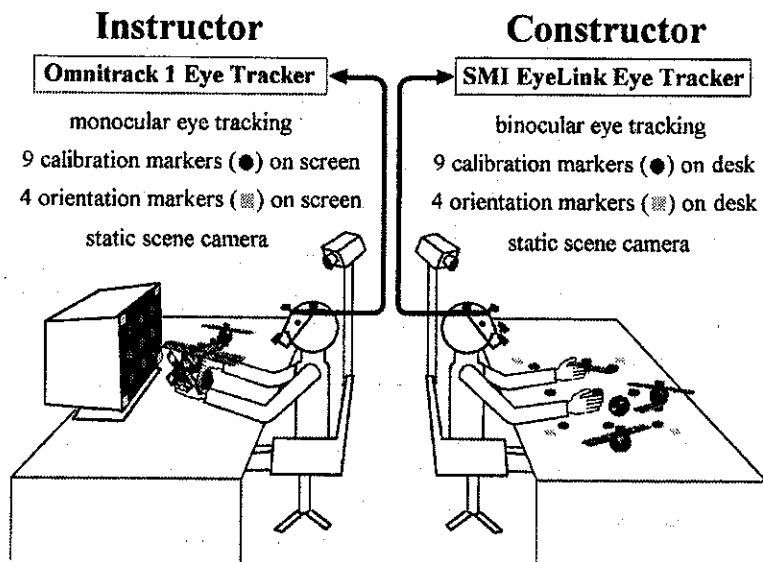


Figure 2: Scheme of the 3D-scenario

the bibliography). Dialogues, one might argue, can be regarded as sequences of turns in which perhaps certain micro- and macro-structures can be distinguished (cf. Asher 1993). But how can this be related to the investigation of agents' eye-movements? Empirical data, transcripts and video-films of two-person task-oriented dialogues reveal that agents do not behave in the way abstract semantics and pragmatics would suggest. As a result of the Russell-Donellan-Kripke discussion about speaker's meaning and abstract meaning, most researchers do know that, of course. Nevertheless, it is not trivial to find out what speakers actually do, and to start developing theories thereof. There has been little research in this area since the late 1970s. However, as we will see, investigation of speakers meaning is difficult, because one has to develop suitable methods for the observation of their doings. Dialogues and video-films provide only a rough idea of what is going on. But let us start from there, nevertheless. The following observations should perhaps be uncontroversial with respect to these and similar data:

- (1) Speakers select domains of interpretation and use them rather flexibly. This is important if we want to understand the use of definite descriptions, anaphora, and all sorts of relational expressions.
- (2) Speakers frequently describe things and situations from an agent-related perspective.
- (3) The use of descriptive vocabulary, especially non-literal expressions (tropes) and neologisms, is induced by the domain under discussion. In short: Specific domains instigate specific wordings.
- (4) The sequence of turns produced in describing the set up of an object depends on the ontological structure an agent "casts" over this object.
- (5) Agents coordinate their wording in order to complete tasks more efficiently.

We refer to these observations respectively as the "flexible domain constraint" (1), the "perspective constraint" (2), the "domain-description constraint" (3), the "ontology constraint" (4), and the "coordination constraint" (5).

If these constraints have some initial plausibility, it seems to be worthwhile finding out more about them. How can we do that? How can we get more reliable information about the mechanisms

of flexible domain selection? As explained above (cf. chapters 1 and 2) a promising method is the analysis of agents' visual foci taken as foci of attention during speech production, for example, while describing an object. The area spotted by attention can be taken as some sort of domain singled out according to relevance criteria. It leaves its traces in the language tokens produced. However, being in the mind, an agent's attention is not directly observable. Hence we must look for its nearest equivalent, and that is where his eyes rest upon. Perhaps one cannot maintain the latter in general, but apparently it can be applied to tasks involving the description of objects *seen*. In short, we identify the focus of attention with sequences of clustered foveal fixations.

In our report we describe three experimental studies involving eye-tracking and the findings they led to: a 2D-blocks world study, a 2D-'airplane' study, and, finally, a 3D-'airplane' study. The currently used 3D-setting seems to be the most promising one for future research. Our first 2D-study was based on task-oriented dialogues, in which an instructor told a constructor to build up a blocks world as shown in figure 3

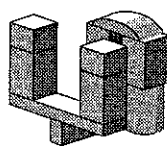


Figure 3: Blocks world

The 2D-Blocks World Setting: Planning, Search, and Interactive Focus-Management

Our empirical work was preceded by the development of the simulated speech exchange system SPEX (cf. Meier and Rieser 1995a,b), a recursively operating notion of focus which can simulate planning activities and verbal actions of agents in straight-forward cases. On the basis of SPEX, the following hypotheses about focus movement were formulated:

H01: Global direction of focus movement

In general, there can be two main directions of focus movement: from the front (the basis of the "see-saw") to the rear (the "archway") or vice versa.

H02: Embedded focus

Focus movement proceeds according to a principle of "focus-embedding": The initial focus covers the whole assembly; the subsequent focus is on the integral part selected according to the direction chosen; within this part, Inst focusses the basic block.

H03: Change of focus

The "support principle" determines how the focus moves upwards from the basic block.

H04: Focus and turn production

While Inst produces a turn, her focus remains fixed on the objects the turn is about.

H05: Focus change, current turn, and following turn

- a. The current focus determines the production of the current turn.
- b. Between turns, a change of focus occurs.
- c. The next focus determines the next turn.
- d. After a reply of Const, Inst may retain the focus position in order to match utterance and perceived situation.

H06: Tests

During tests, either an inverse focus movement contrary to the direction of "support" is used, or the agents go back to the basic block, thus initiating a new canonical focus movement.

H07: Privileged test points

Frequently, tests occur after the agents finished setting up an integral part of the whole assembly.

On the basis of the foveal fixations obtained from the first empirical study, most of these initial hypotheses had to be modified or neglected. While the first supposition, regarding the global direction of focus-movement, was confirmed by the data, parts of the second hypothesis had to be altered. It turned out that in contrast to the predictions of the second hypothesis Inst's focus neither encapsulates the whole figure, nor integral parts of it. Instead of this, it rather seems to act like a scanning device, whose task to provide Inst with an (initial) internal representation of the assembly depicted on the screen can be fulfilled independent of the sequence of foveal fixations. It is only within integral parts of the figure that focus movement is governed by a stricter rule: The "support"-principle, hypothesized to make Inst choose the basic block as the starting point for her description, and in 3. to determine the next block to be focussed within the selected part of the assembly, turned out to be suitable to explain the data obtained from the eye-tracker. However, as the discussion of the fourth and fifth hypothesis with its attention to interactive processes will more detailedly show, this principle had to be considered as a default option, which can be changed according to emerging needs.

Suppositions 4 and 5, which form together an important part, perhaps even the core of the hypotheses, contain predictions that are not uncontroversial. Whereas in connection with the computer simulation carried out by SPEX, the relatively rigid relationship between focus and turn sounds feasible, from a psycholinguistic point of view it is less convincing. The enormous speed of speech production seems to require the planning of each turn, which is, of course, also based on focus information, to be rather independent of its actual utterance. Due to these considerations, we were especially interested in the empirical results regarding the questionable hypotheses, and in fact our expectations to what might be found out were even exceeded. First of all, the fourth supposition, which in a way fixes the focus to the turn it determines, is clearly disproved by the data, for the comparison of language and foveal fixations frequently indicates planning processes. For example, while describing the yellow and the red cuboid in the blocks world, Inst already focusses the two cubes he is going to refer to next.

However, although the focus preceding the respective turn has to be considered the default case, there are also some exceptions to this rule, which occur in connection with search or lexical problems. When Inst is insecure about which word to choose for an object, her focus remains static and in this way matches the fourth hypothesis. If such tests of the semantics of her speech are carried out after the utterance has been finished, they may even lead to "focus loops", i.e. instead of moving ahead, the focus goes back to objects that have already been described.

These findings do not only require supposition 4 to be reformulated, but they also affect the assumptions (a) - (c) stated in connection with the fifth hypothesis, which have to be substantially changed. Prediction 5.(d), however, introduces an aspect that has not yet been considered, for it allows Const to influence Inst's focus movement. This interactive approach, which plays a minor role in the initial hypotheses, gains considerable weight with respect to the results of the empirical study. The data obtained from the eye-tracker strongly suggest that the rules claimed to determine Inst's focus be supplemented with assumptions about the role of the speech situation, i.e. the task-oriented dialogue or, more generally, the discourse. In fact, there are several possibilities of Const's actively contributing to the establishment of Inst's foci: First of all, the rules of discourse require Const to adequately react to the instructions she receives. As long as this back-channel behaviour, which is frequently to be found throughout the transcript, merely contains some ratification, Inst's strategy as it has already been described is not affected. If, however, the latter one is confronted with questions or Const's own proposals, her focus is determined by her partner's speech act.

This influence is insofar always a direct one, as Inst as the recipient of the utterance focusses the object Const refers to. Moreover, the effects of such distractions may also be more extended if

the contents of Const's turn can not so easily be dealt with. In this case, a longer subprocess with the chain of foci being mainly controlled by Const may be necessary before Inst finally returns to her description. Summing up, the considerations regarding the assumptions 4 and 5 show that in order to be empirically adequate, the notion of focus has to take into account the importance of (turn-preceding) planning in human language production, and the fact that focus-management in situated dialogue is highly interactive.

Compared to these findings, the discussion of the sixth and seventh hypothesis is less exciting: While the assumed two potential directions of foci are confirmed by the results of the study, the existence of privileged test-points is not corroborated. Both Inst and Const were rather reluctant to take the trouble to reconsider their achievements as long as no obvious problems occurred. In the face of the human ambition to minimize efforts, this is not surprising.

Altogether the results of our first study induced us to replace the initial hypotheses by the three assumptions stated below¹:

- H11: Asynchrony of planning and production: In the default case, an object is focussed before (not while) it is described.
- H12: Fixed focus in case of lexical problems: If formulating her turn poses a problem for Inst, her focus remains fixed on the object of current interest.
- H13: Coordination of focus: In coordinating their actions, Inst and Const mutually influence their foci of attention.

From Blocks Worlds to the Toy-Airplane Setting: Comparison of 2D-Blocks World Hypotheses with 2D-'Airplane'-Hypotheses

It is obvious that blocks worlds are well suited for the investigation of the instructor's focus of attention. This is due to the following facts: The blocks world chosen has a simple structure admitting essentially to ways of building up the figure: from front to rear or *vice versa*. Furthermore, it consists of elementary objects which do not have an internal structure, i.e. blocks of different colours and shapes. In addition, we have only one relevant ontological relation being at the heart of the block arrangements construction, namely the relation of support: The table surface supports various blocks, and these in turn support others. However, the empirical findings we got from this setting can only be generalized to classes of objects with similar properties, especially concerning the dominating support-relation.

In order to overcome this restriction, we carried out experiments with a toy-airplane that had structured parts, i.e. aggregates made out in turn of elementary parts, and which also exhibits a broader range of ontological relations, i.e. fixed contacts of various sorts among parts or aggregates. In this setting the instructor had at his disposal two 2D-representations of segments of the airplane (both showing the tail from somewhat different angles) and one global 2D-view of it (see fig. 4). We could only record the foveal fixations of the instructor during his description of the plane's tail. Due to the three parts of the airplane-stimulus, we expected the instructor to have a fairly precise idea about the functional relationships among the different parts of the plane: tail, fuselage, wings, and nose. The three pictures together served as a substitute for free object rotation in a 3D-setting. At the same time they provided characteristic perspectives of the plane.

Considering the tail in more detail, we observe the following structural differences to block arrangements: In the plane we have objects of different sorts: perforated cubes and bars, bolts with different heads, and nuts. These objects enter into different structural relations: Perforations (holes) can be used as ports for bolts. Nuts in turn fix bolts and therewith bars. In this way we get different types of contacts between aggregates. The contacts permit the aggregates (including the airplane itself) to be turned around.

We collected the instructor's focus data with respect to the tail and compared them with the simulation hypotheses and the empirical hypotheses of the blocks world study. Contrary to our

¹H 01, although confirmed by the data, is henceforth ignored because it is restricted to the specific blocks world setting.

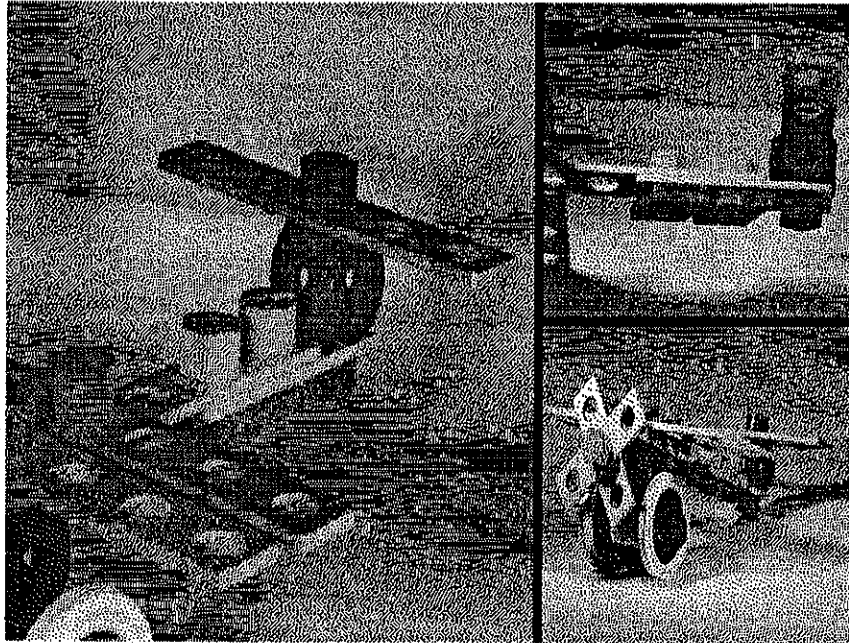


Figure 4: Toy-Airplane Setting

expectation we discovered that focus movements in the new setting were to some extent different from those in the blocks world. This can be attributed to the greater complexity of the tail as well as to the stimulus-representation consisting of the three parts which had to be integrated by the instructor. We label the new empirical hypotheses H2n etc. and compare them to the simulation hypotheses H0n etc. got via SPEX as well as to the empirical hypotheses H1n etc. coming from the blocks world.

- H21: The instructor uses all perspectives presented. Usually, he starts with the side-view of the tail, the fuselage and the wing. The focus-movement from one representation to the next fixes the objects to be mentioned in the instructor's next directive.
- H22: The initial focus acts like a pattern-interpretation process comprising the whole object. Subsequent foci isolate patterns for aggregates making up the tail. This is more in the line of our original simulation hypotheses which yielded the concept of 'embedded focus'.
- H23: Instead of a focus movement following some support-relation, we have different movements 'depicting' the functional relationships among parts.

Coordination of Eye-Movements among Instructor and Constructor

So far, we have only dealt with the instructor's focus of attention in task-oriented dialogue. From research situated in linguistics, logics and the philosophy of language we know that coordination and cooperation are a main characteristics of these dialogues. The studies reported so far also point in this direction: The constructor may at any time direct the eye-movement of the instructor's through check backs, clarification questions or other forms of comment. For example, this can be seen from the following dialogue, made up on the basis of empirical data²

²Observe that the following discourse is, as it stands, no natural datum but inspired by different passages from different dialogues.

- (1) Inst.: *Now, screw the 5-holes bar onto the red cube. Use a red bolt.*
- (2) Const.: *Round or hexagonal?*
- (3) Inst.: *Round.*
- (4) Const.: *The 5-holes bar, which side does it go to?*
- (5) Inst.: *Top, to the top.*

Initially, the instructor focuses on an aggregate consisting of a cube and a 5-holes bar fixed on it with a red round bolt. The constructor takes a 5-holes bar. Of these bars, only one sort is available. Looking for a bolt, he discovers that he has round-headed or hexagonal-headed red bolts, one of which he could use. This situation explains his question (2). The instructor shifts his attention to the bolt involved; hence his answer (3). Now the constructor discovers that he has various options of fixing the bar to the cube. This lies at the heart of his question (4). Most probably, the instructor will have already moved on further, due to the 'asynchrony of planning and production constraint'. He has to go back to the contact area made up by the elements under consideration: bolt, 5-holes bar and cube. Based on his focus, he can answer the constructor's query by issuing (5). Observe that the agents must have made clear beforehand which side of the aggregate has to count as top.

As far as we know, the example provided above is quite realistic. In order to check our intuitions concerning the agent's focus movements, we need a setting where the instructor and the constructor both are equipped with an eye-tracking device (see fig. 2).

Eye-Movements in Dialogue

So far we have investigated into the foveal fixations of instructors in relation to their speech production. Subsequently, we will report about our first 3D-eye-tracker study recording the eye-movements of both agents. We should like to get the following types of information from eye-movements:

- α) the set of fixations making up and in some sense delimiting the content going into a proposition
- β) the domains for the interpretation of smaller units of information making up propositions, such as domains for quantifiers, descriptions or relational expressions
- γ) hints concerning parameters which determine turn-taking in default-cases and non-default-cases.

The results of our first study summarized below are meant to serve as a starting point for further investigation in this direction. Let us turn to the instructor's foveal fixations first. So far, the following hypotheses concerning an instructor's focus of attention in a 2D-'airplane'-setting have been obtained:

H11: Asynchrony of planning and production

H12: Fixed focus in case of problems

H13: Coordination of foci

H22: Initial focus behaves like a pattern interpretation process

H23: Focus exploits functional relationships

In the 3D-setting some phenomena occur which are covered by H11. For example, consider the following passage in our construction dialogue:

(1) Inst. und eine rote eckige Schraube <pause.>, und die rote eckige Schraube
and a red hexagon bolt <pause.>, and the red hexagon bolt
 schiebst du von unten durch das letzte Loch der Dreilochleiste <pause.>
move you from below through the last hole of the 3holes-bar <pause.>

Const. Mhm.
Mhm.

Inst. und oben <pause.> schraubst du einen roten Lochwürfel drauf. <pause.>
and on top <pause.> screw you a red holes-cube on. <pause.>

Inst. Hast du?
Have you?

Translation:

Inst.: *and a red hexagon bolt. And you move the red hexagon bolt through the last hole of the 3holes-bar from the bottom side.*

Const.: *Mhm.*

Inst.: *And on top you screw a red holes-cube. Got it?*

Here the instructor's focus always precedes his speech production: While something is said about the red hexagonal bolt, the focus is already on the 3holes-bar. H12, related to fixed focus, can be demonstrated by the following sequence of turns:

(2) Inst.: und eine rote Schlitzschraube <pause..> und <par 1 beg> schiebst die <par 1 end>
and a red round bolt <pause..> and <par 1 beg> move the <par 1 end>

Const.: <par 1 beg> Eine rote <par 1 end> Schlitzschraube, warte, <par 2 beg>
 <par 1 beg> A red <par 1 end> round bolt, wait, <par 2 beg>
 jaha <par 2 end>
 jaha <par 2 end>

Inst.: <par 2 beg> ne rote <par 2 end> Schlitzschraube, eine
 <par 2 beg> a red <par 2 end> round bolt, a

Const.: Mhm.
Mhm.

Inst.: rote Schraube mit so'm Schlitz.
red bolt with such a slit.

Translation:

Inst.: *and a red round bolt and move the*

Const.: *A red round bolt, wait*
jaha,

Inst.: *A red round bolt, a*

Const.: *Mhm.*

Inst.: *red bolt with some kind of slit*

Here the constructor's turn acts as a check back. Consequently, the instructor's focus remains on the red round bolt.

A situation that shows coordination and fixed focus is the following one: Inst has hexagonal bolts with 'Baufix' written on them. Const has the same type of bolt without the inscription. Because of this difference, we get the following exchange:

(3) Inst.: Auf dieser gelben Sechskantschraube steht "Baufix", hast du so eine? <pause>
On this yellow hexagonal bolt is written 'Baufix'. have you such one? <pause>

Steht "Baufix" drauf.
Is written 'Baufix' on it.

Const.: Nein.
No.

Inst.: Also mu"st du irgendeine nehmen, ja? Also's Wichtigste ist, da"s es ne gelbe
Well must you any take, yes? Well most important is that it a yellow
<pause.> <par 7 beg> äh <par 7 end>
<pause.> <par 7 beg> uhm <par 7 end>

Const.: <par 7 beg> Ja <par 7 end>
<par 7 beg> Yes <par 7 end>

Inst.: Schlitzschraube ist.
round bolt is.

Const.: Mhm.
Mhm.

Translation:

Inst.: *There is 'Baufix' written on this yellow bolt. Have you got one?*
'Baufix' is written on it.

Const.: *No.*

Inst.: *Well, then you have to take just anyone. The most important thing*
is that it is a yellow
uhm

Const.: *Yes.*

Inst.: *round bolt.*

Const.: *Mhm.*

Inst asks for the yellow hexagonal bolt with the inscription 'Baufix' on it. Const declares that she does not have such a bolt. Therefore, Inst has to reconsider his description and so his focus remains on the object under discussion.

The data concerning H23 are quite surprising because foveal fixations seem to behave like invisible deictic gestures, pointing out properties or relations. This can be explained with respect to the following utterance of the instructor:

(4) Inst.: Und äh jetzt schiebst du die rote Schlitzschraube in <pause.> das dritte
And uhm now move you the red round bolt in <pause.> the third
Loch <pause.> und verschraubst es mit dem Lochwürfel <pause.>
hole <pause.> and screw together it with the holecube <pause.>

Translation:

Inst.: *And uhm now you slide the red round bolt into the 3rd*
hole and screw down the holecube with it.

Here Inst's foveal fixations first move through the hole in the bar. Afterwards, putting the aggregate together with the holes-cube is accompanied by specific eye-movements: The action of sliding the bolt into the hole is kind of imitated by the focus.

Coordination of focus is a ubiquitous phenomenon. As an example, consider (5):

- (5) Const.: <par 5 beg> Und die Leisten <par 5 end> im rechten Winkel oder wie?
 <par 5 beg> *And the bars* <par 5 end> *in a right angle or what?*
- Inst.: Ja. Und äh <pause.> ja, im rechten Winkel zu der Dreilochleiste.
Yes. And uhm <pause.> yes in the right angle to the 3holes-bar.
- <par 6 beg> Fünflochleiste. <par 6 end>
 <par 6 beg> *5holes-bar.* <par 6 end>
- Const.: <par 6 beg> Mhm. <par 6 end>
 <par 6 beg> Mhm. <par 6 end>
- Inst.: Mu"s im rechten Winkel zu der Dreilochleiste sein.
Must in the right angle to the 3holes-bar be.
- Const.: Mhm.
Mhm.
- Translation:
 Const.: *And the bars in a right angle, right?*
- Inst.: *Yes. And uhm yes in the right angle as regards the 3holes-bar.*
5holes-bar.
- Const.: *Mhm.*
- Inst.: *Has to make a right angle with the 3holes-bar.*
- Const.: *Mhm.*

Here Const's question determines Inst's focus movement. A 90° -angle is focussed in order to be certain about the orthogonal relation between the 3holes-bar and the 5holes-bar. Then Inst's focus moves ahead again in order to lay the ground for the next speech act.

We got no clear data supporting H22.

Concerning the constructor's side, we could establish the following hypotheses:

H'31: Constructor's free scanning

H'32: Anticipation of future actions to be carried out by focus movement

H'33: Focus directed by instructor's description of objects etc.

H'34: Fixed focus in case of problems encountered.

H'31 means that Const's focus scans through the objects on her side, independent of Inst's directives. For example this happens at the very beginning, before Inst starts his series of directives.

Anticipation read off from eye-movement can be discovered in the following turn of Inst's:

- (6) Inst.: Und äh jetzt schiebst du die rote Schlitzschraube in <pause.> das dritte
And uhm now move you the red round bolt in <pause.> the third
 Loch <pause..> und verschraubst es mit dem Lochwürfel <pause.>
hole <pause..> and screw together it with the holecube <pause.>

Translation:

Inst.: *And uhm now you slide the red round bolt into the 3rd*
hole and screw down the holecube with it.

While Inst produces 'schiebst du die'/move you the, Const already looks for a possible port for the bolt.

The normal procedure on Const's side is that she focuses an object/relation etc. using the information of Inst's utterance (H'33).

In case of problems, Const's focus also remains fixed. In such situations, there is neither free scanning nor anticipation of any sort. For example Const has difficulty in joining nuts and bolts to tie the tail to the fuselage. As long as this problem remains unsolved, her focus stays fixed. This parallels H12 regarding Inst.

With reference to our initial α , β , and γ , we see that we have no observations concerning α . However, we have data concerning β (see H 23) and γ . The results regarding γ can be summarized as follows: In the default case, Inst's focus moves ahead, and part of the information in the new focus is used to yield the content of his next speech act while the content of the previous speech act is realised and hence more or less obsolete for the planning process. If, however, Const's turn determines Inst's eye-movement, Inst's default procedure will stop and he will busy himself again with the content of his last speech act. This reveals a tight connection to mutual belief: An unperturbed default-move ahead seems to indicate that Inst believes it to be mutual belief that his last directive was carried out and his present directive can be carried out without problems. In the nondefault case there is no such belief in mutual belief.

References

General:

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.

Chierchia, G. (1995). *Dynamics of Meaning—Anaphora, Presupposition, and the Theory of Grammar*. Chicago UP.

Clark, H. H. (1996). *Using Language*. Cambridge UP.

Just, M. and Carpenter, P. (1987). *The Psychology of Reading and Language Comprehension*. Allyn & Bacon.

Selected Research Reports:

Clermont, T., Meier, C., M., P., Prestin, E., Rieser, H., Ritter, H., and Velichkovsky, B. (1995a). Augenbewegung, Fokus und Referenz. Technical Report 95/8, SFB 360.

Clermont, T., Meier, C., Pomplun, M., Prestin, E., and Rieser, H. (1995b). Focus and Reference. Videofilm on Eye-Movements and Focussing. Videofilm. SFB 360 "Situierete Künstliche Kommunikatoren", Bielefeld Univ.

Essig, K. (1998). Messung von binokularen Augenbewegungen in realen und virtuellen 3D-Szenarien. Diplomarbeit, Technische Fakultät der Universität Bielefeld.

Heydrich, W. and Rieser, H. (1994). Public Information and Mutual Error. In Kunze, J. and Stoyan, H., editors, *KI-94 Workshops*, pages 110–2. Gesellschaft für Informatik: Saarbrücken. workshop "Modellierung epistemischer Propositionen".

Heydrich, W. and Rieser, H. (1995). Public Information and Mutual Error. Technical Report 95/11, Universität Bielefeld, SFB360.

Meier, C. and Rieser, H. (1995a). Modelling Situated Agents' "Reference Shifts" in Task-Oriented Dialogue. In Dreschler-Fischer, L. and Pribbenow, S., editors, *KI-95 Activities: Workshops, Posters, Demos*, pages 318–21. Gesellschaft für Informatik: Bonn.

Meier, C. and Rieser, H. (1995b). Modelling Situated Agents' "Reference Shifts" in Task-Oriented Dialogue. Technical Report Techn. Rep. 95/11, SFB 360 "Situierete Künstliche Kommunikatoren", Bielefeld Univ.

- Meier, C. and Rieser, H. (1996). Perception, Focus and Resolution of Metonymy. In Gibbon, D., editor, *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference*, pages 305–9.
- Meyer-Fujara, J. and Rieser, H. (1997). Zur Semantik von Repräsentationsrelationen. Fallstudie Eins zum SFB-Flugzeug. Technical Report Techn. Rep. 97/7, SFB 360 "Situierte Künstliche Kommunikatoren", Bielefeld Univ.
- Pomplun, M., Rieser, H., Ritter, H., and Velichkovsky, B. (1997). Augenbewegungen als kognitionswissenschaftlicher Forschungsgegenstand. In Kluwe, R., editor, *Strukturen und Prozesse intelligenter Systeme*. DUV.
- Pomplun, M., Ritter, H., and Velichkovsky, B. (1996). Disambiguating Complex Visual Information: Towards Communication of Personal Views of a Scene. *Perception*, 25:931–948.
- Pomplun, M., Velichkovsky, B., and Ritter, H. (1994). An Artificial Neural Network for High Precision Eye Movement Tracking. In Nebel, B. and Dreschler-Fischer, L., editors, *Lecture notes in artificial intelligence: Proceedings KI-94*, pages 63–9. Springer.
- Rieser, H. (1997). Repräsentations-Metonymie, Perspektive und Koordination in aufgabenorientierten Dialogen. In Umbach, C., Grabski, M., and Hörnig, R., editors, *Perspektive in Sprache und Raum*, pages 1–26. DUV.
- Stampe, D. (1993). Heuristic Filtering and Reliable Calibration Methods for Video-Based Pupil-Tracking Systems. *Behavioral Research Methods, Instruments, and Computers*, 25:137–42.
- Velichkovsky, B., Pomplun, M., and Rieser, H. (1995). Attention and Communication: Eye-Movement-Based Research Paradigms. In Zangemeister, W. H., Stiehl, H. S., and Freksa, C., editors, *Visual Attention and Cognition*. Elsevier.

Mixing Situation Theory and Type Theory to Formalize Information States in Dialogue Exchanges *

Robin H. Cooper
Göteborg University
Göteborg, Sweden
cooper@ling.gu.se

Abstract

When information is exchanged in dialogue it is relevant to keep track not only of the information content about the domain of discussion but also information about linguistic events which are related to this content. Trying to formalize the situation theoretic approach to this has often involved us in rich and dubious ontologies. Here we blend ideas from situation theory together with work on dependent record types which has been conducted within Martin-Löf's type theory and suggest that this gives us a tractable way of characterizing information states that integrate information about semantic content and the linguistic utterances associated with it.

1 Introduction

When information is exchanged in dialogue it is relevant to keep track not only of the information content about the domain of discussion but also information about linguistic events which are related to this content. This idea was part of the earliest work on situation semantics by Barwise and Perry where meaning was treated in terms of a relation between various information resources such as utterance events, background resource situations and visual scenes. The idea has been taken up more recently both in the theoretical literature (e.g. by Ginzburg) and computational literature where quasi-logical forms are proposed which include syntactic information associated with terms.

Trying to formalize the situation theoretic approach to this has often involved us in rich and dubious ontologies, in particular situation theoretic notions of parameters, restricted objects and appropriateness conditions which have often been difficult to understand and which have made the theory difficult to use as a vehicle for expressing analytical insights, let alone pointing the way to a working application.

Recently Cooper (1998, forthcoming a, forthcoming b) has sketched a revision of the view of information states presented in Cooper (1996) and Cooper and Ginzburg (1996) which involves using ideas from Martin-Löf's type theory, in particular the notion of dependent record types introduced by Betarte and Tasistro (forthcoming), Tasistro (1997), Betarte (1998). The results are similar in spirit to other work using type theory for natural language semantics such as Ranta (1994), Ahn (1995) and Ahn and Borghuis (1998). In this paper, I will suggest that this blend of situation theory and type theory gives us a tractable way of characterizing information states that integrate information about semantic content and the linguistic utterances associated with it. First I will give an account of records and dependent record types. Then I will show how

*I am grateful to Bengt Nordström, Gustavo Betarte and other members of the type theory group at Gothenburg for discussion of dependent record types, although they are in no way responsible for what I have done (or failed to do) with them. Work on this paper has been supported by S-DIME (Swedish dialogue move engine), NUTEK/HSFR Language Technology project F305/97 and Maisie (Multimodal access to public information services) (KFB project 1997-0856-059) and TRINDI (Task-oriented instructional dialogue), EC Language Engineering project LE4-8314.

basic information states can be characterized using these notions. Finally, I will show how we can characterize the kind of information states we need in dialogue exchange, that is, information states which relate semantic content to speech events.

2 Records and Record Types

The presentation in this section takes as its starting point the work of Betarte and Tasistro (forthcoming), Tasistro (1997) and Betarte (1998). The reader should consult these works for an exact formulation within Martin-Löf type theory.

The basic idea is represented informally as follows:

If $a_1 : T_1, a_2 : T_2(a_1), \dots, a_n : T_n(a_1, a_2, \dots, a_{n-1})$,

the record:

$$\left[\begin{array}{l} l_1 = a_1 \\ l_2 = a_2 \\ \dots \\ l_n = a_n \end{array} \right]$$

is of type:

$$\left[\begin{array}{l} l_1 : T_1 \\ l_2 : T_2(l_1) \\ \dots \\ l_n : T_n(l_1, l_2, \dots, l_{n-1}) \end{array} \right]$$

Records are sequences of fields which are pairs of labels and values. Record types are sequences of fields which are pairs of labels and types. For a record to be of a given record type it must be the case that for each field in the type there is a field in the record with the same label and the object in the record field must be of the type in the type field. The types in the record type can depend on types occurring earlier in the record type (more technically, they are families of types depending on the record type consisting of the fields preceding them.)

We assume that we have a stock of labels ℓ to which we attach natural numbers as subscripts. (We will normally suppress the subscripted index when it is 0.) A field is a pair whose first member is an indexed label ℓ_i and whose second member is an object (in the case of a record object) or a type (in the case of a record type). We define the extension operation, \oplus as an operation that adds a field onto the end of a sequence of fields, i.e. $\langle f_1, \dots, f_n \rangle \oplus \langle \ell_i, a \rangle = \langle f_1, \dots, f_n, \langle \ell_j, a \rangle \rangle$, where $j = k + 1$ and k is the highest index occurring on ℓ in $\langle f_1, \dots, f_n \rangle$. If r and $\langle f_1, \dots, f_n \rangle$ are records, we shall use $r \oplus \langle f_1, \dots, f_n \rangle$ to represent $(\dots (r \oplus f_1) \oplus \dots \oplus f_n)$.

We shall represent a field $\langle \ell, a \rangle$ in a record object as $[\ell = a]$ and we shall represent a field $\langle \ell, T \rangle$ in a record type as $[\ell : T]$. (We shall leave out the brackets in these representations of fields when no confusion results.) Both record objects and record types will be displayed vertically as in (1).

$$(1) \quad \begin{array}{l} \text{a.} \\ \text{b.} \end{array} \left[\begin{array}{l} l_1 = a_1 \\ l_2 = a_2 \\ \dots \\ l_n = a_n \end{array} \right] \quad \left[\begin{array}{l} l_1 : T_1 \\ l_2 : T_2 \\ \dots \\ l_n : T_n \end{array} \right]$$

We can construct the set *RType* of record types by the following recursive definition assuming that we have some set *Type* of non-record types.

1. $\langle \rangle$ (the empty sequence) is a member of *RType*.

2. If ρ is a record type and T is a function from records of type ρ to $Type \cup RType$, l is a subscripted label, then $\rho \oplus l : T$ is a member of $RType$.

We shall use the notation $\lambda r : T(a(r.l_1, \dots, r.l_n))$ to represent a function from records r of type T to an object a dependent on fields labelled l_1, \dots, l_n ¹ in r . However, this "official" notation leads to a very cumbersome representation as in (2a) which we shall, following Betarte and Tasistro's convention, abbreviate as (2b).

$$(2) \quad \begin{array}{l} \text{a.} \\ \text{b.} \end{array} \left[\begin{array}{l} l_1 : T_1 \\ l_2 : \lambda r : [l_1 : T_1] (T_2(r.l_1)) \\ l_3 : \lambda r : [l_1 : T_1 \\ \quad l_2 : \lambda r : [l_1 : T_1] (T_2(r.l_1))] (T_3(r.l_1, r.l_2)) \\ \dots \\ l_1 : T_1 \\ l_2 : T_2(l_1) \\ l_3 : T_3(l_1, l_2) \\ \dots \end{array} \right]$$

We define what it means for a record r to be of a record type ρ , $r : \rho$, recursively as follows:

1. $\langle \rangle : \langle \rangle$, i.e. the empty record is of the empty record type.
2. If $r : \rho$, $a : T(r)$ and l a label not occurring in r , then $r \oplus [l = a] : \rho \oplus [l : T]$.
3. If $r : \rho \oplus [l : T]$, then $r : \rho$.²

(3a) is an example of a record which has a record as a value and (3b) is a representation of a type for that record.

$$(3) \quad \begin{array}{l} \text{a.} \\ \text{b.} \end{array} \left[\begin{array}{l} \text{rel} = \text{hire} \\ \text{args} = \left[\begin{array}{l} \text{hirer} = a \\ \text{hired} = b \\ \text{ev-time} = t \end{array} \right] \\ \text{pol} = 1 \\ \text{rel} : \text{Rel} \\ \text{args} : \left[\begin{array}{l} \text{hirer} : \text{Ind} \\ \text{hired} : \text{Ind} \\ \text{ev-time} : \text{Time} \end{array} \right] \\ \text{pol} : \text{Bool} \end{array} \right]$$

(Here we assume basic types $Rel(ation)$ ³, $Ind(ividual)$, $Time$ and $Bool(ean)$, that is the type to which 1 and 0 belong.)

An important aspect of this treatment of record types is that subtyping is allowed, i.e. a record with additional fields not mentioned in the type will still be of that type. Thus, for example, (4) is also a record of type (3b).

$$(4) \quad \left[\begin{array}{l} \text{rel} = \text{hire} \\ \text{args} = \left[\begin{array}{l} \text{hirer} = a \\ \text{hired} = b \\ \text{salary} = 100Kecu \\ \text{ev-time} = t \end{array} \right] \\ \text{pol} = 1 \end{array} \right]$$

¹More generally, l_i should represent a path (i.e. a sequence of labels) in r since we allow records to occur as values in record fields.

²Betarte and Tasistro allow subtyping of this nature but define it by defining subsumption on the types and then adding a rule of inference which requires that any object of a given type belongs to a type higher on the subsumption hierarchy.

³We have a terminological problem with "relation" since we are not thinking of *hire* as a relation in the metatheory. When it is important to make the distinction, we will refer to objects of type *Rel* as object relations and to relations in the metatheory as meta-relations.

3 Building in Notions from Situation Theory

The record (3a) corresponds to what is called a basic infon in situation theory. To save space and improve readability I will use the traditional situation theoretic notation for infons to abbreviate records corresponding to basic infons, as in (5). In this notation we are suppressing labels and the polarity field when its value is '1'.

$$(5) \quad \text{a.} \quad \left[\begin{array}{l} \text{rel} = \text{hire} \\ \text{args} = \left[\begin{array}{l} \text{hirer} = a \\ \text{hired} = b \\ \text{ev-time} = t \end{array} \right] \\ \text{pol} = 1 \end{array} \right]$$

b. $\langle\langle \text{hire}, a, b, t \rangle\rangle$

We can use record types to characterize what the appropriate arguments are for object relations. We can define a function *appr* which assigns a record type to objects of type *Rel*.⁴ In (6) are some examples of how *appr* might be defined.

$$(6) \quad \text{a.} \quad \text{appr}(\text{hire}) = \left[\begin{array}{l} \text{hirer} : \text{Ind} \\ \text{hired} : \text{Ind} \\ \text{ev-time} : \text{Time} \end{array} \right]$$

b. $\text{appr}(\text{run}) = \left[\begin{array}{l} \text{runner} : \text{Ind} \\ \text{ev-time} : \text{Time} \end{array} \right]$

Given such a function we can define the type of basic infons as a dependent record type.

$$(7) \quad (\text{Basic}) \text{ Infon} \equiv \left[\begin{array}{l} \text{rel} : \text{Rel} \\ \text{args} : \text{appr}(\text{rel}) \\ \text{pol} : \text{Bool} \end{array} \right]$$

A basic intuition that we have about records is that they can be used to classify objects in the world. For example, if we have a database which contains employee records, we can think of a record indicating a social security number as uniquely identifying a particular employee. We might represent this by introducing a relation, *classify_{ssn,ind}* between records of the type in (8) and individuals.

$$(8) \quad \left[\text{soc-sec-no} : \text{Num} \right]$$

So, for example, we might have the proposition in (9).

$$(9) \quad \text{classify}_{\text{ssn,ind}}(\left[\text{soc-sec-no} = 123456789 \right], \text{Jones})$$

Intuitively, *classify_{ssn,ind}* would be a functional relation since social security numbers uniquely identify people (or at least they are supposed to). However, classification relations need not be functional. For example, employees might be classified by job and salary. So we might introduce a relation *classify_{js,ind}* between records of the type in (10) and individuals.

$$(10) \quad \left[\begin{array}{l} \text{job} : \text{Job} \\ \text{salary} : \text{Ecu} \end{array} \right]$$

We could imagine that both of the propositions in (11) are true.

$$(11) \quad \text{a.} \quad \text{classify}_{\text{js,ind}}(\left[\begin{array}{l} \text{job} = \text{manager} \\ \text{salary} = 100\text{Kecu} \end{array} \right], \text{Jones})$$

b. $\text{classify}_{\text{js,ind}}(\left[\begin{array}{l} \text{job} = \text{manager} \\ \text{salary} = 100\text{Kecu} \end{array} \right], \text{Smith})$

⁴One may wish to think of *appr* as a (meta-)relation rather than a function or to allow disjunctive types in order to allow a single object relation to have more than one appropriate argument assignment.

We will exploit this intuition concerning classification relations in talking about events that are classified by records of the type *Infon*. Following the terminology of situation theory we will refer to events as situations.⁵ Also, following traditional situation theoretic usage we will talk of a situation s supporting an infon σ , in symbols $s \models \sigma$, and will use this to mean that σ classifies s , which on the model of the classification relations we introduced above could be represented as $\text{classify}_{\text{infon}, \text{sit}}(\sigma, s)$, that is in terms of a classification relation between records of type *Infon* and situations. That is, assuming that $s : \text{Sit}$ and hire, a, b and t are as before (12) will represent a proposition.

$$(12) \quad s \models \langle \langle \text{hire}, a, b, t \rangle \rangle$$

The proof of this proposition can be a direct observation of the world, an observation that the event represented by s is in fact one in which a hires b at time t .⁶ We could regard such a proof in formal terms as the checking of a definition which associates situations with sets of (basic) infons. In computational terms we can regard such a proof as a database lookup.

We will assume some standard axioms of situation theory:

$$\text{Extensionality } (\forall \sigma s \models \sigma \leftrightarrow s' \models \sigma) \rightarrow s = s'$$

Consistency If $\sigma : \text{Infon}$, let $\bar{\sigma}$ represent σ with the opposite polarity.

$$s \models \sigma \rightarrow s \not\models \bar{\sigma}$$

$$\text{Directedness } s_1 \models \sigma \wedge s_2 \models \tau \rightarrow \exists s_3 [s_3 \models \sigma \wedge s_3 \models \tau]$$

We take over from type theory the conception of propositions as types of proofs. We then have record types of the form (13).

$$(13) \quad [\text{prf} : s \models \sigma]$$

A record of this type would be (14), given that $\text{prf}(s \models \sigma)$ is a proof of $s \models \sigma$.

$$(14) \quad [\text{prf} = \text{prf}(s \models \sigma)]$$

Following the standard strategy of type theory we will say that any type is *true* just in case there is something of that type (i.e. the type is "inhabited"). That is, types do double duty as existentially quantified propositions. For example, (15) is true just in case a man owns a donkey. (We use *Ind* for the type *Individual* and *Sit* for the type *Situation*.)

$$(15) \quad \left[\begin{array}{l} \text{ref}_0 : \text{Ind} \\ \text{sit}_0 : \text{Sit} \\ \text{prf}_0 : \text{sit}_0 \models \langle \langle \text{man}, \text{ref}_0 \rangle \rangle \\ \text{ref}_1 : \text{Ind} \\ \text{sit}_1 : \text{Sit} \\ \text{prf}_1 : \text{sit}_1 \models \langle \langle \text{donkey}, \text{ref}_1 \rangle \rangle \\ \text{sit}_2 : \text{Sit} \\ \text{prf}_2 : \text{sit}_2 \models \langle \langle \text{own}, \text{ref}_0, \text{ref}_1 \rangle \rangle \end{array} \right]$$

Let us abbreviate (15) as T_{mod} . A function from records of this type to records of the dependent type (16) would be a proof that every man who owns a donkey beats it.

⁵In standard linguistic terminology events refer to a subkind of situation where some change takes place, and are to be distinguished from states (e.g. *John knows Mary*) which correspond to the persistence of a certain state of affairs over a period of time with no change. The word "eventuality" is often used to refer to something like what we are calling situation here, although not normally in the situation theoretic literature.

⁶In some cases the proof may not be direct but may be based on a related observation of the world. For example, suppose that a walks during a time interval t and that t' is an interval wholly contained within t and that we have an observation of $s \models \langle \langle \text{walk}, a, t \rangle \rangle$. Then, if we assume the subinterval property associated with activity relations like *walk*, we would be justified in concluding that the proposition $s \models \langle \langle \text{walk}, a, t' \rangle \rangle$ is true. By separating the observations from the support relation, we can say that while the supports relation (represented by ' \models ') is infinite, any situation is uniquely identified by a finite set of observations.

$$(16) \left[\begin{array}{l} \text{sit}_3 : \text{Sit} \\ \text{prf}_3 : \text{sit}_3 \models \langle\langle \text{beat}, \text{ref}_0, \text{ref}_1 \rangle\rangle \end{array} \right]$$

Abbreviating (16) as T_b , we represent the type of such a function in (17).

$$(17) (T_{mod})T_b$$

This type is *true*, i.e. inhabited, under the same conditions as the classical DRT analysis of donkey anaphora and also the type theory version first reported in Sundholm (1986), modulo the addition of situations.

4 Meaning

We shall take the static meaning of a declarative sentence to be a function from records of a given type to propositions (i.e. record types). The records in the domain of these functions model contexts in something like the sense that the word has been used in the linguistic literature (cf Montague's notion of variable assignments as contexts and the similar notion of context in dynamic semantics). This notion of context as corresponding to a record object is distinct from the use of the word "context" in type theory where it has been used to denote something intuitively similar to record types. I will therefore use "context object" to refer to record objects that play the role of context and "context type" to refer to record types used as contexts. The idea that static meaning for a declarative sentence should be a function from contexts (context objects) to propositions is, of course, the classical Montague-Kaplan view of meaning.

Thus a static meaning for the sentence *Smith hired Jones* could be defined as in (18).

$$(18) \llbracket \text{Smith hired Jones} \rrbracket_{\text{stat}} =$$

$$\lambda r: \left[\begin{array}{l} \text{utt-time} : \text{Time} \\ \text{ev-time} : \text{Time} \\ \text{restr} : \text{ev-time} < \text{utt-time} \\ \text{ref}_0 : \text{Ind} \\ \text{sit}_0 : \text{Sit} \\ \text{prf}_0 : \text{sit}_0 \models \langle\langle \text{named}, \text{ref}_0, \text{"Smith"}, \text{utt-time} \rangle\rangle \\ \text{ref}_1 : \text{Ind} \\ \text{sit}_1 : \text{Sit} \\ \text{prf}_1 : \text{sit}_1 \models \langle\langle \text{named}, \text{ref}_1, \text{"Jones"}, \text{utt-time} \rangle\rangle \\ \text{sit}_2 : \text{Sit} \end{array} \right] \left(\left[\text{prf}_2 : \text{sit}_2 \models \langle\langle \text{hire}, r.\text{ref}_0, r.\text{ref}_1, r.\text{ev-time} \rangle\rangle \right] \right)$$

Here the labels are to be glossed as follows: *utt-time* – "utterance time", *ev-time* – "event time", *restr* – "restriction", *ref_i* – referent, *sit_i* – "situation". The label *sit₂* in (18) corresponds to what in classical situation semantics would be called the "described situation".

Contexts are not always as complete as we would like. We may not know exactly which situation is being described. We may not know who is being referred to by the uses of certain names. Israel and Perry (1990, 1991) point out that in such cases we obtain existential information from the utterance, e.g. "There is a person named Smith who hired a person named Jones". Israel and Perry call this incremental content. We will treat this by saying that a meaning function such as (18) can be coerced when we apply it to a context object with insufficient fields to a function in which fields have been "lowered" from the type of the domain to the resulting proposition. An example of such a coerced function derived from (18) is (19).

$$(19) \lambda r: \left[\begin{array}{l} \text{utt-time} : \text{Time} \\ \text{ev-time} : \text{Time} \\ \text{restr} : \text{ev-time} < \text{utt-time} \\ \text{ref}_0 : \text{Ind} \\ \text{sit}_0 : \text{Sit} \\ \text{prf}_0 : \text{sit}_0 \models \langle\langle \text{named}, \text{ref}_0, \text{"Smith"}, \text{utt-time} \rangle\rangle \\ \text{ref}_1 : \text{Ind} \end{array} \right]$$

$$\left(\begin{array}{l} \text{sit}_1 : \text{Sit} \\ \text{prf}_1 : \text{sit}_1 \models \langle\langle \text{named, ref}_1, \text{"Jones"}, \text{utt-time} \rangle\rangle \\ \text{sit}_2 : \text{Sit} \\ \text{prf}_2 : \text{sit}_2 \models \langle\langle \text{hire, } r.\text{ref}_0, r.\text{ref}_1, r.\text{ev-time} \rangle\rangle \end{array} \right)$$

This corresponds to a case where the context provides an utterance time, an event time preceding the utterance time, an individual named "Smith" and another individual but no naming information about that individual. In this example, any or all of the fields can be lowered in this way, subject only to the general constraint that we cannot lower a field without also lowering any fields that depend on it.

We can define lowering of a particular field as in (20).

$$(20) \text{ If } f = \lambda r: \left(\begin{array}{l} l_i : T_i \\ \dots \\ l_{i-1} : T_{i-1}(l_1, \dots, l_{i-2}) \\ l_i : T_i(l_1, \dots, l_{i-1}) \\ l_{i+1} : T_{i+1}(l_1, \dots, l_{i-1}) \\ \dots \\ l_n : T_n(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_{n-1}) \end{array} \right) \left(\begin{array}{l} l_{n+1} : T_{n+1}(r.l_1, \dots, r.l_i, \dots, r.l_n) \\ \dots \\ l_{n+m} : T_{n+m}(r.l_1, \dots, r.l_i, \dots, r.l_n, l_{n+1}, \dots, l_{n+(m-1)}) \end{array} \right)$$

then $\text{lower}(l_i, f) =$

$$\lambda r: \left(\begin{array}{l} l_1 : T_1 \\ \dots \\ l_{i-1} : T_{i-1}(l_1, \dots, l_{i-2}) \\ l_{i+1} : T_{i+1}(l_1, \dots, l_{i-1}) \\ \dots \\ l_n : T_n(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_{n-1}) \end{array} \right) \left(\begin{array}{l} l_i : T_i(r.l_1, \dots, r.l_{i-1}) \\ l_{n+1} : T_{n+1}(r.l_1, \dots, r.l_{i-1}, r.l_{i+1}, \dots, r.l_n, l_i) \\ \dots \\ l_{n+m} : T_{n+m}(r.l_1, \dots, r.l_{i-1}, r.l_{i+1}, \dots, r.l_n, l_i, l_{n+1}, \dots, l_{n+(m-1)}) \end{array} \right)$$

However, I believe that natural language places constraints on such coercions in the case of pronouns. Consider (21).

(21) [He liked her]_{stat} =

$$\lambda r: \left(\begin{array}{l} \text{utt-time} : \text{Time} \\ \text{ev-time} : \text{Time} \\ \text{restr} : \text{ev-time} < \text{utt-time} \\ \text{pronref}_0 : \text{Ind} \\ \text{sit}_0 : \text{Sit} \\ \text{prf}_0 : \text{sit}_0 \models \langle\langle \text{male, ref}_0, \text{utt-time} \rangle\rangle \\ \text{pronref}_1 : \text{Ind} \\ \text{sit}_1 : \text{Sit} \\ \text{prf}_1 : \text{sit}_1 \models \langle\langle \text{female, ref}_1, \text{utt-time} \rangle\rangle \\ \text{sit}_2 : \text{Sit} \end{array} \right) \left(\begin{array}{l} \text{prf}_2 : \text{sit}_2 \models \langle\langle \text{like, } r.\text{ref}_0, r.\text{ref}_1, r.\text{ev-time} \rangle\rangle \end{array} \right)$$

I believe that all fields in this example except those marked with a label *pronref*, can be lowered by coercion. That is, pronouns must be associated with a referent in the context, though not necessarily one introduced by a linguistic antecedent. A pointing or some other reason for salience in the context may be sufficient to establish a pronoun referent. Note, however, that fields representing gender information associated with the pronoun may be lowered. Currently, I see the use of distinguished labels such as *pronref* as the mechanism for preventing lowering. We shall exploit this idea of the coerced meaning in the presence of incomplete contexts in our account of dynamic meaning.

5 Representing Information States

We can use the machinery that we have built up in this paper to represent an information state as a pairing of a *mental state*, which we will take to be the same kind of function as a meaning, and an *anchor*, a context object. This yields a new (and, I believe, improved) formulation of the notion of information state which has been associated with previous work on attitude reports (Barwise and Perry, 1983, 1985, Cooper, 1996, Cooper and Ginzburg, 1996 and several related proposals in the literature such as Kamp, 1990, Asher, 1993, Crimmins, 1992 and Richard, 1990). We will model information states as records of the form (22).

$$(22) \left[\begin{array}{l} \text{mental-state} = f \\ \text{anchor} = r \end{array} \right]$$

In (23) we give an example of such an information state.

$$(23) \left[\begin{array}{l} \text{mental-state} = \lambda r: \left[\begin{array}{l} \text{utt-time} : \text{Time} \\ \text{ev-time} : \text{Time} \\ \text{restr} : \text{ev-time} < \text{utt-time} \\ \text{ref}_0 : \text{Ind} \\ \text{sit}_0 : \text{Sit} \\ \text{prf}_0 : \text{sit}_0 \models \langle\langle \text{named}, \text{ref}_0, \text{"Smith"}, \text{utt-time} \rangle\rangle \\ \text{ref}_1 : \text{Ind} \\ \text{sit}_1 : \text{Sit} \\ \text{prf}_1 : \text{sit}_1 \models \langle\langle \text{named}, \text{ref}_1, \text{"Jones"}, \text{utt-time} \rangle\rangle \\ \text{sit}_2 : \text{Sit} \end{array} \right] \\ \left(\left[\text{prf}_2 : \text{sit}_2 \models \langle\langle \text{hire}, r.\text{ref}_0, r.\text{ref}_1, r.\text{ev-time} \rangle\rangle \right] \right) \\ \text{anchor} = \left[\begin{array}{l} \text{utt-time} = 28/2/98:14.01 \\ \text{ref}_0 = \text{ind56} \\ \text{sit}_0 = \text{sit5001} \\ \text{prf}_0 = \text{prf}(\text{sit5001} \models \langle\langle \text{named}, \text{ind56}, \text{"Smith"}, 28/2/98:14.01 \rangle\rangle) \\ \text{ref}_1 = \text{ind57} \\ \text{sit}_1 = \text{sit5002} \\ \text{prf}_1 = \text{prf}(\text{sit5002} \models \langle\langle \text{named}, \text{ind57}, \text{"Jones"}, 28/2/98:14.01 \rangle\rangle) \end{array} \right] \end{array} \right]$$

The content of (23) is obtained by coercing the mental state in the manner described above to a function for which the anchor is appropriate and then applying the coerced function to the anchor. The result will be a proposition (a record type) that in effect existentially quantifies over objects which are not supplied to the mental state function by the anchor.

6 Information States in Dialogue Exchange

Suppose that agent *A* wishes to communicate to another agent *B* a simple fact, *Smith hired Jones*, based on a situation that *A* perceived and that *A* decides to communicate this fact by uttering the sentence 'Smith hired Jones'. It seems to me that we need to be able to represent *A*'s information state before she makes the utterance as an information state without any linguistic information, more or less as we have represented this information state above. However, once the utterance has been made, *A* has an additional information state which relates the information to various parts of the utterance, e.g., the person Smith to the particular utterance of "Smith" and so on. It is this detailed relating of the utterance to the non-linguistic information that enables *A* to externalize aspects of her mental state and that enables *B* to recover information about her mental state. We thus need a way of representing mental states which include information about utterance events and their relationship to components in non-linguistic information states. By treating utterances (i.e. speech acts) as situations, that is, the same kind of objects as figure in non-linguistic information states, we can freely mix information about the utterance and what it

describes (refers to). This idea of relating described and utterance situations has its origins in the early situation semantics work of Barwise and Perry (1983). In (24) we show a mental state corresponding to *Smith hired Jones* with added information about the utterance.⁷

(24) λr :	utt-time : Time ev-time : Time restr : ev-time < utt-time u ₀ : Sit syntax ₀ : np(u ₀ , "Smith") ref ₀ : Ind sem ₀ : refer(u ₀ , ref ₀) sit ₀ : Sit res ₀ : resource(u ₀ , sit ₀) prf ₀ : sit ₀ ⊨ ⟨⟨named, ref ₀ , "Smith", utt-time⟩⟩ u ₁ : Sit syntax ₁ : np(u ₁ , "Jones") ref ₁ : Ind sem ₁ : refer(u ₁ , ref ₁) sit ₁ : Sit res ₁ : resource(u ₁ , sit ₁) prf ₁ : sit ₁ ⊨ ⟨⟨named, ref ₁ , "Jones", utt-time⟩⟩ u ₂ : Sit syntax ₂ : tv(u ₂ , "hire") sem ₂ : refer(u ₂ , hire) u ₃ : Sit syntax ₃ : sentence(u ₃ , u ₀ , u ₂ , u ₁) sem ₃ : refer(u ₃ , ⟨⟨hire, ref ₀ , ref ₁ , ev-time⟩⟩) sit ₃ : Sit desc-sit : described-situation(u ₃ , sit ₃) ([prf ₃ : sit ₃ ⊨ ⟨⟨hire, r.ref ₀ , r.ref ₁ , r.ev-time⟩⟩])
--------------------	--

We have introduced the following fields representing linguistic information:

- a field labelled u_i for each constituent utterance of the sentence. Each u_i is typed as a situation.
- a syntax field for each subutterance. The typing requires that in a record of the type required the value of the syntax-field be a proof of a proposition about the constituent structure. Notice that the arguments to syntactic category relations such as *sentence*, *np* are utterance events, not strings or constituent structure tree or feature structures. This, for example, makes it natural to reason about the timing of these events in proving such propositions.
- a semantics-field (sem_i) for each subutterance. Here what is required in the record is a proof about reference.
- a field for the situation described by the sentence

Note that much more information could be added, e.g. phonology, focus, syntactic features etc. This is meant to be an illustrative example of manageable size.

The context in which an agent finds herself may not provide values for all of these linguistic fields. In particular the anchor if the information state of the hearer may be defective. This is probably normally the case, for example, in respect of the described situation field. However, in the presence of a defective anchor the mental state can be coerced by lowering in the way we have

⁷This is meant to illustrate the possibilities rather than provide a complete or adequate syntactic analysis. For the sake of simplification, I have omitted a VP constituent.

described above. In this way, we can obtain predictions not only about the content of an utterance but also about the utterance itself on the basis of incomplete information.

Mental states which relate utterances to information content in this way are important for modelling the way in which information is transferred during the course of a conversation. The agent *A* who wishes to transfer the information to *B* has first to take a non-linguistic information state and make it into a linguistic one by performing an utterance. *B*, by reasoning about the utterance and accepting it, enters into a similar linguistic information state. (There is, of course, lots of scope for error.) For later storage the information state can be stripped of its linguistic information and *B* obtains a non-linguistic information state which is similar to that which *A* began with. This stripping of linguistic information involves removing all fields labelled with *u* or depending on a *u*-label.

7 Conclusion

We have shown how to use dependently type record structures to represent meaning and information states. We have done this by blending together ideas from Martin L of type theory and situation theory. Finally, we have sketched how this view might be used in an account of information transfer in dialogue.

References

- Ahn, Ren  (1995) *Communicating Contexts: A Pragmatic Approach to Information Exchanges*, in Dybjer, Nordstr m and Smith (1995).
- Ahn, Ren  and Tijn Borghuis (1998) *Communication Modelling and Context-Dependent Interpretation: an Integrated Approach*, manuscript, Tilburg University and Eindhoven University of Technology.
- Anderson, C. Anthony and Joseph Owens, eds (1990) *Propositional Attitudes: the Role of Content in Logic, Language and Mind*, CSLI Publications.
- Asher, Nicholas (1993) *Reference to Abstract Objects in Discourse*, Kluwer Academic Publishers, Dordrecht.
- Barwise, Jon, Jean Mark Gawron, Gordon Plotkin and Syun Tutiya, eds, (1991) *Situation Theory and its Applications, Vol. 2*, CSLI Publications, Stanford.
- Barwise, Jon and John Perry (1983) *Situations and Attitudes*, MIT Press.
- Barwise, Jon and John Perry (1985) *Shifting Situations and Shaken Attitudes*, *Linguistics and Philosophy*, Vol. 8 pp. 399–452.
- Betarte, Gustavo (1998) *Dependent Record Types and Algebraic Structures in Type Theory*, Ph.D. thesis, Department of Computing Science, G teborg University and Chalmers University of Technology.
- Betarte, Gustavo and Alvaro Tasistro (forthcoming) *Extension of Martin-L of's type theory with record types and subtyping*, *Proceedings of the conference "25 Years of Constructive Type Theory"*, Oxford University Press.
- Cooper, Robin (1996) *The Attitudes in Discourse Representation Theory and Situation Semantics*, in Seligman and Westerst hl (1996), pp. 137–150
- Cooper, Robin (1998) *Integrating diverse information resources into dialogue updates*, *Proceedings of the Second International Conference on Cooperative Multimodal Communication*, ed. by H. Bunt, R.-J. Beun, T. Borghuis, L. Kievit and M. Verlinden, Tilburg University, pp. 70–82.

- Cooper, Robin (forthcoming a) Using Situations to Reason about the Interpretation of Speech Events, to appear in *Current Issues in Computational Semantics*, ed. by Harry Bunt and Reinhard Muskens, Kluwer Academic Publishers, Dordrecht.
- Cooper, Robin (forthcoming b) Information States, Attitudes and Dialogue, in *Proceedings of the Second Tbilisi Symposium on Language, Logic and Computation*, ed. by R. Cooper and T. Gamkrelidze.
- Cooper, Robin and Jonathan Ginzburg (1996) A Compositional Situation Semantics for Attitude Reports, in Seligman and Westerståhl (1996)
- Crimmins, Mark (1992) *Talk about Beliefs*, MIT Press, Cambridge, Mass.
- Dybjer, Peter, Bengt Nordström and Jan Smith, eds., (1995) *Types for Proofs and Programs: selected papers*, Springer Verlag, Berlin.
- Hanson, P.P., ed. (1990) *Information, Language and Cognition*, Vancouver Studies in Cognitive Science, Vol. 1, University of British Columbia Press, Vancouver.
- Israel, David and John Perry (1990) What is Information? in Hanson, ed. (1990), pp. 1–19.
- Israel, David and John Perry (1991) Information and Architecture, in Barwise, Gawron, Plotkin and Tutiya, eds (1991), pp. 147–159.
- Kamp, Hans (1990) Prolegomena to a Structural Theory of Belief and other Attitudes, in Anderson and Owens (1990).
- Ranta, Aarne (1994) *Type-Theoretical Grammar*, Oxford University Press, Oxford.
- Richard, Mark (1990) *Propositional Attitudes: An Essay on Thoughts and How we Ascribe them*, Cambridge University Press, Cambridge.
- Sundholm, Göran (1986) Proof Theory and Meaning, in Gabbay and Guenther (1986).
- Tasistro, Alvaro (1997) *Substitution, record types and subtyping in type theory, with applications to the theory of programming*, PhD thesis, Department of Computing Science, Göteborg University and Chalmers University of Technology.

The Interplay of Desire and Necessity in Dialogue

Jean-Louis Dessalles

Ecole Nationale Supérieure des Télécommunications

46 rue Barrault - 75013 Paris - France

dessalles@enst.fr

Abstract

The purpose of this paper is to suggest that many argumentative moves in casual dialogues can be explained in terms of conflicting desires and conflicting beliefs, in such a way that some of these moves may be predicted. Participants appraise the different outcomes of the conflicting situation and try to find, together, through dialogue, a solution that they consider as acceptable. We show how realistic dialogues can emerge through a simple recursive process from an initial cognitive conflict. This model is implemented in our program PARADISE which can reconstruct the argumentative moves of some real conversations.

1 Introduction

In many applications, such as Computer Assisted Learning, text generation, contextual help systems and explanation in Knowledge Base Systems, the quality of the computer output crucially depends on its ability to generate a content that will be considered as relevant by human users. It is thus important to understand the process through which we produce the argumentative content of our utterances in everyday dialogues. While this process as a whole still remains mysterious, we claim that some of its aspects, which have a direct influence on acceptability, can be successfully simulated.

We address here the problem of the content of argumentative moves. Therefore we pay no attention, in the first place, to the linguistic form of utterances. While results produced by Conversation Analysis and Discourse Analysis (Goodwin & Heritage 1990, Moeschler 1990, Hirst 1991) may give a reference frame, they do not provide by themselves a sufficient account for argumentative moves. Surface phenomena like the embedding of utterances are not restrictive enough to reveal the process through which arguments are generated (Sadek, 1996). On the other hand, the social context of the interaction (Baker, 1991) or the conversational style of participants (Tannen, 1984) may prove to be too complex to be useful in the first place. The characterisation of speech acts is also both difficult and insufficient for an accurate prediction of arguments (Liddicoat, 1995). Our approach addresses neither the surface level nor the social level, but rather the knowledge level. It is thus close to models based on plan recognition. There is a difference, however. These models take into account not only the knowledge shared by participants, but also second-order knowledge, as in this example from Airenti et al. (1993):

$$R21 : SH_{yx}CINT_{xy}p \& BEL_y \text{ sincere}(x, y, p) \supset BEL_y BEL_x p$$

"if it is shared by y and x that x communicatively intends that some fact p be shared by y and x , and if y believes x to be sincere with him about p , then y believes that x believes that p ."

Such descriptions may be necessary in situations like goal oriented dialogues in which participants are uncertain about each other's knowledge and goals. Our claim, however, is that we can avoid

the complexity introduced by epistemic statements and still account for the argumentative frame of many casual dialogues. This does not mean that no assumption is made about the participants' beliefs. On the contrary: we must have an accurate representation of what the participants know about the domain they are talking about. Simply, we make the simplifying assumption that they do not develop explicit hypotheses about what their interlocutors know or ignore. As we will see, this assumption will not appear as too restrictive in many cases.

Authors who develop computational models of argumentation first analyse the content of arguments uttered in natural contexts and study their logical effect on preceding arguments (Guez 1990, Flowers et al. 1982). The challenge is then to justify the role of each argument using a minimal knowledge, as in (Quilici, 1992). The originality of our approach comes from the strict separation between domain knowledge and dialogic principles, and from the small number of such principles.

In what follows, we introduce the notion of "cognitive conflict" by showing that many dialogues begin with an incompatibility between desires or beliefs. Then we illustrate the role of successive arguments by showing their effect on the current conflict. This leads us to the main claim of this paper: the logical structure of many dialogues results from a recursive process in which attempts to escape from a conflict alternate with the occurrence of new conflicts. This process has been implemented in our program PARADISE. Lastly, we discuss the implications and limits of this approach.

2 Cognitive Conflicts

Some dialogues are presented as resulting from the opposition between conflicting goals supported by different locutors (Flowers et al. 1982, Baker 1991). We want to extend this perspective by showing that conflicts most often do not originate between participants, but among participants' beliefs. We call them "cognitive conflicts", since they involve thoughts rather than individuals. Moreover, for our concern here which is to predict the content of argumentative moves, considering cognitive conflicts as such will prove to be more relevant than considering the social interaction itself. The following example, taken from Tannen (1984:62), will illustrate this broader notion of conflict.

- (1) Context: A, B and C were speaking about sociology, and B showed a fairly good knowledge of Erving Goffman's books. A and C are surprised, since they thought this author was known only among specialists.
- A1: But anyway. ... How do you happen to know his stuff?
B1: Cause I read it.
C1: What do you do?
A2: [? ?] are you in ... sociology or anything?
B2: Yeah I read a little bit of it.
A3: Hm ?
B3: I read a little bit of it.
A4: I mean were you... uh studying sociology?
B4: No.
A5: You just heard about it, huh ?
B5: Yeah. No. I heard about it from a friend who was a sociologist, and he said read this book, it's a good book and I read that book 'n
A6: I had never heard about him before I started studying linguistics.
B6: Really ?
A7: Yeah.

This conversation is interesting because B's strange behaviour reveals an important mechanism which is involved in the process of argument generation. Tannen explains B's reaction at the social level: B felt aggressed by sudden and intrusive questions like A1, A2 and C1, and did not answer properly until B5. What is particularly interesting for our concern is precisely the difference of

acceptability between B1 and B5. B1 is not acceptable, and is repeatedly discarded by A and C (replies C1, A2, A4 and A5). We claim that B5 is relevant because it solves a cognitive conflict, while B1 is unacceptable because it leaves the conflict unchanged. The cognitive conflict arises in A and C's mind because they do not understand how B could know Goffman's books without being himself a sociologist, as revealed by Tannen who is A in this excerpt. If we consider that A and C hold the following belief:

$$\textit{knows}(X, \textit{goffman's books}) \Rightarrow \textit{sociologist}(X)$$

then the belief that B is not a sociologist conflicts with his knowledge of Goffman's books. In other words, the fact that B is a sociologist appears to be both false and true: B is believed to have another profession, but at the same time his knowledge of Goffman presents him as an expert in sociology.

Now we understand why B1 is not admissible. It has no effect on the cognitive conflict. The fact that B read Goffman's books is indeed a cause of his knowing of the books, but it does not affect the conflict between this belief and the belief that he is not a sociologist. With B5, the situation changes. We have the intuitive feeling that the conflict is solved, and that B5 acted as a genuine explanation. To arrive at a more formal description, we should consider that B5 changes the knowledge hold by participants by introducing a new premise. You may know Goffman's books either by being a sociologist or by being recommended his books:

$$\begin{aligned} \textit{knows}(X, \textit{goffman's books}) \\ \Rightarrow (\textit{sociologist}(X) \textit{ or } \textit{recommends}(Y, X, \textit{goffman's books})) \end{aligned}$$

We see that the conflict is cancelled. B may now know Goffman's books without being a sociologist. No contradiction ensues. This effect of B5 is precisely what makes this reply relevant. The architecture of this excerpt is much better understood if, before considering the social situation, we describe it at the knowledge level. This reveals a cognitive conflict which is apparent from the very beginning, and which seems to constitute the proper motivation of later replies. We will claim that many dialogues consist of successive alternations between cognitive conflicts and their solution.

3 Reconstructing Conversations

The reconstruction of arguments that were really uttered by human interlocutors during a spontaneous interaction is the best way to check the accuracy of a model that describes and predicts the content of argumentative moves. We will show in detail how the following conversation can be reconstructed at the argumentative level. In order to do so, we will first analyse how arguments are generated at the knowledge level.

(2) Context: A is repainting doors in his home. He decided to remove the old paint first, which proves to be a hard work (translated from French)

- A1: I have to repaint my doors. I've burned off the old paint.
It worked OK, but not everywhere. It's really tough work! [...]
In the corners, all this, the mouldings, it's not feasible ! [...]
- B1: You have to use a wire brush
- A2: Yes, but that wrecks the wood
- B2: It wrecks the wood ... [pause 5 seconds]
- A3: It's crazy! It's more trouble than buying a new door.
- B3: Oh, that's why you'd do better just sanding and repainting them.
- A4: Yes, but if we are the fifteenth ones to think of that
- B4: Oh, yeah ...
- A5: There are already three layers of paint
- B5: If the old remaining paint sticks well, you can fill in the peeled spots with filler compound
- A6: Yeah, but the surface won't look great. It'll look like an old door.

We observe that this conversation starts with the expression of a cognitive conflict. A wants to repaint his doors. As a consequence, he must burn off the old paint, which means working hard (because of the presence of mouldings). However, A would prefer avoiding such effort. The conflict results from two incompatible desires : A wants to remove the old paint, and he wants to avoid tough work. The relevance of B1 appears then clearly : by using a wire brush, A can remove the old paint easily, and the conflict vanishes. A acknowledges this fact in A2, but also points to a new conflict : using a wire brush is desirable (to avoid tough work), but it is incompatible with the wish of having a smooth surface, since the wood gets wrecked. This conversation is a typical example of an alternation between conflicts and solutions.

The linguistic form of utterances is not relevant here. We retain only the argumentative frame, which can be sketched this way:

- (3) A1: repaint, burn-off, mouldings, tough work
B1: wire brush
A2: wood wrecked
A3: tough work
B3: sanding
A5: several layers
B5: filler compound
A6: not nice surface

The challenge is to predict each of these argumentative moves using a domain knowledge and general argumentative principles. The difficulty of the task should not be underestimated. The domain knowledge should not be designed for the purpose of the reconstruction. It should be validated in principle by an external expert. Also, the argumentative principles hold for any conversation and should not be changed to fit the actual utterances. Consider first the following domain knowledge. It is given in propositional logic and causal links (clauses C1-C8) do not mention time indexes. This is for the sake of simplicity, since what we want to illustrate is more the management of conflicting necessities than knowledge representation.

- (C1) *burn_off*[& *not wood_wrecked*] → *nice_surface*
- (C2) *filler_compound*[& *sanding*] → *nice_surface*
- (C3) *sanding*[& *not several_layers*] → *nice_surface*
- (C4) *burn_off*[& *mouldings* & *not wire_brush*] → *tough_work*
- (C5) *wire_brush*[& *burn_off*] → *not tough_work*
- (C6) *wire_brush*[& *burn_off* & *wood_soft*] → *wood_wrecked*
- (C7) *wood_wrecked* → *not nice_surface*
- (C8) *repaint*[& *nice_surface*] → *nice_doors*
- (C9) *actions*(*repaint, burn_off, wire_brush, sanding, filler_compound*).
- (C10) *atypical*(*[wood_soft, several_layers]*).
- (C11) *undesirable*(*tough_work, 10*)
- (C12) *undesirable*(*not nice_doors, 20*)

The premises in brackets in the causal links constitute the context in which these links apply. Some propositions are marked as actions (C9). This means that their truth value can be freely determined. Their default value is false. Other propositions are listed as atypical (C10). Their default value is also negative. Lastly, some propositions are marked as undesirable (C11 and C12). The associated value is indicative. Only their relative ranking is relevant and they may change during the argumentative process. The problem is to show how a simple process can make use of this knowledge to generate the arguments that were really given.

4 From Knowledge to Conversation

The model, implemented in the program PARADISE, relies on two basic mechanisms: abduction and necessity management. We will illustrate how these mechanisms are involved in the reconstruction of the preceding conversation.

Example of Argument Generation

Before arriving at B1, the program must decode the situation and the motivation of actions described in A1. When A1 is given as input, the program first instantiates *repaint* as true. It recognises it as an action, using C9. By making an abduction, it infers the underlying motivation, thanks to C8: the effect does not exist before the action, therefore we have *not nice_doors* initially. This situation is undesirable with intensity 20 (C12). *nice_doors* becomes desired with intensity 20, and a cognitive conflict occurs. It vanishes immediately since *repaint* produces the desired effect, *nice_doors*, from which it acquires the necessity value 20. At this point, the program has nothing to do. The cognitive conflict has been solved and no causal link can be activated. Notice that many of the inferences performed by the program have no effect on its output. They are however necessary for the correct computation of arguments.

A1 continues with *burn_off*, which is also an action. The program similarly uses C1 to infer *not nice_surface* through abduction, and the conflict *nice_doors* / *not nice_doors* reappears since C8 is now blocked. But it vanishes immediately thanks to C1 which restores *nice_surface*. This term, like *burn_off*, inherits a necessity value 20.

A1 then reveals *mouldings* and *tough_work*. The program comes upon another cognitive conflict: *tough_work* has a necessity value 20 that is inherited from *burn_off*. But *not tough_work* is desired with the intensity 10 (C11). These conflicting desires and necessities call for a solution. This solution is obtained through abduction, thanks to C5. *wire_brush* is an action. It can be made true, and produces the wanted effect, *not tough_work*. It receives the necessity 10, the conflict is solved and B1 can be uttered.

Basic Principles used in Argument Computation

The program is given two basic abilities: abduction and the ability to propagate necessity coefficients (a deductive ability is of course assumed, but it could be part of another module). Abduction consists in inferring causes from effect. The current implementation includes only a basic form of abduction, since the it is not a proper part of our model. An abduction from D using the causal link:

$$A[\& B \& C] \rightarrow D$$

is only possible if A , B and C are true or unknown. A is abducted first. The abducted term must of course be unknown. Another form of abduction consists in assuming that the effect of a new action was false before the action. For instance if A is an action and D is unknown, *not D* is assumed before A is performed.

The computation of necessity coefficients, which constitute the core of our model, can be described using a few principles:

- (P1) if a new fact T is introduced that has an undesirability N , *not T* is produced and receives the necessity N .
- (P2) an effect inherits the necessity of its cause.
- (P3) the negation of a cause inherits the necessity of the negation of its effect.
- (P4) An abducted term is given the necessity of its effect

These principles control the process of argument generation. When no conflict is detected, necessity coefficients are simply propagated, according to P1-P4. A conflict arises when (T, N_1) and $(\text{not } T, N_2)$ occur at the same time. N_1 and N_2 are the respective necessities of T and *not T*.

- (P5) In such a conflicting situation, the program attempts to "save" the weaker term (here *not T* if $N_1 > N_2$) by looking for a cause through abduction. This is what allowed the program to find *wire.brush* and to produce B1.
- (P6) When no such solution can be found, the weaker term of the conflict is negated. P3 applies and the conflict shifts to other terms.

For instance, when A2 (wood gets wrecked) is given, it reveals a conflict that the program did not anticipate. The causal link C6 had not been triggered because *wood.soft* was atypical, but now the program makes the abduction *wood.soft* through C6. A deduction through C7 leads to a conflict between $(\text{not nice.surface}, 10)$, produced by P2, and $(\text{nice.surface}, 20)$ which was previously memorised. No new abduction can produce *nice.surface*. As a consequence, *nice.surface*, which has a greater necessity, "wins". According to P6, *not wood.wrecked* is produced through C7 and P3, and the conflict shifts to $(\text{wood.wrecked}, 10)$ and $(\text{not wood.wrecked}, 20)$. For the same reason, *not wood.wrecked* is enforced, and the conflict is transferred to $(\text{wire.brush}, 10)$ and $(\text{not wire.brush}, 20)$. *wire.brush* loses the battle, and we are back to the situation we had before B1. No wonder that A3 echoes A1.

The remainder of the conversation is reconstructed according to the same principles. We must suppose that A's insistence on *tough.work* increases the undesirability of this term to, say, the value 30. *not burn.off* inherits the necessity 30 through C4 and P3, which is higher than *burn.off*'s necessity 20. The conflict is transferred by P6 to *nice.surface* which is set to false. There, according to P5, an abduction can be done using C3. It restores *nice.surface* and justifies B3. Similarly, after the blocking of C3 by *several.layers*, B5 is found through a further abduction using C2.

5 Technical Remarks

Our first remark concerns the knowledge representation, which is quite basic in the previous example. A realistic use of the system would of course require a more general representation involving variables, time indexes, quantification, explicit default conditions, etc. The knowledge module, however, is not a proper part of our system. It should rather be part of a semantic module. What is relevant to our model is the ability of this semantic module to detect unusual situations (e.g. a cause with an impossible effect) and to perform abductions. For our purpose here, which was to explain how the management of necessities allows the generation of relevant arguments, the use of an elementary knowledge representation was sufficient.

The simplicity of the dialogic principles, on the other hand, is a positive characteristic of our model. Many other approaches give the priority to an explicit representation of intentions, goals and plans pursued by interlocutors. Notions like interfering plans (Quilici, 1992), argumentative orientation or argumentative strength (Guez, 1990); preconditions, sub-plans and sub-goals, support or threat for a plan (Young, Moore & Pollack, 1994) are not made explicit in our model. Some of them emerge from the necessity management and from the abductive process. Our model is, in itself, a kind of planning program, since its output can be seen as a plan. However, it does not manipulate programs as such. For instance, when uttering B5, the program has obtained a whole plan for an apparently satisfactory repainting of the doors. Such plans result from a poorly efficient process if we compare it with classical planning programs (Kambhampati, 1997). However, as we will claim, our model is much more plausible from a cognitive perspective.

Another kind of representation which is absent in our model is second-order knowledge: knowledge about others' knowledge and about plans. In our approach, this meta-knowledge often proves to be inessential. Whenever it is necessary, our project is to represent and to use it exactly as domain knowledge.

A further remark concerns the non-monotony of the system. This non-monotony is essential for the resolution of cognitive conflicts. Default values, which are negative for actions and atypical facts, and the fact that unknown terms do not block deductions (as *wood_wrecked* in C1 when B1 is produced) let the door open to knowledge revision when new elements are introduced. For instance, after introducing *wire_brush*, which was abducted to get out of the conflict *tough_work* / *not_tough_work*, the term *tough_work* is no longer true because C4 becomes blocked.

Our last remark is about the deterministic aspect of necessity management. This is the main difference between this approach and planning programs, which use heuristics and a nondeterministic goal generation. What plays the role of goals, in our model, is the weaker term of a conflict. The determination of goals is thus fully deterministic. Nondeterministic aspects are limited to the abducted algorithm. Even there, the search is strongly constrained, since the effect is given as input. The use of knowledge is thus progressive and controlled, which allows to consider the use of realistic size knowledge bases.

6 Cognitive Plausibility of the Model

The first positive argument for the cognitive plausibility of our approach is its simplicity. The few principles used are quite basic from a psychological point of view. Our abductive ability, though still mysterious in its functioning, is recognised as a powerful human ability (Johnson et al. 1994, Josephson & Josephson 1994). The few principles underlying necessity management are quite natural hypotheses. Differential necessity values and inheritance through causal links can be inferred from knowledge revision phenomena (Castelfranchi & D'Aloisi, 1991). We are indeed less prone to believe that D. Scarlatti met Mozart if we know that he was born the same year as Bach. Our ability to evaluate and to solve conflicts can be seen in our capacity to balance pros and cons. Necessities are also apparent in our use of modal expression (as in "you have to use a wire brush").

The present model has been inspired by the observation and analysis of natural conversations (Dessalles, 1993). Most spontaneous conversations are organised around cognitive conflicts that

are made manifest when new topics are introduced. Some conflicts can be described as apparent logical impossibilities, i.e. paradoxes. This is what happened in the conversation about Goffman's books. Other conflicts are about undesirabilities, as was the case with our main example.

A third possibility, which can also be considered as a cognitive conflict, but which our model does not consider, is the mention of an improbable event. For instance:

(4) From Tannen (1984):

A1: Speaking of which they had the Loud Family. Remember the Loud Family? On Saturday Night Live? [TV program]

B1: What was the Loud Family?

A2: Dju hear about that? THEY TALK LIKE THIS.

B2: I know lots of people in New York who talk like that.

Neither A1-A2, which introduce an improbable story, nor B2, which attempts to lower the improbability, can be computed in our current implementation. This would require significantly different mechanisms based on qualitative probabilities.

Another limit of the current implementation comes from the fact that meta-knowledge is not available. This may be problematic in situations in which participants are uncertain about what others know. Also, utterances like A6 in the conversation about Goffman ("I had never heard about him before I started studying linguistics") are difficult to predict. A6 was just a way of making the logical context of the conversation more explicit. By generalising A's case, we obtain the fact that non-linguists do not know Goffman, which is roughly the context that made B's knowledge of Goffman paradoxical. A's motivation for uttering A6 might be to make her surprise more manifest to B. This kind of initiative is still beyond our modelling capabilities.

7 Conclusion

We presented here a model of argument generation based on the notion of cognitive conflict. Such a conflict arises when several beliefs are incompatible, or when some fact considered as necessary is incompatible with another fact which is desired. Such conflicts are made manifest during conversation, and participants try together to solve them. Quite often a solution creates a new conflict. This is the recursive aspect of dialogue, which may cause some conversations to last for half an hour on the same topic (as in one example of our corpus which consists of 355 utterances). The basic mechanisms involved in our model were kept simple. When a cognitive conflict is recognised, participants try to make abduction from the weaker term. If this fails, they give up and assert the falsity of this term. This generally causes the conflict to shift to other terms. The new conflict is handled the same way, until a satisfactory solution is found or, as was the case in our main example, until participants are unable to make further abductions.

Further work is required to include meta-knowledge and probabilities into the model. We also plan to link the model to a proper semantic module that would allow a dynamic production of causal links. Another objective is to make the implementation robust enough to use it in real human-machine interactions, like Computer Assisted Learning dialogues.

References

- Airenti, G., Bara, B. G., and Colombetti, M. (1993). Conversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17:197-256.
- Baker, M. (1991). An analysis of cooperation and conflict in students' collaborative explanations for phenomena in mechanics. In Tiberghien, A. and Mandl, H., editors, *Knowledge Acquisition in Physics & Learning Environments*. Springer Verlag.
- Castelfranchi, C. and D'Aloisi, D. (1990). A cognitive approach to knowledge revision. In Kohonen, T. and Fogelman-Souli, F., editors, *COGNITIVA 90 -Proceedings of the Third Cognitive Symposium (Madrid)*, pages 31-38, Amsterdam. North Holland, ed. 1991.

- Dessalles, J.-L. (1993). *Modle cognitif de la communication spontane, appliqu l'apprentissage des concepts - PhD Thesis*. ENST - 93E022, Paris.
- Flowers, M., McGuire, R., and Birnbaum, L. (1982). Adversary arguments and the logic of personal attacks. In Lehnert, W. G. and Ringle, M. H., editors, *Strategies for Natural Language Processing*, pages 275-294, Hillsdale N.J. L.E.A.
- Goodwin, C. and Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology*, 19:283-307.
- Guez, S. (1990). *A Computational Model for Argument Understanding*. 13th International Conference on Computational Linguistics (Coling 90), Vol 3.
- Hirst, G. (1991). Does conversation analysis have a role in computational linguistics. *Computational Linguistics*, 17(2):211-239.
- Johnson, T. R., Krems, J., and Amra, N. K. (1994). A computational model of human abductive skill and its acquisition. In Ram, A. and Eiselt, K., editors, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 463-468, Hillsdale, N.J. Lawrence Erlbaum Associates.
- Josephson, J. R. and Josephson, S. G. (1994). *Abductive inference: computation, philosophy, technology*. Cambridge University Press.
- Kambhampati, S. (1997). Refinement planning as a unifying framework for plan synthesis. *AI magazine*, pages 67-97.
- Liddicoat, A. (1995). Argumentation as an interactional process in conversation. *Australian review of applied linguistics*, 18(2).
- Moeschler, J. (1990). *Conversation: cohrence et pertinence*. CNET, Paris.
- Quilici, A. (1992). *Arguing about planning alternatives*. Coling-92 - Proceedings of the fifteenth International Conference on Computational Linguistics, Nantes.
- Sadek, D. (1996). *Le dialogue homme-machine : de l'ergonomie des interfaces l'agent intelligent dialoguant*. Arago 18. OFTA, Paris.
- Tannen, D. (1984). *Conversational Style - Analyzing Talk Among Friends*. Ablex Publishing Corporation, Norwood.
- Young, R. M., Moore, J. D., and Pollack, M. E. (1994). Towards a principled representation of discourse plans. In Ram, A. and Eiselt, K., editors, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 946-951, Hillsdale, N.J. Lawrence Erlbaum Associates.

Automatic Evaluation of Spoken Dialogue Systems

Wieland Eckert, Esther Levin, Roberto Pieraccini
AT&T Labs Research
180 Park Avenue
Florham Park, NJ, 07932
E-mail: eckert@research.att.com

Abstract

We advocate an objective evaluation methodology for the automated evaluation of spoken dialogue systems that eliminates manual interaction and reduces annotation errors and personal bias. The evaluation is performed by observing interactions between the system and a simulated user. We argue that user simulation is an inexpensive and feasible method for optimizing a dialogue system in the lab. Using a simulated user we can conduct dialogues until the performance measure reaches a predetermined confidence level. A simulated user not only exercises the dialogue system and points out defects, it also helps predict the success of a modified dialogue strategy.

1 Introduction

An increasing number of speech dialogue systems are being deployed (Fraser and Dalsgaard 1996; Gorin et al. 1996; Lamel et al. 1996; Kellner et al. 1996; Pieraccini et al. 1997; Sadek et al. 1996). These systems deal with information retrieval tasks using unconstrained speech. However, there are no commonly agreed methods to evaluate a system, or to compare the performance of competing systems for the same application. Several approaches have been proposed (Simpson and Fraser 1993; Danieli and Gerbino 1995; Hirschman and Thompson 1996; Walker et al. 1997), but the current literature shows that none of them has been commonly adopted by system developers. Still, everyone reports performance measures using different metrics. Recent activities of the EAGLES groups (Gibbon et al. 1998) have not resolved this problem.

We tackle some of these difficulties by proposing an *automatic evaluation framework* that uses a *simulated user* to interact with the dialogue system. Using a simulated user that interacts with the system, we obtain results that are unbiased, accurate, and inexpensive. Evaluation is done automatically and therefore quickly, and with less effort than manual evaluation. Comparison of different dialogue systems (or different strategies within the same system) is simple and conclusive.

2 Dialogue Processing as a Feedback System

We adopt the view that a dialogue partner can be modeled as a feedback system. It is quite common to decompose this dialogue system into several modules which resemble the different fields of research involved. Typically, we deal with speech recognition, natural language understanding, dialogue control, language generation, and text to speech synthesis modules. Figure 1 illustrates this approach of specialized modules. Furthermore, we assume that both participants are treated equally, i.e. *both* are described by interacting modules. Feedback to one dialogue partner is provided by the other participant. Each of them can be seen as a system that follows their particular goal and receives feedback which, in turn, influences the goals and strategies.

An idealized interaction would involve an error-free communication channel: Speech recognition would be exactly the inverse operation of speech production and information is neither lost nor

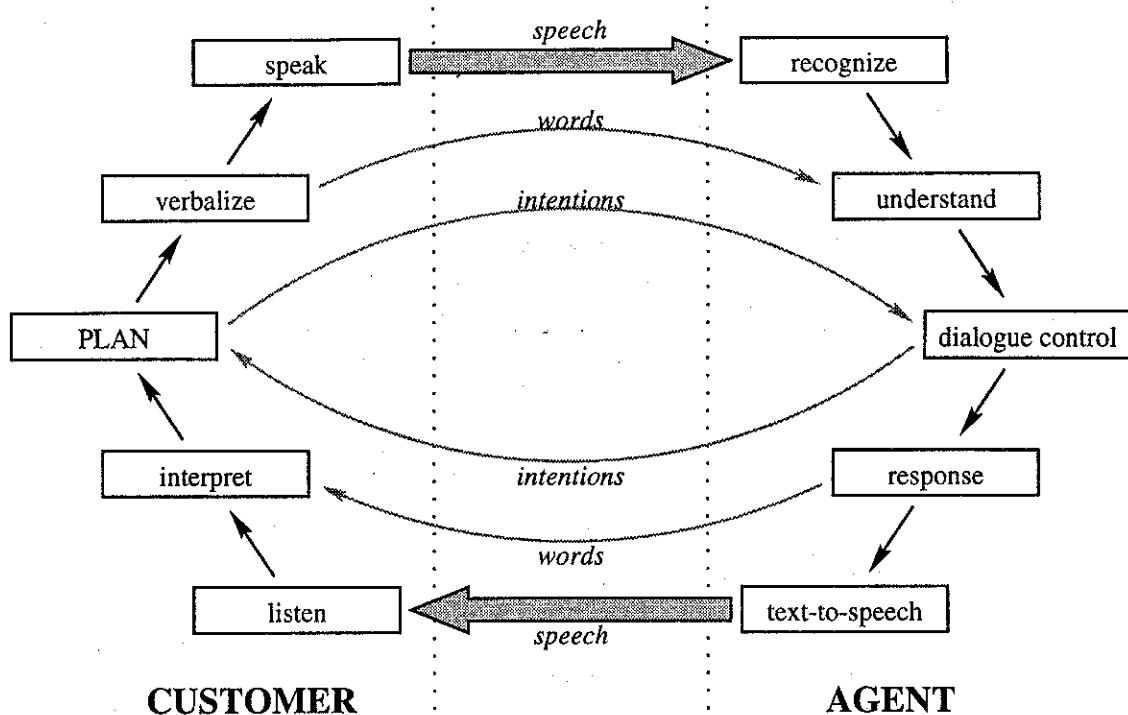


Figure 1: Dialogue interaction can be modeled using several different levels: speech signals, sentences, and intentions.

added. This reduction changes the view of dialogue interactions from exchanging acoustic speech signals to communicating via text (word sequences). In yet another step of simplification we may want to assume ideal understanding capabilities. Then we can model dialogues by means of exchanging information on the level of intentions (also known as *dialogue acts*). We began our research on this level of exchanging intentions. Having understood the benefits and limitations of an idealized interaction we started to extend our framework to deal with the understanding module as well. An example dialogue is shown in Figure 2. Corresponding intentional descriptions¹ in Figure 3 are only given to illustrate the data exchange mechanism.

It is clear, that *both* dialogue partners are

- **responsive**, they consider the current input for generating their output;
- **adaptive**, they change their goals (global strategies) or their methods (local strategies) according to the dialogue history;
- **goal driven**, a customer wants some information, an agent wants a happy customer;
- **cost conscious**, they use some measure of the "cost" or "reward" of different alternatives, usually described by the overall satisfaction of the speaker.

The assessment of the quality of a dialogue system needs to eliminate as many variabilities as possible. A most general approach is to only observe the external behavior (black box evaluation) without detailed knowledge about internal proceedings. The only method to gain insight about a system without knowing internals is by exploration (probing the behavior), i.e. conducting dialogues with the system. White box evaluation would need to describe and evaluate internal processing. Details about internal procedures are not really helpful for comparative evaluation of different dialogue systems. Furthermore, any comparison between different dialogue systems

¹In this paper we refrain from taking a position about "good" or "bad" formalisms or semantic representations.

System:	This is the AT&T CHRONUS Air Travel Information System. How may I help you?
User:	I need a flight from Philadelphia with T W A.
System:	Can you please tell me your destination airport?
User:	To Oakland
System:	There are no flights from Philadelphia to Oakland served by Trans World Airlines. Instead there are flights served by American Airlines, Delta Air Lines, United Airlines and US Air. Do you have any preference?
User:	either one is fine
System:	Oakland is served by 2 airports: Metropolitan Oakland International and San Francisco International. Please choose one.
User:	to Oakland
System:	Oakland is served by 2 airports: Metropolitan Oakland International and San Francisco International. Please choose one.
User:	Oakland airport
System:	Flight UA355/UA673 leaves at 5 53 PM and arrives at 10 05 PM. Flight UA131/UA551 leaves at 9 05 AM and arrives at 1 39 PM. Other questions?
User:	no thanks

Figure 2: Example dialogue in the domain of flight information inquiries.

must also consider the user populations that were used to evaluate these systems. Objective comparison is possible when the same population is used for both systems. Thus, any systematic bias is minimized.

3 Automatic Evaluation

We advocate an objective description of a dialogue system's capabilities. Ideally, one would like to specify a *characteristic function* or *transfer function* of this system (cf. control theory). While we usually don't know the transfer function of a complex dialogue system, we can get a glimpse of it by exploration. This is exactly what a corpus of dialogues shows: the results of applying the system's transfer function to a distribution of inputs. An external observation mechanism can record and extract features of the unknown transfer function.

Since evaluation is based on interactions between the system and users, we have to make sure that these interactions are typical of the projected use of the system, i.e. that the actual operating point of the system is observed. As in all experimental setups, we face the dilemma of considering the user's ability to "learn the ropes", i.e. to adapt to the system's strategies. Some evaluation methodologies ignore the user's adaptation by considering only novice or expert users. Using an automatic evaluation approach we can specify characteristics of a *typical* user population that contain various different behavior patterns, ensuring a proper mix of novice, intermediate, and expert users.

Figure 4 shows a block diagram of our evaluation environment. We observe the information exchange between an existing dialogue system and a simulated user. According to a population profile a particular set of values of simulation parameters is picked and used to drive the user simulation. This simulation is terminated when a predetermined number of dialogues is conducted. Another useful termination criterion is the stability of the evaluation results, e.g. when the performance measure reaches a particular confidence level.

Definition of a single, global *performance* figure is not trivial since it may vary depending on the customer's needs. We propose to use a compound quality measure $q(d)$ of a dialogue d that assigns a single value to a dialogue. This compound measure might consist of several components that are weighted according to the customer's demands. For our experiments we used a compound

```

S: ANNOUNCE: OPEN
U: SUBJECT: FLIGHT
  ORIGIN_CITY: PPHL
  AIRLINE: TW
S: REQUEST_CONSTRAINT: DESTINATION
U: DEST_CITY: OAK
S: ANNOUNCE: FAIL
  RELAXED_CONSTRAINT: AIRLINE
  CHOICE_TOKEN: AIRLINE
  ORIGIN_CITY: PPHL
  DEST_CITY: OAK
  AIRLINE: TW
  DATA: AIRLINE: {AA | DL | UA | US}
U: RELEASE: AIRLINE
S: CHOICE_TOKEN: DEST_AIRP
  DEST_CITY: OAK
  NDATA: 2
  DATA: DEST_AIRP: {OAK | SFO}
U: DEST_CITY: OAK
S: CHOICE_TOKEN: DEST_AIRP
  DEST_CITY: OAK
  NDATA: 2
  DATA: DEST_AIRP: {OAK | SFO}
U: DEST_AIRP: OAK
S: DATA: ORIGIN_TIME:1753 DEST_TIME:2205 FLIGHT_NUMBER:UA355/UA673
  DATA: ORIGIN_TIME:905 DEST_TIME:1339 FLIGHT_NUMBER:UA131/UA551
  YESNO: QUESTION
U: DUMMY: NO

```

Figure 3: Intentions of the dialogue in Figure 2. The representation consists of (sequences of) attribute value pairs that describe the information content conveyed in the utterance.

measure that is a weighted sum of simple components c_i ,

$$(1) \quad q(d) = \sum_i k_i c_i(d)$$

where the components c_i specify elementary cost functions that can be determined easily for a particular dialogue. Examples for costs are the dialogue length in turns, cost for database access, or penalty for not replying with an answer. While it is still not obvious how to specify the weights k_i , a reasonable approach shown in (Walker et al. 1997) is to select predictor coefficients that have found most predictive for an overall user satisfaction.

The overall performance measure $perf(D, U)$ of a dialogue system D interacting with a user population U is then the expected value of the quality function $q(d)$. It is calculated by weighting the performance for each possible dialogue d with the probability that it occurs, and can be approximated by evaluating this integral over a small, representative corpus C .

$$(2) \quad perf(D, U) = \mathcal{E}_d\{q(d)\}$$

$$(3) \quad = \int_d P(x) q(x) dx$$

$$(4) \quad \approx \sum_{d \in C} \frac{1}{|C|} q(d)$$

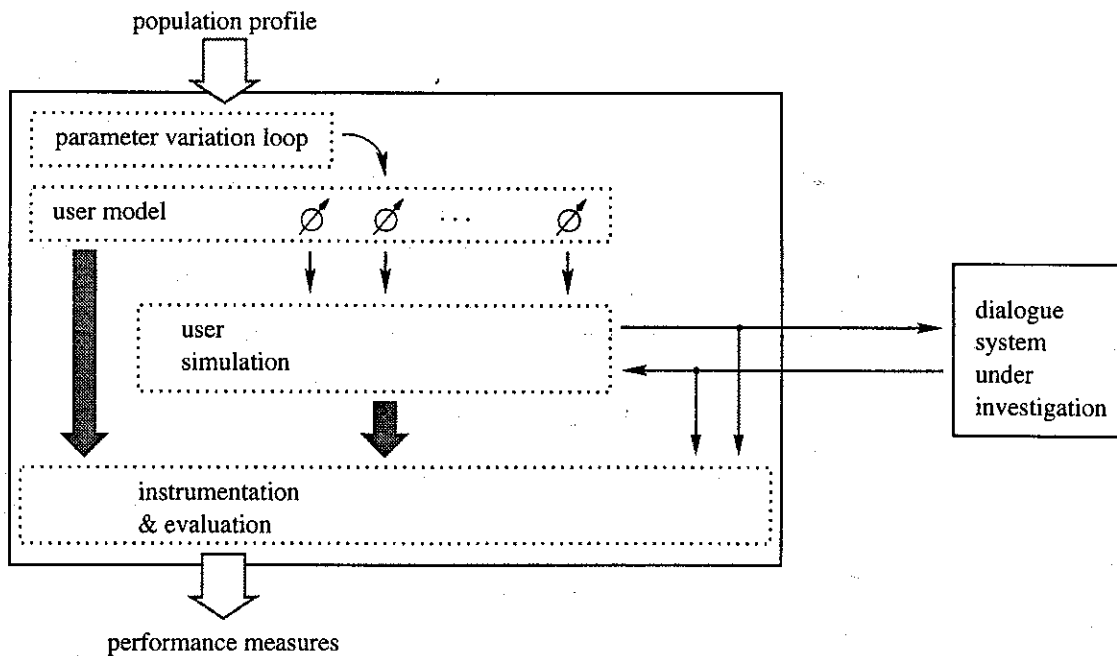


Figure 4: Block diagram of our evaluation environment.

All corpus based evaluation methodologies have to use the latter approximation and they have to rely on C being representative. In fact, the approximation $P(d) \approx 1/|C|$ is coarse and can only hold for large corpora with duplicated entries. Otherwise one has the added inconvenience of estimating $P(d)$.

Note, that a dialogue system does not allow for arbitrary dialogues, it follows a strategy. In general, a dialogue system has the property of stabilizing a dialogue, i.e. after any unexpected or unintelligible user response the system has a strategy dictating how to continue in order to reach the dialogue goal. The same holds for the user's response. Both are caused by the fact that the dialogue partners want to reach their respective goals. Cooperative dialogue partners have even fewer opportunities to diverge since their goals are correlated. This means that for cooperative behavior (of both user and system), the approximation (4) will converge much faster than for non-cooperative behavior. In other words: the evaluation of stereotypical (i.e. uninteresting, dumb) dialogues is easier than the assessment of diverse (i.e. interesting, intelligent) dialogue behavior.

4 User Simulation

The automatic evaluation approach outlined in the previous section can be applied to any existing corpus of dialogues. However, serious limitations are, that

- a corpus must exist, i.e. data was collected during user trials;
- the corpus must be representative, i.e. trials with users of the projected population, not the designers of the system;
- the corpus must be large enough to obtain significant results.

Additionally, for every slight change in the system's dialogue strategy a new data collection is necessary, because even with only a different prompt users may respond in a different way, resulting in a big change of the dialogue continuation.

We present a new approach for dealing with these problems: We employ a *simulated user* that interacts with the dialogue system. Simulation overcomes all the above mentioned limitations of

a static corpus. We can conduct an arbitrary number of dialogues, and for a modified dialogue system we can rerun the simulation easily. We can imagine having an infinite corpus of dialogues that is generated on demand.

There is no doubt that adequate simulation can accurately represent all the properties of the underlying sample the simulation model is based on.² In order to generate "interesting" dialogues we want to have some variety, not stereotypical behavior, i.e. we need nondeterministic models. Some generalization is obtained when dialogues are described in terms of their statistical properties, which also ensures consistent dialogue behavior with some predetermined number of outliers. This is the reason why we do not want to employ rule-based simulation models. Rule-based approaches are, in general, used to model the ideal behavior of a system. In this context we need, however, a model that is capable of representing variations from ideal behavior. Actually, we need a model that describes *real users* as accurately as possible in order to generate reliable performance figures.

The main problem is to find a sound, stochastic representation of dialogue behavior. Dialogues can be described at several levels of abstraction. Most commonly used are the levels of speech (acoustic signals, utterances), syntax (lexical, word sequences, sentences), and intentions (semantics, dialogue acts). These levels coincide with the modularization of most current dialogue systems containing a speech recognizer, natural language parser, and a dialogue manager component, cf. Figure 1. If we adopt a stochastic simulation of user responses, we might want to factor the resulting probabilities into

$$(5) \quad P(\text{utterance}) = P(\text{acoustics}|\text{syntax, semantics}) \\ \cdot P(\text{syntax}|\text{semantics}) \\ \cdot P(\text{semantics})$$

for any utterance, assuming independence of these description levels. Furthermore, we have to introduce conditions on the dialogue context

$$(6) \quad P(\text{utt}_{\text{user},t}|\text{utt}_{\text{system},t-1}, \text{utt}_{\text{user},t-2}, \text{utt}_{\text{system},t-3}, \dots)$$

that represent the *dialogue strategy*. This representation becomes tractable when we make assumptions about conditional independence of representational levels, and limitations of dialogue context to a length of n turns

$$(7) \quad P(\text{utt}_{\text{user},t}|\text{utt}_{\text{system},t-1}, \dots, \text{utt}_{\dots,t-n})$$

The length n of previous utterances considered for the current response might be considered the user's short term memory.

In our implementation we limit the context length to $n = 1$, i.e. we make a user response dependent only on the previous system utterance. Using the descriptive level of intentions, we have to determine the bigrams

$$(8) \quad p_{ij} = P(\text{utt}_{\text{user},t} = I_i|\text{utt}_{\text{system},t-1} = I_j)$$

of replying with a response I_i when receiving the stimulus I_j from the system (I_i and I_j denote sets or sequences of intentions). We assume that the process is time invariant, i.e. these probabilities do not depend on the absolute value of t . This bigram model can be extended canonically to cover longer context lengths.

Additional parameters in our user modeling approach deal with conversational customs, like After a dialogue lasts more than X turns a user just hangs up unhappily or In the initial utterance the user gives Y pieces of information without being asked for them. Other parameters are the probability of an over-informative response, or the probability of going into a sub-dialogue. Again, these parameters are modeled by (normal) densities which are specified for a population.

Using stochastically generated simulated user interactions we can run a large number of dialogues and reach results that are significant according to a predetermined confidence level. Due

²Proof: one can store the whole sample and show exact behavior.

to the nature of stochastic modeling we can be sure that all results will eventually converge, and we can run experiments until a predefined confidence level is reached.

Note, that in equation (3) we consider all possible dialogues. Since the dialogue system and the user both constrain the set of dialogues, we can rewrite this equations and calculate the performance of a dialogue system by considering all possible user populations, i.e. we calculate the expected value

$$(9) \quad perf_D = \mathcal{E}_U\{perf(D, U)\}$$

of the system D conducting a dialogue d with any user u of any population U . Without loss of generality, we can actually consider each user having only a single dialogue with the system when we consider a simple renaming of any subsequent dialogue of the same individual. This assumption collapses all possible user populations into a single, representative population. User's learning experiences are not hurt since this is already represented in the statistics of the population. For the matter of evaluation a dialogue system it does not matter if a subsequent dialogue is conducted with the same, now more experienced individual, or with a different one who is better accustomed to this task in the first place.

5 Obtaining Simulation Models

Ideally, one would extract the probabilities of these stochastic models from a corpus of human-machine dialogues. Learning algorithms that are well understood, e.g. for the extraction of language models for speech recognizers, can also be applied to estimate the parameters of our user models. However, we face the problem that a given corpus might be too small for reliable estimates of the model parameters. Either we have to work with bad estimates, or reduce the model complexity. Simpler models can be obtained, for instance, by reducing the context length n , or by making additional independence assumptions. The trade off is either having a sophisticated model using unreliable parameters, or a simplified model with highly reliable parameters. This dilemma is well known.

The worst case for obtaining simulation models is having no dialogue corpus at all. This typically happens when a new dialogue application is envisaged and no previous studies are available. In this case, a simple model with unreliable parameters can be built manually by guessing typical user reactions. Even though this model would shift the operating point of the system substantially, a properly designed dialogue system would try to "help this inexperienced user" and respond accordingly, thus stabilizing the dialogue. We see the whole process of evaluating a dialogue system as an incremental process going through several iterations, starting with a manually generated model based on "common sense" when a dialogue corpus is unavailable or too limited.

In our case we have the official ATIS corpus at hand. While investigating this corpus we faced the following problems:

- While the corpus is large enough to estimate language models (i.e. working on word sequences), on the intentional level it is sparse.
- In the ATIS corpus all information is usually given in the initial user utterance (class A) and there are rarely any followup utterances necessary (class B). As a consequence there is a large imbalance of training material for initial and follow-up utterances.
- The characteristics of the ATIS corpus differ from real user dialogues. ATIS dialogues are quite short and contain a lot of information (kind of complete inquiries) in each utterance, whereas our experiments show longer dialogues with less information per utterance.

We concluded that the official ATIS corpus is based on a very different dialogue strategy, and therefore of limited use. Class A sentences are a rich source for estimating parameters for initial utterances, but a large class of probability densities is not represented in this corpus at all, for instance yes/no-questions or simply requesting the departure time from the user. Thus we had to

hand craft some of the probabilities but have been inspired by the characteristics of the corpus and by common sense.³ While we are not satisfied with this kind of guesswork, we see it as a bootstrap process for obtaining better models. Standard machine learning textbooks ((Sutton and Barto, 1998, Section 9.2)) show how to obtain an improved model of the user while interacting.

Even with only rough estimates of response probabilities we found that these simulated dialogues are perfectly reasonable, i.e. we can easily explain the behavior and draw parallels to real user's behavior. We attribute that to both partners having the property of following their goals. A minor or even a major deviation has no catastrophic effects on the dialogue, although it is still visible in the performance results.

In equation (9) we calculate the performance of a dialogue system dealing with a variety of users. As a dual problem one could calculate the performance of a user population that is exposed to a variety of dialogue systems

$$(10) \text{perf}_U = \mathcal{E}_D\{\text{perf}(D, U)\}$$

A simulation model S would have to show the same performance as the underlying population U of real users. Thus, a metric for the quality of a simulation model is the mean square error.

$$(11) E(S, U) = \mathcal{E}_D\{\text{perf}(D, S) - \text{perf}(D, U)\}^2$$

A proper model design and refinement technique would try to minimize this kind of model approximation error. If this error approaches 0, it means that simulated dialogues and interactions with real users are indistinguishable. Another point of view is to consider (10) as the degree of goal correlation between the system and the user. Collaborating agents usually reach a higher performance measure than opposing agents. If we assume collaborating agents, a drop in performance within a dialogue may be caused by misrecognition or misunderstanding. Therefore, a dialogue system may compare a user's actual response with the predicted response to adapt the system's strategy in order to reach better overall performance.

6 Utility

There are several reasons for employing automatic evaluation methods and interactions with simulated users:

- We want an *objective* methodology that eliminates any bias of manual evaluation.
- Automatic evaluation is inexpensive and quick. Large amounts of data can be processed more quickly and reliably than with human intervention.
- Automatic evaluation using simulated users enables comparative evaluation of different dialogue systems, or different strategies of the same system.
- Improvements of a dialogue system can be performed in the lab. Costly trials with real users can be reduced.

The first two items are quite obvious since they follow directly from the automation of labor expensive and tedious manual work.

While there is considerable literature about evaluation of a particular dialogue system, comparative evaluation of two or more systems has not been tackled, yet. Even for the same application (e.g. ATIS) there was no comparison possible since researchers adopted different dialogue strategies, experimented with different user populations, and reported different performance metrics. Using a simulation environment like ours, one can connect to a variety of dialogue systems and run experiments with a consistent population of simulated users.

This framework allows for documenting the improvement of a dialogue system in the lab. In the following we outline the reasoning behind this method. Let's assume we have an automatic dialogue

³We will have a better foundation for these estimates when actual dialogue corpora are available.

system D_1 and we collect a corpus $C_U(D_1)$ of dialogues with real users. Based on that corpus a simulation model $S(C_U)$ can be built. Now that we have a user simulation tool and an automatic evaluation environment, we are able to optimize the existing dialogue system according to some given objective performance measure. The resulting system D_2 is better on the simulation model, i.e. $perf(D_2, S) > perf(D_1, S)$. This optimized dialogue system can be exposed to real users and a test corpus $C_U(D_2)$ is collected. We claim that the simulation is useful, iff the performance is not worse than that of the previous system, i.e. $perf(D_2, C_U(D_2)) \geq perf(D_1, C_U(D_1))$. Even though the simulation model is not necessarily *accurate* in the sense that it represents the user population *exactly*, it is still useful for optimizing an existing dialogue system. Most likely this process will be iterated several times, leading to better dialogue systems as well as better simulation models.

Keep in mind that the whole optimization can be done in the lab and only for testing purposes we do have to validate the system using interactions with real users. On the other hand, we can easily test different dialogue strategies in the lab and get an idea about their estimated performance in the field.

7 Our Implementation

We started our experiments in the ATIS domain (Eckert et al. 1997) and on the intentional level, i.e. a simulated user dealing with a dialogue manager in terms of intentions. These experiments essentially exercise the dialogue manager's ability to complete a dialogue, even with unorthodox user behavior. It assumes perfect recognition and perfect understanding. The majority of the dialogues went well, but a number of problems in the system's dialogue strategy were found and fixed. These problems occurred when the stochastic simulation encountered a situation that was not foreseen and never exercised by the developers. Thus, simulation did help debugging and optimizing the dialogue manager.

One defect in our dialogue strategy was, for example, that the system was too polite and never hung up on the user. The system would repeat infinitely that it did not find a solution for the user's particular request, but suggested a relaxation of some parameter, e.g. airline, instead. It is a perfectly reasonable strategy to repeat this information only n times and give up on a stubborn user. We found these defects by looking at the compressed performance statistics of a simulation run. This particular defect showed its footprint in an unusual large amount of dialogues with very many turns. After looking at one of these long dialogues we explained the defect and fixed it easily. We performed more experiments with different user populations (Eckert et al. 1997) in order to exercise the AMICA dialogue manager (Pieraccini et al. 1997). These simulations provided us with valuable hints for improving the strategies of the system.

Next, we went to the syntactic level and sent word sequences to the natural language understanding component which passed on its results to the dialogue manager. Thus, we test the robustness of NLU in combination with dialogue management, assuming perfect speech recognition. An example dialogue of our simulation experiments is shown in Figure 5. Again, we found several shortcomings of our NLU component that did not show up in previous tests with expert users. Even though we only used a very simple, template based text generation method to transform the user's intentions into word sequences, we found a few instances of simulated hesitations and interruptions where the NLU component could not extract all intentions of the user. As a result, these dialogues tend to take more turns than those assuming perfect understanding. Ideally, one would expect identical performance when transmitting intentions or when communicating in word sequences. The task of a robust speech understanding system is to minimize this difference as far as possible.

As a logical consequence one may think of using synthesized speech for the simulated user in order to exercise the complete speech dialogue system with its speech recognizer. While this approach is definitely beneficial to the understanding of interacting using complex media, we think that there are still many open research issues to be solved before we benefit from this kind of simulation results.

System:	This is the AT&T CHRONUS Air Travel Information System. How may I help you?
User:	Can you get me a flight from Denver to Pittsburgh Airport?
System:	There are too many flights that match your request. I need additional information. Flights from Denver to Greater Pittsburgh International are served by 2 airlines: Delta Air Lines and US Air. Do you have any preference?
User:	The cheapest
System:	At what time do you want to leave?
User:	The earliest flight
System:	Flight DL296/DL842 leaves at 0 55 AM and arrives at 10 AM.
User:	Other questions?
User:	hangup

Figure 5: Actual Example dialogue that was conducted using our simulated user.

8 Discussion

Evaluation of a (speech) dialogue system is like exploring the unknown. Using a black box approach, we don't have an analytical description of the exact behavior of the dialogue system. Additionally, we can only gather statistical data of our user population. Therefore, we can not calculate the performance of the system using an analytical approach, but a stochastic approach to evaluation dialogue systems is necessary. We have to probe the dialogue system. In order to control as many parameters as possible, we employ a simulated user with stochastically well defined behavior. User simulation and automatic evaluation is the key to enable comparative evaluation of dialogue systems.

Simulation using stochastic models is a well known field. Our original work is the application of stochastic user simulation to evaluate dialogue systems, and the automatic evaluation of the simulation results. Stochastic modeling is an assumption that is made in our approach, but these models adequately represent the underlying corpus.

One of the first evaluation methodologies for speech dialogue systems was described in (Simpson and Fraser 1993). While it intended to enable comparative studies of different dialogue systems, It did not address the problem of diverse user populations. Any experimental result is prone to some bias caused by the user sample picked. In order to normalize results one would have to determine the prior distributions of user characteristics, and either factor out their influence or normalize to some standards. Using a simulated user we provide every examined dialogue system with the same user characteristics, repeatedly. Additionally, experiments with a different user population can be run easily.

Recent work in the field of machine learning techniques led to the development of an automatic, learning dialogue agent (Levin et al. 1997). This learning approach utilizes a cost measure for any action taken and finds an optimal strategy how to interact with a user. While the idea is fascinating, the drawback is that a very large number of dialogues is necessary to learn a good strategy. Obviously, one would not take advantage of human dialogue partners (especially if one might loose customers), so a simulation approach is necessary. We were able to show in (Levin et al. 1997) that user simulation is a useful strategy for learning dialogue strategies.

9 Summary & Further Work

In this paper we presented a new approach toward the automatic evaluation of speech dialogue systems. This approach alleviates two of the dominant difficulties: (1) manual work is greatly reduced by employing an *automatic evaluation process* which leads to less transcription errors and higher processing speed, and (2) comparative evaluation is possible because of a *simulated user* interacting with the system in a consistent way following stochastic constraints.

We showed that a sound model of user behavior can be deduced using stochastic properties

that can be determined from a corpus. Factorization of conditional probabilities into the layered levels of acoustic, syntactic, and semantic knowledge representations gives us a decomposition into tractable sub-problems. We started our experiments on the level of semantic information, i.e. the dialogue system and the user simulation both communicate descriptions of their intentions. Experiments to extend this to the next level of text strings are in progress. We used the ATIS task as a test bed for our experiments and connected this evaluation environment to our AMICA system. Our experiments were successful in the sense that we found (and fixed) several shortcomings of the dialogue system and of the natural language understanding module. Even with very rough approximations of user models we were able to show very reasonable behavior of the user simulation. Additionally, we employed Reinforcement Learning to obtain a dialogue strategy that is optimal with respect to the characteristics of the simulation population.

Future work will tackle the problem of obtaining good simulation models. A very interesting approach is to extract the model while interacting with a real user. This user model would be exploited in idle times to provide the grounds for "what-if" simulations which would improve the system's performance for previously unseen dialogues. One would need to leap from hand crafted dialogue strategies to automatic, learning systems. Finally, there is hope that speech dialogue systems can become intelligent.

References

- Danieli, M. and Gerbino, E.: 1995, Metrics for Evaluating Dialogue Strategies in a Spoken Language System, in *Proceedings of the 1995 AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pp 34-39
- Eckert, W., Levin, E., and Pieraccini, R.: 1997, User Modeling for Spoken Dialogue System Evaluation, in Furui et al. (1997), pp 80-88
- Fraser, N. M. and Dalsgaard, P.: 1996, Spoken dialogue systems: A European perspective, in ISSD 96 (1996), pp 25-36
- Furui, S., Juang, B.-H., and Chou, W. (eds.): 1997, *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*, Santa Barbara
- Gibbon, D., Moore, R., and Winski, R. (eds.): 1998, *Handbook of Standards and Resources for Spoken Language Systems*, Vol. 3, Spoken Language System Assessment, Walter de Gruyter, Berlin
- Gorin, A. L., Parker, B. A., Sachs, R. M., and Wilpon, J. G.: 1996, How may I help you?, in IVTTA 96 (1996), pp 57-60
- Hirschman, L. and Thompson, H.: 1996, Overview of evaluation in speech and natural language processing, in R. Cole (ed.), *Survey of the State of the Art in Human Language Technology*, Cambridge University Press, Cambridge
- ISSD 96: 1996, *Proceedings International Symposium on Spoken Dialogues*, Philadelphia, Acoustical Society of Japan
- IVTTA 96: 1996, *Proceedings of the IEEE Third Workshop on Interactive Voice Technology for Telecommunications Applications*, Basking Ridge, IEEE Communication Society
- Kellner, A., Rueber, B., and Seide, F.: 1996, A voice-controlled automatic telephone switchboard and directory information system, in IVTTA 96 (1996), pp 117-120
- Lamel, L. F., Gauvain, J. L., Bennacef, S. K., Devillers, L., Foukia, S., Gangolf, J. J., and Rosset, S.: 1996, Field trials of a telephone service for rail travel information, in IVTTA 96 (1996), pp 111-116

- Levin, E., Pieraccini, R., and Eckert, W.: 1997, A Stochastic Model of Computer-Human Interaction for Learning Dialogue Strategies, in Furui et al. (1997), pp 72-79
- Pieraccini, R., Levin, E., and Eckert, W.: 1997, AMICA: The AT&T Mixed Initiative Conversational Architecture, in *Proc. European Conf. on Speech Communication and Technology*, pp 1875-1878, Rhodes, Greece
- Sadek, M. D., Ferrieux, A., Cozannet, A., Bretier, P., Panaget, F., and Simonin, J.: 1996, Effective human-computer cooperative spoken dialogue: The AGS demonstrator, in *ISSD 96 (1996)*, pp 169-172
- Simpson, A. and Fraser, N.: 1993, Black Box and Glass Box Evaluation of the SUNDIAL System, in *Proc. European Conf. on Speech Communication and Technology*, pp 1423-1426, Berlin, Germany
- Sutton, R. and Barto, A.: 1998, *Reinforcement Learning, An Introduction*, MIT Press
- Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A.: 1997, PARADISE: A Framework for Evaluating Spoken Dialogue Agents, in *Proc. Conf. of the Association for Computational Linguistics*, pp 271-280, Madrid, Spain

Distributed Knowledge*

Jelle Gerbrandy
ILLC/Department of Philosophy
University of Amsterdam
gerbrand@phil.uva.nl

Abstract

In this paper, two definitions for the notion of *distributed knowledge* in possible worlds semantics are discussed and compared. In the first definition distributed knowledge is characterized by intersection of information states, the second defines the concept as ‘those sentences that are logical consequences of the beliefs of the agents.’ I will argue that the effect of the first kind of definition depends on the ontological view one has about possible worlds, and that the second kind of definition depends on the expressive power of the language. I will also show that the logic of the two operators is the same, in the sense that the logics have the same weakly sound and complete axiom system.

1 Distributed Knowledge

Consider a situation with two agents, a and b , and suppose that a has the information that p is the case, b knows that p implies q , but neither a nor b knows that q is the case. Even though neither one of the agents knows that q , there is a sense in which the information that q is the case is already present in their information states taken together: q is a logical consequence of the information that the two agents have. One way of formulating this is to say that the information that q is present in the ‘system’ consisting of both agents in a ‘distributed’ form: the information that q is distributed over information states of a and b . The standard term for such kind of knowledge is distributed knowledge.

The distributed knowledge between a and b is the information of a and b ‘together.’ To give a formal analysis of this notion, we need a notion of adding two information states. This is the central question of this paper: what does it mean to add the information contained in one information state to the information contained in another one?

We can also put the question in a more concrete form. Suppose both a and b each have a certain amount of information, and communicate everything they know to a third agent, who initially has no information at all. What is the information state of this third agent c after a and b have communicated everything they know to c ?

Kripke Semantics

I will study the question of how to define the notion of distributed knowledge using standard Kripke semantics. In particular, we will be concerned with giving a semantics for the following language:

Definition 1.1 Given a set of agents \mathcal{A} , and a set of propositional variable \mathcal{P} , the language \mathcal{L}^D of epistemic logic with distributed knowledge is given by the following definition:

$$\Phi := p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid D_B\phi$$

where $p \in \mathcal{P}$, $a \in \mathcal{A}$, and $B \subseteq \mathcal{A}$.

*This paper has benefited from comments of Joseph Halpern and Yde Venema. Responsibility for mistakes and misunderstandings is all mine, of course.

The set of sentences in which the operator D does not occur will be denoted by the symbol \mathcal{L} , the language of classical modal logic. We use the standard abbreviations: $\phi \rightarrow \psi$ for $\neg(\phi \wedge \neg\psi)$, and $\phi \vee \psi$ for $\neg\phi \rightarrow \psi$. \square

The language contains propositional variables \mathcal{P} , conjunction \wedge , negation \neg , and operator K_a and D_B for each $a \in \mathcal{A}$ and $B \subseteq \mathcal{A}$. The intended interpretation of a sentence of the form $K_a\phi$ is that agent a 'knows' or 'believes' that ϕ , and a sentence of the form $D_B\phi$ should be read as " ϕ is distributed knowledge between the agents of B " or " ϕ is implicitly believed in the group B ."

Here, the terms 'agent', 'knowledge' and 'belief' are to be read in a very loose sense. An 'agent' can be any kind of object for which it makes sense to say that it has certain information: humans, but also robots, database systems; and, in a more abstract way, computer processes and computer parts. Saying that human agents know and believe certain things is not a very controversial thing to do; ascribing knowledge and belief to more abstract agents is more controversial. The literature on computer science and artificial intelligence contains a fairly sophisticated and well-motivated theory of agenthood, with Fagin et al. (1995) providing a systematic method for ascribing knowledge to such abstract agents.

One way of providing a semantics for the language \mathcal{L}^D is in terms of Kripke models (Kripke 1963).

Definition 1.2

A *pointed Kripke model* (K, w) is a quadruple $(W, \{\overset{a}{\rightarrow}\}_{a \in \mathcal{A}}, V, w)$, where W is a set of possible worlds, w is a distinguished element of W (the point of evaluation), $\overset{a}{\rightarrow}$ is a relation on W for each $a \in \mathcal{A}$, V is a valuation function that assigns a truth-value (either 0 or 1) to each pair of a world $v \in W$ and a propositional variable $p \in \mathcal{P}$. \square

The definitions of satisfaction of sentences of \mathcal{L} in a model is as follows.

Definition 1.3

$$\begin{aligned} (K, w) \models p & \text{ iff } V(w, p) = 1 \\ (K, w) \models \phi \wedge \psi & \text{ iff } (K, w) \models \phi \text{ and } (K, w) \models \psi \\ (K, w) \models \neg\phi & \text{ iff } (K, w) \not\models \phi \\ (K, w) \models K_a\phi & \text{ iff for all } v \text{ such that } w \overset{a}{\rightarrow} v : (K, v) \models \phi \end{aligned}$$

The intuition behind this semantics as a semantics for epistemic logic derives from Hintikka (1962): we can model the belief of an agent as 'the set of possibilities that are compatible with the belief of that agent', or, alternatively but not incompatibly, as 'the set of possibilities that, with respect to the beliefs of the agent, could be (models of) the real world.' In a Kripke model, this intuition is reflected by the relations $\overset{a}{\rightarrow}$: if $w \overset{a}{\rightarrow} v$ in a Kripke model K , this means that v is epistemically possible for a in world w . In the following, I will often refer to the set of worlds v such that $w \overset{a}{\rightarrow} v$ as the *information state* of a in w , and sometimes write $w(a)$ for this set. The clause in the definition simply says that a believes that ϕ in world w just in case ϕ is true in all worlds in the information state of a in w .

It is often assumed that beliefs are introspective; that each agent knows exactly what he believes. This assumption corresponds to taking the following two axiom schemes as valid: $K_a\phi \rightarrow K_aK_a\phi$ (if a believes that ϕ , she knows that she believes that ϕ) and $\neg K_a\phi \rightarrow K_a\neg K_a\phi$ (if a does not believe that ϕ , she knows that she doesn't believe that ϕ). These two axioms are valid in all Kripke models in which the accessibility relations are transitive (if $w \overset{a}{\rightarrow} v \overset{a}{\rightarrow} u$, then $w \overset{a}{\rightarrow} u$ for all a, w, v and u) and euclidean (if $w \overset{a}{\rightarrow} v$ and $w \overset{a}{\rightarrow} u$, then $v \overset{a}{\rightarrow} u$). Another way to formulate this property is to say that each agent considers only worlds possible in which she is in exactly the same information state that she is actually in, i.e. that for each $v \in w(a)$, it holds that $v(a) = w(a)$.

Knowledge, as opposed to belief, also has the property of being *factive*, if a knows ϕ , then ϕ must actually be the case. This property is captured by the axiom scheme $K_a\phi \rightarrow \phi$, and corresponds

to the property that $w \xrightarrow{a} w$ for each w , i.e. that \xrightarrow{a} is reflexive for each a . Reflexivity says that each agent considers the actual world possible, i.e. that $w \in w(a)$ for each w .

Possible world semantics gives rise to a very idealized notion of belief and knowledge. In particular, it holds that if ψ is a logical consequence of ϕ , then in all models in which $K_a\phi$ is true –i.e. where agent a knows that ϕ is true– $K_a\psi$ is true as well. In other words, the belief (and the knowledge) of an agent is closed under logical consequence. This does not conform at all to the way these concepts are used in daily life: people do not always see the consequences of our beliefs, and know that other people are limited in the same way. This fact is particularly apparent when we look at mathematical truths: once we know the basic axioms of arithmetic, we do not know that Fermat’s last theorem is true, although it is a logical consequence of the axioms. This mismatch between the formal semantics and the pretheoretic concepts of knowledge and belief has been called the *problem of logical omniscience*, and different semantics have been proposed to solve it. The proposals to deal with the problem of logical omniscience are of a wide variety; some examples of semantics that can with some plausibility be called variations on the possible worlds view those that introduce ‘impossible possible worlds’, the so-called ‘awareness logics,’ and perhaps situation theory can be classified under this heading as well.

However, the operators K_a , even if they are not a perfect reflection of the notions of knowledge and belief, still can be of interest for a theory of the more abstract concept of *information*. This is what Barwise (1989) proposes: to use the term ‘information’ instead of ‘knowledge’ as the informal notion that best corresponds to the interpretation of K_a . In contrast with knowledge and belief, it does make sense to say that *information* is closed under logical consequence: if a sentence logically follows from the information you already have, then in some sense, you also have the information that this sentence is true, even though you may not be aware of it. In the following, I will use the terms ‘belief,’ ‘knowledge’ and ‘information’ more or less interchangeably; more precisely I will use the term ‘belief’ for what perhaps is better characterized as ‘information that may not be correct,’ and the term ‘knowledge’ for information that is also true.

Distributed information as intersection.

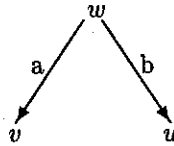
Let us consider first what the ‘inventors’ of the notion of distributed knowledge have to say. Halpern and Moses (1990) define the notion of distributed knowledge between two agents a and b in a Kripke model (K, w) like this:

Definition 1.4 For Kripke models (K, w) :

$$(K, w) \models D_{\{a,b\}}\phi \text{ iff } (K, v) \models \phi \text{ for all } v \text{ such that } w \xrightarrow{a} v \text{ and } w \xrightarrow{b} v$$

This notion is called ‘implicit knowledge’ in Halpern (1987). The intuition behind the definition is straightforward: in w , the worlds accessible by \xrightarrow{a} are the worlds compatible with the information of a and the worlds accessible by \xrightarrow{b} are the worlds compatible with the information of b . The information distributed between them can be characterized by the set of worlds that are compatible with both the information of a and that of b : those worlds accessible from w both by \xrightarrow{a} and \xrightarrow{b} . Reformulating this: the information that is distributed between two information states $w(a)$ and $w(b)$ is characterized by the intersection of $w(a)$ and $w(b)$. Or, as Fagin et al. (1995) put it: “we combine the knowledge of the agents in group \mathcal{B} by eliminating all worlds that some agent in \mathcal{B} considers impossible.”

What is not immediately obvious, perhaps, is that the plausibility of this definition depends very much on the ontological view one has on possible worlds. To illustrate this point, consider a very simple model K with three worlds w, v and u , two agents a and b , an accessibility relation $\xrightarrow{a} = \{(w, v)\}$ and $\xrightarrow{b} = \{(w, u)\}$, and a valuation function that the same truth-values to all propositional variables in v and u .



In this model, the set of worlds that a considers possible is a singleton set containing one world v in from which no further worlds are accessible. The set of worlds that b considers possible is also a singleton set containing one world, u , in which the same propositional variables are true as in v , and in which no further worlds are accessible. Of course, this example is not an example of a Kripke model that models the beliefs or the knowledge of an agent. One can easily give adapt the example to get a transitive and euclidean model: simply add a and b -edges $v \xrightarrow{a} v$, $v \xrightarrow{b} v$, $u \xrightarrow{b} u$ and $u \xrightarrow{a} u$. Changing the example into a proper model of knowledge, in which the accessibility relations are equivalence relations takes a little more work. Such an example can be found in van der Hoek et al. (to appear). For the sake of simplicity, I will stick to the original example in the following, and just disregard these issues. The argument I will give does not depend on them.

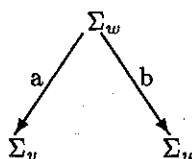
The information that a has in w is given by the singleton set $\{(K, v)\}$, and the information of b is given by the singleton set $\{(K, u)\}$. Since the two worlds are different, the intersection of state of a and that of b is empty, and hence, $(K, w) \models D_{\{a,b\}} \perp$: the distributed knowledge of the two agents is inconsistent. On the other hand $\{(K, u)\}$ and $\{(K, v)\}$ have exactly the same structure, in the sense that their generated submodels are isomorphic (in particular, this means that the same sentences of \mathcal{L}^D are true in the two models). The definition 1.4 of distributed knowledge, then, only makes sense if we take a view on Kripke models in which the difference between isomorphic worlds is somehow essential for the information that is represented by them: there must be something that distinguishes the world v from the world u that is relevant to the information of a and b in w , and this distinction is not visible by looking at the value of the propositional variables and the information states of the agents alone (since these are the same in u and v).

This is indeed the view that is implied by the framework that is developed by Fagin et al. (1992) and in the fourth chapter of Fagin et al. (1995). In this model, the possible worlds in a Kripke model have an internal structure themselves: they are descriptions of 'states', basically 'ways the world could be.' It is this internal structure of the possible world that makes the world 'what it is': the values of the propositional variables and the accessibility relations in the Kripke model are an extra logical layer that makes it possible to speak about possible worlds in the language of epistemic logic, but this extra layer may represent the internal structure of the possible worlds only partially. If the structure of the Kripke model (that is, the model *modulo* the worlds in that model) gives no clue as to how to distinguish two possible worlds, this is still no reason to consider them 'the same.' In the example above, the distributed knowledge of a and b should indeed be inconsistent under this conception: after all, their information states contain *different* worlds.

This view on Kripke models is very different then that of Hintikka, for example. Hintikka (1962) identifies possible worlds with *model sets*: descriptions of possible states of affairs in some logical language. For our purposes, we can identify a possible world with a set of sentences of \mathcal{L} that is maximal consistent in some logic, say K45, the modal logic of belief. We can then define a *Hintikka model* as a Kripke model in which each world is a maximal consistent set of sentences, and for each Σ and $\Gamma \in W$, it holds that if $\Sigma \xrightarrow{a} \Gamma$ and $K_a \phi \in \Sigma$, then $\phi \in \Gamma$.¹

Returning to our example, we can represent the Hintikka model that corresponds with it as follows:

¹I am just giving a rough definition to illustrate the point.

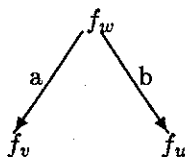


In this picture, Σ_w is the set of sentences that is true at w , and Σ_v and Σ_u are the sentences true at v and u respectively. Note that each sentence that is true at v is true at u as well, so drawing the Hintikka model as above is a bit misleading: the worlds Σ_v and Σ_u are in fact the same world. That means that the intersection of the worlds accessible for a from Σ_w with those accessible for b is not empty. In other words, in the Hintikka model that corresponds to our example, the distributed knowledge between a and b is not inconsistent.

The difference between the two viewpoints leads to a difference at the logical level as well. If our set of propositional variables is finite, we can describe the set of sentences Σ_w corresponding the world w up to uniqueness by a single sentence; Σ_w is finitely axiomatizable. Let ψ_w be that sentence. Our example shows that $\neg D_{\{a,b\}} \perp$ is a semantical consequence of ψ_w with respect to all Hintikka models, but not with respect to Kripke models.

Hintikka's view is not the only alternative to Kripke semantics. Elsewhere (Gerbrandy and Groeneveld 1997; Gerbrandy 1997, see also Barwise and Moss 1996), I have developed a semantics for epistemic logic that takes a slightly different take on the notion of a possible world in a Kripke model. The idea behind it is that a world in a Kripke model is a description of a possible way the world could be, and this description is completely exhausted by the values of atomic sentences and the information states assigned to each of the agents. If one takes this idea seriously, then one can identify a possible world in a Kripke model with a *possibility*: a function that assigns to each propositional variable a truth-value and to each agent a set of possibilities. To make this work formally needs some rather sophisticated mathematical tools, with which I will not be concerned here.

For example, the possible world v in the model of our example would correspond to the function f_v that assigns to each p the same value that it gets at v , and assigns to both a and b the state that is the empty set, i.e. $f_v(a) = f_v(b) = \emptyset$. The function f_w corresponding to w assigns to a the set $f_w(a) = \{f_v\}$, and to b the set $f_w(b) = \{f_u\}$. In a picture:



It holds that f_v and f_u are the *same* functions. So, again, the intersection of $f_w(a)$ and $f_w(b)$ is not empty: the distributed knowledge of a and b is not inconsistent.

As an intermediate conclusion we can say that the analysis of distributed knowledge in Kripke models as 'intersection of information states' is ontology dependent, in the sense that different views on the role of possible worlds lead to different notions of distributed knowledge. I do not think this is a very big problem, but the considerations above do show that the truth values of sentences of the form $D\phi$ depend on the way one models the information states of the agents involved. In particular, the truth values of such sentences depend on the view of 'what a possible world is.' In this respect, the semantical interpretation of D is not in the same way 'ontology-independent' as other epistemic operators, such as 'common knowledge,' are.

Adding information as logical consequence.

I will now approach the question of distributed information in a more syntactic way. If we are given two information states, we can combine them by taking the logical consequence of sentences that

are accepted in either state. With respect to distributed knowledge, this means that a sentence is distributed knowledge if and only if it is a logical consequence of the sentences that are known by the agents.

More formally, we can model this by taking all sentences of \mathcal{L} that are accepted in either $w(a)$ or in $w(b)$, and say that ϕ is distributed knowledge between a and b just in case it is a logical consequence of this set of sentences. I.e. we can extend the semantical definition with the following clause:²

Definition 1.5

$$(K, w) \models D_{\{a,b\}}\phi \text{ iff } \{\psi \in \mathcal{L} \mid (K, w) \models K_a\psi \text{ or } (K, w) \models K_b\psi\} \models \phi$$

The reason for looking at consequences of sentences of \mathcal{L} , as opposed to \mathcal{L}^D , is that in the latter case, the definition would be circular: the right hand side of the definition would quantify over all sentences believed by a or b in w , which includes the sentence $D_B\phi$. Since \mathcal{L} does not contain the operator D , then we can see it as a definition.

If we define distributed knowledge in this way, then which sentences are distributed knowledge depends on the expressive power of the language: the stronger the language, the more sentences will be distributed knowledge. For example, suppose $w(a)$ and $w(b)$ are both singleton sets consisting of w_a and w_b respectively. If w_a and w_b are indistinguishable in the language \mathcal{L} but distinguishable in a stronger language \mathcal{L}^+ , then relative to \mathcal{L} , the distributed knowledge of a and b is consistent (a and b consider the same sentences of \mathcal{L} true), but relative to \mathcal{L}^+ , their distributed knowledge is not consistent (there is a sentence of \mathcal{L}^+ of which a believes it is true, but b believes it is false).

If the notion of distributed knowledge between a and b is meant to capture the amount of information a and b can communicate, using the language \mathcal{L} , to a third agent, then the syntactic approach seems to be the right one: this third agent will get exactly the information that can be expressed by the language in which the agents are communicating. The fact that the meaning of D_B depends on the expressive power of the language is very natural in this case: if the agents communicate using the language \mathcal{L} , then the amount of information that they can communicate depends on the expressive power of the language \mathcal{L} . But this argument is only valid if we assume that the agents do communicate within \mathcal{L} : this is, in general, not the case in the model of Fagin et al. (1995).

2 A comparison of the definitions

In this section I will compare the operator D of definition 1.4 and the operator D of definition 1.5. To make the discussion a bit easier, I will write use \models_1 for the interpretation of \mathcal{L}^D as in definition 1.4, and I will use \models_2 to denote satisfaction relation defined in definition 1.5. Also, I will extend the definition to apply to all groups B in the obvious way:

$$\begin{aligned} (K, w) \models_1 D_B\phi & \text{ iff } \text{for all } v \text{ such that } w \xrightarrow{a} v \text{ for each } a \in B, \\ & \text{it holds that } (K, v) \models_1 \phi \\ (K, w) \models_2 D_B\phi & \text{ iff } \{\psi \in \mathcal{L} \mid (K, w) \models K_a\psi \text{ for some } a \in B\} \models_2 \phi \end{aligned}$$

The comparison is divided in two subsections. Under the heading ‘truth,’ I will show that the two definitions do not assign the same truth-values to sentences, and identify two different classes of Kripke models in which the differences between the two definitions collapse. Under the heading ‘logic,’ there is a proof that if the the set of propositional variables is infinite, then \models_1 and \models_2 give rise to the same logic, in the sense that the same sentences are valid under both conceptions.

Truth

If we consider the relations \models_1 and \models_2 , it is not hard to see they are not equivalent.

²Humberstone (1985) characterizes what he calls ‘collective knowledge’ like this.

The following proposition shows that the operator D of \models_1 is, in a sense, weaker than the operator D of \models_2 : any sentence of \mathcal{L} that is distributed knowledge under the second conception is also distributed knowledge under the first:

Proposition 2.1 For all ϕ of \mathcal{L} :

If $(K, w) \models_2 D_B \phi$ then $(K, w) \models_1 D_B \phi$, but not vice versa.

proof: In this proof and the ones that follow, I will write w instead of (K, w) when this is not likely to lead to confusion, and write $w(a)$ for the set $\{(K, v) \mid w \xrightarrow{a} v\}$.

Assume $w \models_2 D_B \phi$. Then $\{\psi \in \mathcal{L} \mid w \models K_a \psi \text{ for some } a \in \mathcal{B}\} \models \phi$. Take any $v \in \bigcap_{a \in \mathcal{B}} w(a)$. Clearly, for any ψ such that $w \models K_a \psi$ for some $a \in \mathcal{B}$, it holds that $v \models \psi$. So, by assumption, $v \models \phi$, and since v was arbitrary, it follows that $w \models_1 D_B \phi$.

For the negative result, the model of our example on page 113 is an example of a model in which $(K, w) \models_1 D_{\{a,b\}} \perp$, but $(K, w) \not\models_2 D_{\{a,b\}} \perp$. \square

In certain models, however, the operators are equivalent: in models that are *full* and in models that are *distinguishing*, the differences between the two operators collapse.

Proposition 2.2 (equivalence results)

1. A Kripke model K is *full* just in case for each w in K and each set of sentences Γ it holds that if $\{\psi \in \mathcal{L} \mid w \models K_a \psi \text{ for some } a \in \mathcal{B}\} \subseteq \Gamma$ and Γ is \models_2 -satisfiable (i.e. there is a model in which all sentences of Γ are \models_2 -true) then there is a $v \in \bigcap_{a \in \mathcal{B}} w(a)$ such that $v \models_2 \psi$ for all $\psi \in \Gamma$. It holds that if K is full, then:

$$(K, w) \models_1 \phi \text{ iff } (K, w) \models_2 \phi$$

2. A Kripke model K is *distinguishing* just in case for each w in K , each $v \in \bigcup_{a \in \mathcal{A}} w(a)$ and each $a \in \mathcal{A}$, there is a sentence ϕ_a of \mathcal{L} such that $v \models \phi_a$ iff $v \in w(a)$.

If K is distinguishing, then³

$$(K, w) \models_1 \phi \text{ iff } (K, w) \models_2 \phi$$

proof: For the first item, suppose that K is full. We prove the result by induction on ϕ , where the only interesting case is when ϕ is of the form $D_B \psi$.

We show that for w in K it holds that:

$$(K, w) \models_1 D_B \psi \text{ iff } (K, w) \models_2 D_B \psi.$$

Assume that $w \not\models_2 D_B \psi$. Then $\{\psi \in \mathcal{L} \mid w \models K_a \psi \text{ for some } a \in \mathcal{B}\} \not\models_2 \psi$. But then, the set $\Gamma := \{\psi \in \mathcal{L} \mid w \models K_a \psi \text{ for some } a \in \mathcal{B}\} \cup \{\neg \psi\}$ is \models_2 -satisfiable, and since w is full this means that there must be some $w_\Gamma \in \bigcap_{a \in \mathcal{B}} w(a)$ in which all sentences of Γ are \models_2 -true. But then in particular, $w_\Gamma \models_2 \psi$, so by induction hypothesis $w_\Gamma \models_1 \psi$, and therefore $w \not\models_1 D_B \psi$.

For the other direction, the reasoning is the same as in proposition 2.1: suppose $w \models_2 D_B \psi$. Then by definition, $\{\psi \in \mathcal{L} \mid w \models \Box_a \psi \text{ for some } a \in \mathcal{B}\} \models_2 \psi$. Then, *a fortiori*, for each $v \in \bigcap_{a \in \mathcal{B}} w(a)$, $v \models_2 \psi$, and therefore, by induction hypothesis, $v \models_1 \psi$.

For the second item, the proof is by induction on the on the number of occurrences of D_B -operators in ϕ , with a subinduction on the structure of ϕ . The only interesting cases in the induction is when ϕ is of the form $D_B \psi$.

So, suppose K is distinguishing, and that we have proven the result for all sentences that contain at most as many occurrences of D as ϕ does.

Suppose first that $w \models_1 D_B \psi$. Then, for each $v \in \bigcap_{a \in \mathcal{B}} w(a)$ it holds that $v \models_1 \psi$. Now take any $a \in \mathcal{B}$. Since K is distinguishing, there is a sentence ϕ_a of \mathcal{L} such that for each u such that

³van der Hoek et al. (to appear) prove a weaker result that gave me the idea for this one.

$w \xrightarrow{b} u$ for some b , it holds that $u \models \phi_a$ iff $w \xrightarrow{a} u$. So, clearly, $w \models_1 K_a \phi_a$, and, since $\phi_a \in \mathcal{L}$, also $w \models_2 K_a \phi_a$. Also, $w \models_1 K_a (\bigwedge_{b \in \mathcal{B}} \phi_b \rightarrow \phi)$. Since this sentence contains less occurrences of D than $D_B \phi$ does, it follows that $w \models_2 K_a (\bigwedge_{b \in \mathcal{B}} \phi_b \rightarrow \phi)$.

Note that it holds for all χ and $a \in \mathcal{B}$ that if $w \models_2 K_a \chi$, then $w \models_2 D_B \chi$. So, in particular, $w \models_3 D_B \phi_a$ for each $a \in \mathcal{B}$, and $w \models_3 D_B (\bigwedge_{a \in \mathcal{B}} \phi_a \rightarrow \phi)$. It also holds that if $w \models_2 D_B \chi$ and $w \models_2 D_B (\chi \rightarrow \chi')$, then $w \models_2 D_B \chi'$. Combining all this, it follows that $w \models_2 D_B \phi$, as we wanted to prove.

The other direction is goes as before. □

The condition of being 'full' is subsumed under the property of Kripke models that information states can be characterized by a set of sentences, i.e. that information states consist of *all* models of a particular set of sentences. If we assume that the beliefs of the agent can be expressed in the object language, then this is a natural consequence of the slogan that 'the beliefs of an agent are modeled by the set of possibilities compatible with his beliefs.' The condition of being distinguishing is subsumed under the property that information states consist of all models of some *finite* set of sentences.

In any case, the differences between \models_1 and \models_2 are relevant only in models in which information states cannot be characterized by a set of sentences: otherwise, the notions collapse.

Logic

We have seen above that when our set of propositional variables is finite, the logic of \models_1 is not the same as the logic of \models_2 . When we have infinitely many propositional variables, the situation is different, however. In this case the logic of both relations are the same, in the sense that any sentence that is valid under \models_2 is also valid under \models_1 , and vice versa. We can show this by using the properties of fullness and being distinguishing of the previous section.

Proposition 2.3 For each sentence ϕ of \mathcal{L}^D :

$$\text{For all } (K, w): (K, w) \models_1 \phi \text{ iff for all } (K, w): (K, w) \models_2 \phi.$$

Moreover, this equivalence holds also if we restrict the quantification to all transitive, euclidean and/or reflexive models.

proof:

[\Rightarrow] Suppose $(K, w) \not\models_2 \phi$. Then, with lemma 2.4, we can find a full model (K', w') such that $(K', w') \not\models_2 \phi$. But then, $(K', w') \not\models_1 \phi$.

[\Leftarrow]. Suppose there is a K such that $(K, w) \not\models_1 \phi$. Then can use lemma 2.5 and find a distinguishing model (K', w') such that $(K', w') \not\models_1 \phi$. But if (K', w') is distinguishing, this implies that $(K', w') \not\models_2 \phi$. □

Lemma 2.4 For each (K, w) there is a full model $(K, w)'$ such that $(K, w) \models_2 \phi$ iff $(K, w)' \models_2 \phi$. Moreover, we can find a $(K, w)'$ that is euclidean, transitive and/or reflexive just in case (K, w) is.

proof: The proof is just a simple variation on the canonical model construction of the standard completeness proof (cf. proposition 3.1. Define K' as follows. For its domain, K' has all maximal \models_2 -satisfiable sets, i.e. all sets of sentences Σ for which there is a model in which all sentences of Σ are true, and which is maximal in the sense that adding any sentence to Σ results in a set that is not satisfiable. The accessibility relations of K' are given by: $\Sigma \xrightarrow{a} \Gamma$ iff for all ψ , if $K_a \psi \in \Sigma$, then $\psi \in \Gamma$, and the valuation function assigns to each Σ and p the value 1 just in case $p \in \Sigma$.

For a euclidean, transitive and/or reflexive model, we construct K' with sets that are satisfiable in the appropriate models.

It holds that:

$$(K', \Sigma) \models_2 \phi \text{ iff } \phi \in \Sigma$$

The proof works by first showing the result for all sentences of \mathcal{L} by induction on ϕ , and then showing that the result holds for all sentences of \mathcal{L}^D , again by induction on ϕ . The details of the proof are similar to those in the standard Henkin proof of the completeness of classical modal logic. The case where ϕ is of the form $D_B\psi$ runs as follows:

$$\begin{aligned} (K', \Sigma) \models_2 D_B\psi &\text{ iff} \\ \{\chi \in \mathcal{L} \mid (K', \Sigma) \models K_a\chi \text{ for some } a \in \mathcal{B}\} \models_2 \psi &\text{ iff (induction hypothesis)} \\ \{\chi \in \mathcal{L} \mid K_a\chi \in \Sigma \text{ for some } a \in \mathcal{B}\} \models_2 \psi &\text{ iff (since } \Sigma \text{ is maximal)} \\ D_B\psi \in \Sigma & \end{aligned}$$

Now, clearly, the set of sentences that are true in (K, w) is satisfiable. Let Σ be this set. Then (K', Σ) is a full model such that $(K, w) \models_2 \phi$ iff $(K', \Sigma) \models_2 \phi$ for each $\phi \in \mathcal{L}^D$. \square

Lemma 2.5 For each (K, w) such that $(K, w) \models_1 \phi$, there is a distinguishing model (K', w') such that $(K', w') \models_1 \phi$.

Moreover, if K is transitive and euclidean (and reflexive), then we can choose K' to be transitive and euclidean (and reflexive) as well.

proof: Suppose $(K, w) \models_1 \phi$. We know from the completeness proof that we can find a countable model $(K, w)'$ such that $(K, w)' \models_1 \phi$. Let, for each v in the domain of K' , $p_{a,w}$ be a propositional variable that does not occur in ϕ . Here we use the fact that the set of propositional variables is infinite.

Now let K'' be exactly as K , except that $V(v)(p_{w,a}) = 1$ iff $w \xrightarrow{a} v$.

Clearly, K'' is distinguishing: it holds for each v and w in the domain of K that $v \in w(a)$ iff $v \models p_{a,w}$. Of course, if K' is transitive, euclidean and/or reflexive, then so is K'' .

Since ϕ does not contain any of the propositional variables $p_{a,w}$, it follows by a standard argument that $(K', w) \models \phi$ iff $(K'', w) \models \phi$. So, (K'', w) is the model we are looking for. \square

So, we have proven that \models_1 and \models_2 give rise to the same notion of validity. I state here, without proof, that this result can be extended: the set of sentences of \mathcal{L}^D true in all Hintikka models, or in all possibilities, is the same set as the set of \models_1 (or \models_2) valid sentences.

3 Completeness.

We have seen that, if our language has infinitely propositional variables, the two notions of validity \models_1 and \models_2 are the same. In the following, we will write $\models \phi$ iff $(K, w) \models_1 \phi$ for each (K, w) , and $\models_{DK45} \phi$ and $\models_{D55} \phi$ for their transitive and euclidean, and transitive, euclidean and reflexive counterparts.

The minimal modal logic K has the following axioms and rules:

K1 $\vdash \phi$ whenever ϕ is a truth-functional tautology.

K2 $\vdash (K_a\phi \wedge K_a(\phi \rightarrow \psi)) \rightarrow K_a\psi$.

MP From $\vdash \phi$ and $\vdash \phi \rightarrow \psi$, conclude that $\vdash \psi$.

Nec From $\vdash \phi$ conclude that $\vdash K_a\phi$

Adding the following two axioms to the logic of K provides a sound and complete axiomatization of all valid sentences of \mathcal{L}^D .

D1 $D_{\{a\}}\phi \leftrightarrow K_a\phi$.

D2 $(D_B(\phi \rightarrow \psi) \wedge D_C\phi) \rightarrow D_D\psi$ if $B \subseteq D$ and $C \subseteq D$.

We let DK consist of the axioms of K together with $D1$ and $D2$.

Axiom $D1$ says that the knowledge that is distributed in the ‘group’ consisting of a only is just the knowledge of a . Axiom $D2$ says that if a certain group has distributed knowledge of ϕ , and another group has distributed knowledge that ϕ implies ψ , then both groups together have distributed knowledge of ψ as well.

To consider one implication of these axioms: if $\mathcal{B} \subseteq \mathcal{C}$, then $\vdash D_{\mathcal{B}}\psi \rightarrow D_{\mathcal{C}}\psi$ for all ψ .⁴ This validity corresponds with the intuition that if a sentence is distributed knowledge in a group \mathcal{B} , it will also be distributed knowledge in any group larger than \mathcal{B} . One can see this as a generalization of the maxim that ‘two know more than one.’

Also, these axioms imply that $D_{\mathcal{B}}$ is a normal modal operator, in the sense that a necessitation rule for $D_{\mathcal{B}}$ is a derived rule.⁵ and that D distributes over implication.⁶

The logic DK45 is given by adding the following two axioms to DK:

D4 $D_{\mathcal{B}}\phi \rightarrow D_{\mathcal{B}}D_{\mathcal{B}}\phi$.

D5 $\neg D_{\mathcal{B}}\phi \rightarrow D_{\mathcal{B}}\neg D_{\mathcal{B}}\phi$.

While the logic DS5 is given by adding the following axiom to DK45:

DT $D_{\mathcal{B}}\phi \rightarrow \phi$.

Proposition 3.1 (Completeness)

DK is a sound and complete axiomatization of \models .

DK45 is a sound and complete axiomatization of \models_{DK45}

DS5 is a sound and complete axiomatization of \models_{DS5} .

proof: I’ll present a sketch of a completeness proof that is based on the completeness proof Fagin et al. (1992) give for DS5. Let a ‘pseudomodel’ be a model of the form $(K, (\xrightarrow{\mathcal{B}})_{\mathcal{B} \subseteq \mathcal{A}}, V)$, and define ‘pseudo-satisfiability’ as a relation between sentences of \mathcal{L}^D and pseudo-models by treating the operators $D_{\mathcal{B}}$ as quantifying over $\xrightarrow{\mathcal{B}}$ -accessible worlds. With respect to pseudo-satisfiability, the operators $D_{\mathcal{B}}$ are just classical modal operators. We can just use the standard techniques to show that any DK (or DK45 or S5) consistent theory can be pseudo-satisfied in a pseudo-model (which is transitive, euclidean and reflexive, if necessary). It is also not very hard to check that the rules are sound for pseudo-models that have the property that if $\mathcal{B} \subseteq \mathcal{C}$, then $\xrightarrow{\mathcal{B}} \subseteq \xrightarrow{\mathcal{C}}$. The canonical model has this property. That means that the logic is sound and complete with respect to pseudo-models the property that if $\mathcal{B} \subseteq \mathcal{C}$, then $\xrightarrow{\mathcal{B}} \subseteq \xrightarrow{\mathcal{C}}$. Since $D_{\mathcal{B}}$ is just another classical modal operator in these models, it follows that we can construct finite canonical models, and therefore that the logic is decidable.

We can also use standard techniques to show that any pseudo-model can be unraveled into a model that looks like a tree, and in which the same sentences are true.

Now turn the unraveled pseudo-model into a model for \mathcal{L}^D by taking the same worlds and valuation function, and setting $w \xrightarrow{a} v$ in the new model just in case there is a \mathcal{B} such that $a \in \mathcal{B}$ and $w \xrightarrow{\mathcal{B}} v$ in the pseudo-model. We can now show that any sentence that is pseudo-satisfied at some world w in the restored unraveled pseudo-model is satisfied at w in the new model.

We have now proven that DK is complete with respect to all models. Since we need only an unraveled pseudo-model of finite ‘depth’ to satisfy a given sentence (the length of the longest path in the tree need not be greater than the maximal nesting of modal operators in the sentence), DK has the finite model property.

The new model does not have the properties associated with belief or knowledge. To show that every DK45-consistent set is satisfiable in a belief model, we can simply take the transitive and

⁴If $\mathcal{B} \subseteq \mathcal{C}$, and $a \in \mathcal{B}$, then $\vdash (D_{\{a\}}(\psi \rightarrow \psi) \wedge D_{\mathcal{B}}\psi) \rightarrow D_{\mathcal{C}}\psi$ is a special case of $D2$. With axiom $D1$, it follows that $D_{\{a\}}(\psi \rightarrow \psi)$, so, by propositional logic, we can conclude that $\vdash D_{\mathcal{B}}\psi \rightarrow D_{\mathcal{C}}\psi$

⁵Assume that $a \in \mathcal{B}$. Then $\vdash \psi$ implies that $\vdash K_a\psi$, which implies by $D1$ that $\vdash D_{\{a\}}\psi$, which implies by $D2$ that $\vdash D_{\mathcal{B}}\psi$.

⁶This is a special case of $D2$, where the sets \mathcal{B} , \mathcal{C} and \mathcal{D} are all the same.

euclidean closure of the accessibility relations in the tree model, and for DS5, we take the reflexive transitive and euclidean closure. Lemma 3.2 guarantees that this can be done: the new model is safe. \square

Lemma 3.2 Let $\mathcal{K} = (K, (\xrightarrow{a})_{a \in \mathcal{B}})$ be a model such that for each $w \in K$, the set of sentences that are true at w is DK45-consistent. We define the relation $\overset{\mathcal{B}}{\rightsquigarrow}$, for each $\mathcal{B} \subseteq \mathcal{A}$ relative to \mathcal{K} , as follows:

$s \overset{\mathcal{B}}{\rightsquigarrow} t$ iff there are $s_0 \dots s_n$ and $s'_1 \dots s'_m$ such that for each $a \in \mathcal{B}$: $s_0 \xrightarrow{a} s_1 \xrightarrow{a} \dots \xrightarrow{a} s_n$, $s_0 \xrightarrow{a} s'_1 \xrightarrow{a} \dots \xrightarrow{a} s'_m$ (with $n \geq 0$ and $m \geq 1$) and $s_n = s$ and $s'_m = t$. Note that $\overset{\mathcal{A}}{\rightsquigarrow}$ is the transitive and euclidean closure of \xrightarrow{a} .

We say that \mathcal{K} is *safe* when $s \overset{\mathcal{B}}{\rightsquigarrow} t$ iff for each $a \in \mathcal{B}$, $s \overset{\mathcal{A}}{\rightsquigarrow} t$. Models that look like a tree are safe, for example, and also models that are transitive and euclidean.

It now holds that:

$$(K, (\overset{\mathcal{A}}{\rightsquigarrow})_{a \in \mathcal{B}}, V), w \models \phi \text{ iff } (K, (\xrightarrow{a})_{a \in \mathcal{B}}, V), w \models \phi.$$

proof: By induction on ϕ . The interesting case is when ϕ is of the form $K_a\psi$, where we use the fact that the theory of w is DK45-consistent. Once we have this, the case where ϕ is of the form $D_B\psi$ is straightforward with the assumption that \mathcal{K} is safe. \square

I only have a partial answer to the question whether each consistent sentence is satisfiable in a finite model. Note that the question loses some of its urgency by the fact that we have already shown that the logics are decidable.

Proposition 3.3 (finite model property)

Every DK-consistent sentence is true in some finite model.

Suppose ϕ contains is such that if an occurrence of D_B is in the scope of an occurrence of D_D , then either $\mathcal{B} \subseteq \mathcal{C}$, $\mathcal{C} \supseteq \mathcal{B}$, or $\mathcal{B} \cap \mathcal{C} = \emptyset$. Then, if ϕ is DK45-consistent, there is a finite introspective model in which ϕ is true, and if ϕ is DS5-consistent, there is a finite reflexive and introspective model in which ϕ is true.

proof: I suggested in the proof of proposition 3.1 how each DK-consistent set is satisfied in a finite model; this is also proven in Gargov and Passy (1990). One can also prove this more directly. My proof of the second statement is rather complicated: it would take me several pages to go through all the details. I have chosen to simply omit it. \square

The literature contains a whole range of proofs for modal logics ‘with intersection.’ The following is an attempt at an overview. There is a very elegant completeness proof in Gargov and Passy (1990) for a logic they call ‘Boolean modal logic,’ of which the language \mathcal{L}^D is only a small fragment. They also show that DK has the finite model property. Their proof works just as well for our case. Passy and Tinchev (1991) also study a richer language than the one considered here. They give a completeness proof that may also work for DK. Since the two last-mentioned articles get their inspiration from propositional dynamic logic as opposed to epistemic logic, these proofs apply only to the logic DK; the authors are not concerned with proving completeness with respect to transitive or euclidean models that are typical of epistemic semantics. And it is exactly this that makes the completeness proofs of DK45 and DS5 so difficult. Fagin et al. (1992) give a proof of the completeness of DS5; Fagin et al. (1995) claim that the logic is complete also for the language with all D_B -operators for each $\mathcal{B} \subseteq \mathcal{A}$, but do not give a proof. van der Hoek and Meyer (1992) and van der Hoek and Meyer (1997) contain a completeness proofs for DK45 and DK as well, but only for a language with a single operator D_A , with \mathcal{A} the set of *all* agents. Since their proofs are rather long and opaque, it is not immediately obvious how their techniques can be used to work for the full language \mathcal{L}^D . Finally, Yde Venema (personal communication) has a proof of the completeness of DK using the ‘step-by-step’ method, that can be extended to cover the cases of DK45 and DS5 as well.

4 Conclusions

In this article, I have compared two different schemes for defining the semantics of an operator that expresses distributed knowledge in a Kripke model. The first definition, which defines operator as quantifying over the intersection of the accessibility relations of the agents involved has been shown to be ontology dependent. The second definition, where distributed knowledge is defined in terms of logical consequence, is shown to be dependent on the expressive power of the language. Moreover, the different definitions are not interchangeable *salva veritate*. However, under certain aspects the differences between the definitions can be disregarded: in particular, when information states always consist of all models of a certain set of sentences. It is shown that the different definitions have the same weakly sound and complete axiomatization. This shows that the issue of what a good definitions of distributed knowledge is cannot be decided on the basis of logic alone (i.e. we cannot choose the one definition over and above another one on the basis of our intuitions about which sentences logically follow from others, because in this respect, the definitions are equivalent), but must be decided at the semantical level (which sentences are true in which models).

References

- Barwise, J. (1989). On the model theory of common knowledge. In *The Situation in Logic*, number 17 in CSLI Lecture notes, pages 201–220. CSLI Publications, Stanford.
- Barwise, J. and Moss, L. S. (1996). *Vicious Circles*. CSLI Publications, Stanford.
- Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge*. The MIT Press, Cambridge (Mass.).
- Fagin, R., Halpern, J. Y., and Vardi, M. Y. (1992). What can machines know? on the properties of knowledge in distributed systems. *Journal of the Association for Computing Machinery*, 39(2):328–376.
- Gargov, G. and Passy, S. (1990). A note on Boolean modal logic. In Petkov, P. P., editor, *Mathematical Logic*, Proceedings of the Summer School and Conference on Mathematical Logic, honourably dedicated to the ninetieth anniversary of Arend Heyting (1989–1980) held September 13–23, 1988, in Chaika (near Varna), Bulgaria, pages 299–309. Plenum Press.
- Gerbrandy, J. (1997). Dynamic epistemic logic. *ILLC preprint LP-1997-04*. To appear in the proceedings of the Second Conference on Information-Theoretic Approaches to Logic, Language, and Computation.
- Gerbrandy, J. and Groeneveld, W. (1997). Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169. Also available as an ILLC Report LP-1996-10.
- Halpern, J. Y. (1987). Using reasoning about knowledge to analyze distributed systems. In Traub, J., Grosz, B., Lampson, B., and Nilsson, N., editors, *Annual review of computer science*, Vol. 2, pages 37–68. Annual Reviews Inc., Palo Alto, California.
- Halpern, J. Y. and Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the Association for Computing Machinery*, 37(3):549–587.
- Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press.
- Humberstone, I. L. (1985). The formalities of collective omniscience. *Philosophical Studies*, 48:401–423.
- Kripke, S. A. (1963). A semantical analysis of modal logic I, normal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 9:63–96.

- Passy, S. and Tinchev, T. (1991). An essay in combinatory modal logic. *Information and Computation*, 93:263–332.
- van der Hoek, W. and Meyer, J.-J. C. (1992). Making some issues of implicit knowledge explicit. *International Journal of Foundations of Computer Science*, 3(2):193–223.
- van der Hoek, W. and Meyer, J.-J. C. (1997). A complete epistemic logic for multiple agents: combining distributed and common knowledge. In Bacharach, M. O. L., Gerbrand-Varet, Z. A., Mongin, P., and Shin, H. S., editors, *Epistemic Logic and the theory of Games and Decisions*, pages 35–68.
- van der Hoek, W., van Linder, B., and Meyer, J.-J. (to appear). Group knowledge isn't always distributed. *To appear in: Mathematics for the Social Sciences*.

Questions in Update Semantics*

Jeroen Groenendijk
ILLC/Department of Philosophy
Faculty of Humanities
University of Amsterdam
groenend@philo.uva.nl

Abstract

The paper presents a static and dynamic version of an update semantics for a language of predicate logic enriched with simplex interrogatives. The static version gives rise to the same logic as the denotational semantics given for the same language in G&S (1997). The dynamic semantics has some additional features, which concern anaphoric relations, and present us with a notion of existential quantification which seems to correspond to the specific use of indefinites. At meta-logical level an absolute notion of discourse coherence is defined which covers the usual notion of answerhood as a special case. More fine-grained notions of relevance and coherence are given as well, and are applied to simple interrogative dialogues.

1 Introduction

In this paper I combine in a conservative way the denotational semantics for a language of predicate logic enriched with simplex interrogatives given in sections 4.4 and 4.5 of G&S (1997) (based on G&S (1984)), with an update semantics for predicate logic in the line of GSV (1996). I give both a static and a dynamic version of the update interpretation, which differ only in the interpretation of existential quantification. The combination is conservative in that under the static update interpretation, the resulting logic coincides with the logic the denotational semantics gives rise to.

Some of the basic ideas of the work presented here have much in common with Jäger (1995) and Hulstijn (1997), who also combine the partition semantics for questions of G&S (1984) with update semantics. The present short paper left no room for a comparison, I hope to make up for that in a longer version.

The paper is organized as follows. In Section 2, I discuss the syntax of the logical language, and the simple type of dialogues that we will study here. In Section 3, I say what the ordinary states of GSV (1996) are. In Section 4, I turn these into structured states in which questions can be modeled. In Section 5, I give both a classical and a dynamic update interpretation for the language. In Section 6, I compare the notions of entailment for the different interpretations. In Section 7, I discuss answerhood as a special case of coherence of interrogative dialogues, in section 8, I make coherence more finegrained by linking it with a comparative notion of relevance, and I apply it to some simple examples in Section 9. In Section 10, I say a few words about

*Preferably, this paper should be read as an addendum to G&S (1997), and with some familiarity with GSV (1996). The paper has a rather long history of oral presentations, and in the mean time some things I said already entered in the work of others. The material got more or less its present shape during my stay in Jerusalem in December 1997 and January 1998, as a visiting scholar at the Institute for Advanced Studies at the Hebrew University, taking part in the Research Group on Semantics, organized by Edit Doron and Fred Landman. Besides to them, and many others who commented on earlier presentations, I owe special thanks to Paul Dekker, Martin Jansche and Craige Roberts, with whom I had extensive e-mail exchanges about the material. And, of course, I should mention that much in here is intimately related to the previous work on questions and on dynamics I did together with Martin Stokhof. The research reported here, is part of the project *Sources and Streams of Information*, supported by NWO, the Dutch Organization for Scientific Research.

an alternative interpretation of existential quantification that the semantics gives rise to, where quantification not only introduces a variable, but also the issue what object its value is.

2 Syntax

I consider only simplex interrogatives, added to a language of predicate logic:

Definition 1 (Syntax) Let L_1 be a language of predicate logic.

The language L_2 is the smallest set such that:

If $\phi \in L_1$, then $\phi \in L_2$;

If $\phi \in L_1$, \vec{x} a sequence of n variables ($0 \leq n$), then $?\vec{x}\phi \in L_2$.

I use $\phi?$ as a meta-variable ranging over all interrogatives. Unless indicated otherwise, ϕ , ranges over all sentences in L_2 , and τ over sequences thereof.

The language does not contain interrogatives embedded under other operators, nor can the interrogative operator be iterated. Yes/no-interrogatives and single and multiple constituent interrogatives are constructed in one go by a single operation. Querying always ranges over the domain as a whole, 'restricted querying', as in *Which A are B*, cannot be represented in the language.

I only consider mention-all interpretations of constituent interrogatives, they ask for the specification of the denotation of an n -place relation ($0 \leq n$).

Update semantics lends itself naturally to the interpretation of discourse. We consider a specific type of dialogue:

Definition 2 (Interrogation) An *interrogation* in L_2 is a sequence of sentences τ in L_2 . All interrogatives in τ are uttered by single agent, and all indicatives in τ are uttered by one other agent.

Given the strict casting, we do not have to indicate who says what in representing an interrogation. Some examples and their English glosses:

- (1) a. $?\exists xPx; \neg\exists xPx$
 'Does someone have the property P ?'; 'No.'
- b. $?xPx; Pa \wedge Pb \wedge \forall x(Px \rightarrow (x = a \vee x = b))$
 'Who have the property P ?'; 'a and b.'
- c. $?x(x = a); Pa \wedge \forall x(Px \rightarrow (x = a))$
 'Who is a?'; 'The one with the property P .'
- d. $?xyRxy; Rab \wedge Rac \wedge \forall x\forall y((Rxy \rightarrow ((x = a \wedge y = b) \vee (x = a \wedge y = c)))$
 'Who stands in the relation R to whom?'; 'a to b and c'

The logical language deals only with full sentential answers. Intended exhaustiveness of an answer has to be made explicit in the logical formulation.

Under the dynamic interpretation of the language, we will be able to account for certain anaphoric relations across sentences in an interrogation:

- (2) a. $\exists xPx; ?Qx; \neg Qx$
 'Someone has the property P '; 'Does he have Q , too?'; 'No.'
- b. $\exists xPx; ?y(y = x); x = a$
 'Someone has the property P '; 'Who is it?'; 'It is a'.

However, the anaphoric relations indicated in the following examples will not be accounted for:

- (3) a. $?\exists xPx; ?Qx$
 ‘Does someone have the property P ?’; ‘Does he have Q , too?’
 b. $?xPx; ?yRyx$
 ‘Who has the property P ?’; ‘Who stands in the relation R to him?’

I take interrogatives, like negation, to be scope islands, binding blocks. Quantifiers inside the scope of an interrogative, and the interrogative operator itself, will not be enabled to bind variables outside their scope. For lack of space, I cannot go into this, but I believe that unlike the examples in (2), those in (3) exemplify the kind of anaphoric relations you find in modal subordination. I will not deal with subordination here, but I will set up the semantics in such a way that it is clear that when combined with an analysis of subordination, these anaphoric relations could in principle be handled.

3 Information

In update semantics, we define the interpretation of a language in terms of update functions, (partial) functions from information states to information states. If we want to model questions in update semantics, the basic issue is: what is the information change potential of a question? The answer we give to this is that a question *structures* information.

For the update interpretation of L_2 we use a notion of structured information states, which we construct from ordinary information states. Here, I give only a rather informal characterization of the latter, and refer to GSV (1996) for explicit definitions.¹

An ordinary information state is a set of possibilities, where a possibility consists of a possible world, and an assignment function, defined only for the variables which are in active use. Which variables are in active use is fixed for a state as such, the assignments in its possibilities are all defined for the same variables.

Let α be a non-logical constant of the language or a variable, i a possibility. We write $i(\alpha)$ for the denotation of α in i , where $i(\alpha)$ is determined by the world in i if α is a constant, and by the assignment in i if α is a variable which is in the domain of the assignment. If the latter is not the case, $i(\alpha)$ is undefined. Let $\vec{\alpha}$ be a sequence of n terms, then by $i(\vec{\alpha})$ we mean $\langle i(\alpha_1), \dots, i(\alpha_n) \rangle$, where if for some $k \leq n$: $i(\alpha_k)$ is undefined, then $i(\vec{\alpha})$ is undefined.

The set of possibilities is partially ordered by an extension relation: i' is an extension of i , $i \leq i'$ iff i and i' contain the same world, and if a variable x is in i then x is in i' and $i'(x) = i(x)$. By $i[x/e]$ we mean the possibility which is like i , except that the variable x has been added and is assigned the value e . Let \vec{x} be a sequence of n variables, and \vec{e} an n -tuple of objects, then by $i[\vec{x}/\vec{e}]$ we mean $i[x_1/e_1] \dots i[x_n/e_n]$. Typically: $i < i[x/e]$.²

In terms of $i[x/e]$ I define a notion of random assignment of objects e in the domain D to a variable x in a state s :³

Definition 3 (Assignment) $s[x] = \{i[x/e] \mid i \in s \ \& \ e \in D\}$

Let \vec{x} be a sequence of n variables, then by $s[\vec{x}]$ we mean $s[x_1] \dots [x_n]$.

Performing an assignment operation $s[x]$ leads to a real extension of s . The relation of extension between ordinary states is defined as follows:

Definition 4 (Extension)

s' is an extension of s , $s \leq s'$ iff $\forall i' \in s': \exists i \in s: i \leq i'$

Although $s < s[x]$, s does subsist in $s[x]$. Subsistence is defined as follows:

¹In my informal presentation, I try not to mention ‘referent systems’, but I freely allow to recycle a quantifier, and assume the presence of a referent system in the states.

²This essentially hinges on the secret use of referent systems.

³Random assignments are argued to be inappropriate for a language with epistemic modalities in GSV (1996). But see Aloni (1997) for an interesting discussion and an alternative.

Definition 5 (Subsistence) Let $s \leq s'$, and $i \in s$

i subsists in s' , $i \prec s'$ iff $\exists i' \in s': i \leq i'$

s subsists in s' , $s \prec s'$ iff $\forall i \in s: i \prec i'$

Subsistence of s in s' means that no possibility in s has been eliminated, s' may be a real extension of s , but the additional information in s' may only concern the addition of assignments of values to new variables.

4 Structured Information

Questions structure information states, a question raises an *issue*. I model an issue in a state as an equivalence relation on the possibilities in that state. If two possibilities i and j are related, then the state is indifferent as to whether things are more like i than like j . Raising an issue amounts to disconnecting possibilities in a state. And resolving an issue amounts to eliminating for some i , all possibilities j that i is related to, which includes i itself.

Since an equivalence relation on a set corresponds to a partition of that set, we can also look upon a structured state as a set of alternatives, a set of non-empty non-overlapping subsets of the set possibilities in a state, the union of which equals the set of all possibilities in the state. Resolving an issue amounts to not just eliminating some possibilities, but eliminating all possibilities in one or more alternatives, thereby eliminating these alternatives as such.

Although partitions give a nicer picture, I use equivalence relations in the definitions. And rather than taking a structured state as a pair consisting of an ordinary state s and an equivalence relation σ on it, I identify structured states with the relation σ as such. Nothing is lost, we can extract s from σ .⁴

Definition 6 (Structured States)

Let I be the set of possibilities, S be the set of ordinary states.

σ is a *structured state* iff for some $s \in S: \sigma$ is an equivalence relation on s .

The *data* in σ , $\sqrt{\sigma} = \{i \in I \mid \langle i, i \rangle \in \sigma\}$.

All issues are resolved in σ if $\sigma = \sqrt{\sigma}^2$. Such a state is called a *state of indifference*. Consider the following operation on structured states:

Definition 7 (Specific Assignment)

$\sigma[x] = \{\langle \iota_1[x/e], \iota_2[x/e] \rangle \mid \iota \in \sigma \ \& \ e \in D\}$

If the domain consists of more than one object, and σ is not the absurd state, then $\sigma[x]$ is not a state of indifference. If $d, e \in D$, $d \neq e$, and $\langle i, j \rangle \in \sigma$, then both $i[x/d]$ and $j[x/d]$ are in the data of $\sigma[x]$, but $\langle i[x/d], j[x/e] \rangle \notin \sigma[x]$. The effect of specific assignment is not only that a variable is (re-) introduced, but it also raises the issue what the value of the variable is. (Except for a short excursion on specificity in Section 10, I am not going to employ specific assignments in the main body of the paper.)

Whereas ordinary states can only be extended in two ways: by eliminating possibilities, and by (re-)introducing a variable; for structured states there is a third way of extension: by raising an issue. These three ways of extension are covered by the following definition:

Definition 8 (Structured Extension)

$\iota \leq \iota'$ iff $\iota_1 \leq \iota'_1 \ \& \ \iota_2 \leq \iota'_2$.

$\sigma \leq \sigma'$ iff $\forall \iota \in \sigma: \exists \iota' \in \sigma': \iota \leq \iota'$.

Whereas in the case of ordinary states, subsistence of a state s in a state s' means that no possibilities are eliminated, and at most new variables have been added, subsistence in the case

⁴In the end, we have to keep data and issues apart, and let issues structure (a larger part of) the whole logical space. Someone who has to answer a question needs (at least) this richer type of information to be able to figure out what an appropriate answer should look like.

of structured states means also that, although an issue may have been raised, it is not a material issue, but only concerns the question what value a variable may have.

Definition 9 (Structured Subsistence) Let $\sigma \leq \sigma'$, $\iota \in \sigma$.

ι subsists in σ' , $\iota < \sigma'$ iff $\exists \iota' \in \sigma' : \iota \leq \iota'$

σ subsists in σ' , $\sigma < \sigma'$ iff $\forall \iota \in \sigma : \iota < \sigma'$

Although $\sigma[x]$ is a real extension of σ , σ does subsist in $\sigma[x]$. The issue that a specific assignment raises, is not a material issue.

5 Update Semantics

In stating an update semantics, I follow the way in which we constructed the language (but see Section 10). First I define an update interpretation for the indicative part L_1 of the language, relative to ordinary states. In terms of that, I define an update interpretation for the interrogative language L_2 relative to structured states.

Except that I use random assignment in the interpretation of existential quantification, the update interpretation of L_1 runs as in GSV (1996):

Definition 10 (Dynamic Update Semantics L_1) Let s be an ordinary state, $\phi \in L_1$. The update of s with ϕ , $s[\phi]$, is recursively defined as follows:

$$\begin{aligned} s[Rt] &= \{i \in s \mid i(t) \in i(R)\} \\ s[t_1 = t_2] &= \{i \in s \mid i(t_1) = i(t_2)\} \\ s[\neg\phi] &= \{i \in s \mid i \not\prec s[\phi]\} \\ s[\phi \wedge \psi] &= s[\phi][\psi] \\ s[\exists x\phi] &= s[x][\phi] \end{aligned}$$

The update with an atomic clause is undefined, in case one of the terms is a variable which is not present in the state. This percolates up into the other clauses in an obvious way, which I do not bother to state explicitly. The semantics has the update property: if $s[\phi]$ exists, then $s \leq s[\phi]$.

Given the way the existential quantifier is defined, we get the familiar effects of dynamic predicate logic: existential quantifiers can bind to their right beyond their syntactic scope, but if the quantifier occurs inside a binding block, its binding capacity does not go beyond it. Negation is a binding block: a quantifier inside its scope cannot bind variables outside of it. The syntactic notion of a binding block corresponds to the semantic notion of an eliminative update. ϕ constitutes an eliminative update iff $\forall s$: if $s[\phi]$ exists, then $s[\phi] \subseteq s$.

For reasons of comparison, I also supply a static update interpretation $[\phi]^c$ for L_1 . Only the clause for existential quantification reads differently:

Definition 11 (Classical Update Semantics)

$$s[\exists x\phi]^c = \{i \in s \mid i < s[x][\phi]^c\}$$

Unlike $s[\exists x\phi]$, $s[\exists x\phi]^c$ is an eliminative update. This means that under the classical interpretation every formula is a binding block, binding is always restricted to variables in the syntactic scope of a quantifier.

Before I define the update semantics for L_2 , I first define relative to L_1 the auxiliary notion of an abstract. An abstract is a function from states to sets of sequences of objects:

Definition 12 (Abstracts) Let \vec{x} be a sequence of n variables ($0 \leq n$), $\phi \in L_1$.

$\lambda\vec{x}\phi$ an n -place abstract in L_1 . $s[\lambda\vec{x}\phi]$ is the function f from s into D^n s.th.:

$$\forall i \in s : f(i) = \{\vec{e} \in D^n \mid i[\vec{x}/\vec{e}] < s[\vec{x}][\phi]\}$$

If we replace $[\phi]$ by $[\phi]^c$, we obtain a classical interpretation $s[\lambda\vec{x}\phi]^c$ for abstracts.

Consider a simple example. In a state s , λxPx denotes the function which for each $i \in s$ gives the objects $e \in D$ such that $e \in i(P)$. Note, $s[x][Px] = s[\exists xPx]$. $s[\lambda xPx]$ gives for each possibility $i \in s$ the objects e such that $i[x/e] \in s[\exists xPx]$. (This holds dynamically, not classically: $s[x][\phi]^c \neq s[\exists x\phi]^c$.)

In case $n = 0$, we get $\lambda\phi$, where for each $i \in s$, $s[\lambda\phi](i)$ will either give the empty set, in case $i \not\prec s[\phi]$, or the set with the empty sequence as only element, in case $i \prec s[\phi]$. So, if $s[\lambda\exists xPx](i) = \emptyset$, this means that there is no object with the property P in i , otherwise, it means there is at least one such object in i .

In general, for $i \in s$, $s[\lambda\vec{x}\phi](i)$ is the denotation of an n -place relation in i . We can compare for $i, j \in s$ whether or not the denotation of the relation in question is the same in i and j . We use this in the interpretation of interrogatives.

I define the update interpretation of L_2 relative to structured states by lifting the interpretation of L_1 relative to ordinary states to structured states, adding a clause for interrogatives. (But see Section 10.) Replacing $[]$ by $[]^c$, gives a classical version.

Definition 13 (Dynamic Update Semantics L_2) Let $\phi \in L_1$.

$$\begin{aligned} \sigma[\phi] &= \{\iota' \in (\sqrt{\sigma}[\phi])^2 \mid \exists \iota \in \sigma: \iota \leq \iota'\} \\ \sigma[?\vec{x}\phi] &= \{\iota \in \sigma \mid \sqrt{\sigma}[\lambda\vec{x}\phi](\iota_1) = \sqrt{\sigma}[\lambda\vec{x}\phi](\iota_2)\} \end{aligned}$$

The first clause gives the 'lift' of the interpretation of L_1 from ordinary to structured states. Since for a structured state σ , $\sqrt{\sigma}$ is an ordinary state, $\sqrt{\sigma}[\phi]$ brings us back to the interpretation we gave for L_1 , and outputs an ordinary state. Then we consider all pairs of possibilities $\langle i, j \rangle$ for $i, j \in \sqrt{\sigma}[\phi]$, and take those pairs that are an extension of some pair in the original σ to be in $\sigma[\phi]$.

The second clause gives the interpretation for the interrogatives $?\vec{x}\phi$ of L_2 . It eliminates, i.e., disconnects, those pairs $\langle i, j \rangle \in \sigma$, such that the denotation of the relation $\lambda\vec{x}\phi$ in i is different from its denotation in j . Note that $\sigma[\phi?]$ is an eliminative update: if $\sigma[\phi]$ exists, $\sigma[\phi?] \subseteq \sigma$. Interrogatives are binding blocks. Note also that $\sigma[\phi?]$ has no effect on the data: $\sqrt{\sigma} = \sqrt{\sigma[\phi?]}$.

As in the denotational semantics, possibilities are grouped together where the denotation of a particular relation is the same. Such a group forms an alternative. But note that sets of possibilities are richer structures than just the sets of possible worlds of the denotational semantics. We can also group together sets of possibilities where a particular variable has the same value. An interrogative like $?y(y = x)$ does just that.

I add a rule for the interpretation of sequences of sentences in a language L , which follows the pattern of conjunction.

Definition 14 Let $\tau = \phi_1; \dots; \phi_n$ be a discourse in L , $[]$ an update interpretation for L . $[\tau] = [\phi_1] \circ \dots \circ [\phi_n]$, where \circ is function composition.

This definition tells us how to update a state with an interrogation. Who's state? The state of the interrogator. Simple versions of update semantics, as the one at hand, only model the point of view of a hearer. It is the hearer who updates. And if we have to choose, it is the interrogator rather than the person heard, who is updated.⁵

6 Entailment

I state the definition of entailment for the dynamic interpretation of L_2 . But by choosing the right type of states and interpretation function, it applies equally well in the other cases. Entailment is defined in terms of subsistence:⁶

⁵A third party overhearing the dialogue is another option.

⁶We can also take a discourse as conclusion.

Definition 15 (Entailment)

τ entails ϕ , $\tau \models \phi$ iff $\forall \sigma$ such that $\sigma[\tau][\phi]$ exists: $\sigma[\tau] \prec \sigma[\tau][\phi]$

The entailment relation applies uniformly, not only with respect to our two languages and our two ways of interpretation, but also with respect to the two syntactic categories, indicatives and interrogatives.

In particular the last item in the following fact, indicates that the semantics for interrogatives presented here, is indeed a conservative combination of update semantics for predicate logic in the line of GSV (1996), and the denotational semantics of questions of G&S (1997):

Fact 1 (Comparing entailment)

1. Under the classical update interpretation for L_1 , $\tau \models \phi$ corresponds to entailment in denotational classical predicate logic.
2. Under the dynamic update interpretation of L_1 , $\tau \models \phi$ corresponds to entailment in relational dynamic predicate logic.
3. Compared to the intensional denotational semantics for questions of G&S (1997). Under the classical update interpretation of L_2 :
 - (a) If τ is a sequence of interrogatives, $\tau \models \psi?$ corresponds to the entailment relation between questions
 - (b) If ϕ is an indicative, $\phi \models \psi?$ corresponds to the answerhood relation.

Anything non-familiar that we can expect, must have to do with the possibility that a quantifier in an indicative can bind a variable in a subsequent interrogative, as exemplified in (2). The first of these two examples was the interrogation:

(4) $\exists xPx; ?Qx; \neg Qx$

When we read (4) as an interrogation in the appropriate way, the last utterance is an answer to the issue raised in the second utterance.

Note, first of all, that it makes no sense to put this as $\neg Qx \models \exists xPx; ?Qx$, which we do in the denotational semantics to assert that the relation of answerhood holds. From a dynamic perspective, order matters, and answers come *after* questions have been posed, and not before. In dynamic semantics, the entailment relation itself is dynamic. We will have to give a different account of answerhood, than in terms of $\phi \models \psi?$. Dynamically, the latter means that *after* having been told that ϕ , it makes no sense anymore to ask $\psi?$.

What we can say about what (4) entails dynamically is the following:

- (5)
- a. $\exists xPx; ?Qx; \neg Qx \models \exists x(Px \wedge \neg Qx)$
 - b. $\exists xPx; ?Qx; \neg Qx \models ?\exists x(Px \wedge \neg Qx)$
 - c. $\exists xPx; ?Qx; \neg Qx \not\models \forall x(Px \rightarrow \neg Qx)$

In (5a) we find precisely what information the discourse provides to the interrogator. As (5c) indicates, the interrogator cannot conclude, after having been told that there is someone walking in the park, having asked whether he is wearing blue suede shoes, and having received a negative reply to that, that whoever is walking in the park is not wearing such shoes, or, for that matter, that there is just one such individual. It is just the one person that the first speaker apparently has in mind, who is not wearing such shoes. The question of the interrogator apparently has the role of trying to figure out whether of the people the first speaker could mean, it is (the) one who is wearing these extravagant shoes. And the answer eliminates just that possibility.

Likewise for the second example in (2):

(6) $\exists xPx; ?y(y = x); x = a$

The discourse in (6) only provides the information that a has the property P , not that he is the only one, and hence it does (at most) provide a partial answer to the question who have the property P .

- (7) a. $\exists xPx; ?y(y = x); x = a \models Pa$
 b. $\exists xPx; ?y(y = x); x = a \not\models Pa \wedge \forall x(Px \rightarrow x = a)$
 c. $\exists xPx; ?y(y = x); x = a \not\models ?xPx$

Note that the following does not hold either:

- (8) $\exists x(Px \wedge \forall y(Py \rightarrow x = y)); ?y(y = x); x = a \not\models ?xPx$

What even does not hold is:

- (9) $\exists x(Px \wedge \forall y(Py \rightarrow x = y)); ?y(y = x); x = a \not\models ?x(x = a)$

The interrogator may have no idea who a is. So, after having been told that a is the one who has the property P , it can make sense to ask who a is.⁷

These examples concerning dynamic entailment should do for the moment. A more systematic study has to await another occasion. There is still another task lying ahead of us. Dynamically, $\phi \models \psi?$ does not provide us with an appropriate notion of answerhood. So, we have to provide one.

7 Coherence

We discuss answerhood in the wider context of logical coherence of interrogations. The link between the two is that, if a discourse $\tau; \phi$ is to be coherent, then if τ raises an issue, and ϕ is an indicative, ϕ should address the issue raised by τ . In principle, ϕ should provide at least a partial answer to an issue raised by τ . The question whether an indicative ϕ is an answer to an interrogative $\psi?$, then boils down to the question whether $\psi?; \phi$ is coherent.

We want the notion of coherence to be compositional in the sense that we can say that a sequence of sentences $\phi_1; \dots; \phi_n$ is coherent iff ϕ_i is coherent relative to $\phi_1; \dots; \phi_{i-1}$, for $1 \leq i \leq n$.⁸

We will not ask much of the interrogator, she may raise new issues at liberty, which do not have to be related to what she asked before, nor to what the interrogated person has brought forward. Of the person heard we do require full cooperation.

A minimal requirement for coherence of a discourse is consistency:

Definition 16 (Consistency)

- ϕ is consistent with σ iff $\sigma[\phi]$ exists, and $\sigma[\phi] \neq \emptyset$
 ϕ is consistent with τ iff $\exists \sigma: \phi$ is consistent with $\sigma[\tau]$.

For any interrogative $\phi?$, and discourse τ , $\phi?$ is consistent with τ as soon as $\phi?$ is defined in $\sigma[\tau]$. Consistency requires little from the interrogator. And from the person heard not very much is required either, $p \vee \neg p$ would already be fine, at any point. But such an equivocation is excluded if we require that next to being consistent with τ , ϕ is informative after τ .

Definition 17 (Informativeness)

- ϕ is informative in σ iff $\sigma[\phi]$ exists and $\sigma \not\vdash \sigma[\phi]$
 ϕ is informative after τ iff $\exists \sigma: \phi$ is informative in σ .

What informativeness of a sentence ϕ at some point in a discourse requires is that ϕ is not already entailed by the preceding discourse. For an interrogative this means that it should not raise an issue that has already been raised, or is already resolved by what went before.

⁷Questions of identity are not an easy affair, In Gerbrandy (1997) it is argued that the way we look at it here is far from satisfactory.

⁸The notion of coherence defined in GSV (1996), lacks this feature of compositionality.

Although informativeness requires of the person heard, that she potentially provides some information to the interrogator, it does not require that the information she provides has any relation to the issues the questioner has raised.

Definition 18 (Resolution)

ϕ is *resolvent* in σ iff $\sigma[\phi]$ exists, and $\exists i \in \sqrt{\sigma} : \forall j : \langle i, j \rangle \in \sigma \Rightarrow j \notin \sqrt{\sigma}[\phi]$.

For an indicative to be resolvent in σ it is not enough to eliminate just some possibilities in σ , but all possibilities in some alternative in σ are to be eliminated, thereby at least partially resolving an issue in σ . Only indicatives, and hence only the person heard, can resolve an issue.

Definition 19 ((Maximal) Resolution) Let $\phi \in L_1$.

ϕ is *(maximally) resolvent after τ* iff

for every indifferent state σ such that $\sigma[\tau][\phi]$ exists:

if $\sigma[\tau]$ is not an indifferent state;

and ϕ is consistent with $\sigma[\tau]$;

(and ϕ is informative in $\sigma[\tau]$)

then ϕ is resolvent in $\sigma[\tau]$

In case τ consists of indicatives only, resolution by ϕ is trivially guaranteed. Indicatives can raise no issues, if we start out from an indifferent state, $\sigma[\tau]$ will still be indifferent. As long as no questions have been posed, or as soon as all issues are resolved, the person heard can bring up whatever she likes.

Consider the simple case where τ consists of a single interrogative $\psi?$ Then the definition requires that for every initial state σ such that $\psi?$ raises an issue in σ , if ϕ is consistent with σ and σ does not subsist in $\sigma[\phi]$, then ϕ resolves an issue in $\sigma[\psi?]$. As soon as ϕ is consistent with σ , and not yet supported by σ , ϕ should eliminate at least one whole block in the partition $\psi?$ makes on σ .

Suppose ϕ is consistent and informative per se to begin with. Then ϕ will certainly be consistent with and informative in the minimal state. And suppose that $\not\models \psi?$, i.e., the minimal state updated with $\psi?$ is not a state of indifference. Then the definition requires that ϕ should eliminate at least one block in the partition made by $\psi?$ in the minimal state, i.e., on the whole logical space. (And under the assumption that ϕ is consistent, leave, at least part of one block there as well.) This requires nothing else than that ϕ gives a partial answer to $\psi?$ in the denotational semantic sense of the word.

If we drop the condition that ϕ is informative in $\sigma[\tau]$ from the definition of resolution, we obtain a notion of complete answerhood, rather than partial answerhood. Actually, under a classical update interpretation of the language, that notion coincides with the notion of (complete) answerhood in terms of entailment.

Fact 2 (Answerhood) Let $\phi \in L_1$, \models_c classical entailment.

$\phi \models_c \psi?$ iff ϕ is maximally resolvent after $\psi?$

Finally, we combine consistency, informativeness and resolution in a logical notion of coherence:

Definition 20 ((Maximal) Coherence) Let $\phi \in L_1$.

ϕ is *(maximally) coherent after τ* iff

ϕ is consistent with τ ; and

ϕ is informative after τ ; and

ϕ is (maximally) resolvent after τ

A coherence requirement for interrogatives, requires only informativeness.

As we wanted, the notion of coherence is compositional:

Definition 21 (Compositional Coherence) Let $\tau = \langle \phi_1, \dots, \phi_n \rangle$.

τ is *coherent* iff ϕ_i is coherent after $\langle \phi_1, \dots, \phi_{i-1} \rangle$, for $1 \leq i \leq n$.

The notion of coherence we defined is absolute. It only distinguishes between non-coherence, and maximal and minimal coherence, and nothing in between.

8 Relevance

Under the absolute notion, both the interrogations $?p; \neg p$ and $?p; \neg p \wedge q$ count as maximally coherent. And so do for example both $\neg q; ?p; p \vee q$ and $\neg q; ?p; p$. But they don't seem equally perfect.

Definition 22 (Comparative Informativeness)

ϕ is at least as informative after τ as ψ iff

$\forall \sigma$: if ψ is informative in $\sigma[\tau]$, then ϕ is informative in $\sigma[\tau]$.

ϕ is at least as informative as ψ iff ϕ it is at least as informative as ψ after $\langle \rangle$.

Under a classical interpretation, ' ϕ is at least as informative as ψ ' is the same as ' ϕ classically and statically entails ψ '.

The notion of comparative informativeness makes it possible to distinguish between the interrogations $?p; \neg p$ and $?p; \neg p \wedge q$. The answer $\neg p$ is less informative after $?p$ than $\neg p \wedge q$. However, the extra information that q , is irrelevant to the issue raised by $?p$. The combination of the two clauses of comparative informativeness makes it possible to distinguish between the two.

Things will work out in such a way that $?p; \neg p$ counts as more coherent than $?p; \neg p \wedge q$. However, $?p; \neg p; q$ remains maximally coherent. After first having resolved the issue whether p , the person heard is free to start providing whatever information she wants. It is counted as more coherent to keep such extra information out of the answer as such.

Likewise, after $\neg q; ?p$, the answers $p \vee q$ and p are equally informative. However, the latter appears to be a more coherent way to address the issue raised. It seems that if we can choose between two answers which in the context are equally resolvent, the logically weaker answer, the answer which is less informative out of context, is preferred.

A second point to note is that the absolute notion of coherence, although it can distinguish between the minimal and the maximal case, does not in general have a preference for more exclusive answers over less exclusive ones. This is covered by the following comparative notion:

Definition 23 (Comparative Resolution)

ϕ is at least as resolvent after τ as ψ iff

$\forall \sigma$: if ϕ is resolvent in $\sigma[\tau]$, then ψ is resolvent in $\sigma[\tau]$.

After $?xPx$, the answer $Pa \wedge Pb$ counts as more resolvent than just Pa , and the latter as less resolvent than $Pa \wedge \forall x(Px \rightarrow (x = a))$. But the last and the first are not related in this way.

I put the notions of comparative informativeness and resolution to cooperative work in a comparative notion of relevance:

Definition 24 (Relevance) ϕ is at least as relevant after τ as ψ iff

1. ϕ is at least as resolvent after τ as ψ or
2. ϕ and ψ are equally resolvent after τ and
 - (a) ϕ is at most as informative after τ as ψ or
 - (b) ϕ and ψ are equally informative after τ and
 - ϕ is at least as informative as ψ

Relevance favours sentences which eliminate more alternatives. Among sentences which eliminate the same alternatives, relevance favours the ones which eliminate less possibilities after a state has been updated with τ . Among those that do the same in that respect as well, relevance favours the

logically stronger sentences. In terms of the notion of relevance, the examples discussed in this section can all be accounted for.

Since they eliminate *all* alternatives, sentences inconsistent with the preceding discourse count as more relevant than those that are consistent with it, but such sentences are disqualified by the following notion of comparative coherence:

Definition 25 (Comparative Coherence) Let ϕ, ψ be consistent with τ .

ϕ is at least as coherent after τ as ψ iff ϕ is at least as relevant after τ as ψ .

The notion of comparative coherence does comparable work to the notions which compare answers defined in G&S (1997) and in Ch. 4 of G&S (1984). As the examples given above make clear, the notion enables us to compare interrogations as to how well the issues that are being raised are addressed and resolved.⁹

9 Some Examples

A warning ahead. I intend the examples below only as illustrations of the notions defined above. I make no claims to the effect that they provide an alternative explanation as compared to other approaches.

Consider the following example. In a neutral context, (10a) is ambiguous between (10b) and (10c):¹⁰

- (10) a. Alf rescued Bea. And no-one else.
 b. $Rab; \forall x(Rxb \rightarrow (x = a))$
 c. $Rab; \forall x(Rax \rightarrow (x = b))$

Both (10b) and (10c) are, of course, coherent sequences. Both are consistent, informative, and (trivially) resolvent, given initial indifference.

However, after the interrogatives in (11a) and (12a), or with the intonational information indicated by underlining in (11a) and (12a), (11a) corresponds with (11b) only, and (12a) only with (12c):

- (11) a. (Who rescued Bea?) Alf rescued Bea. And no-one else.
 b. $?xRxb; Rab; \forall x(Rxb \rightarrow (x = a))$
 c. $?xRxb; Rab; \forall x(Rax \rightarrow (x = b))$
- (12) a. (Whom did Alf rescue?) Alf rescued Bea. And no-one else.
 b. $?xRax; Rab; \forall x(Rxb \rightarrow (x = a))$
 c. $?xRax; Rab; \forall x(Rax \rightarrow (x = b))$

Note that if we swap the interrogatives in (11a) and (12a), the resulting sequences are unacceptable. In a presuppositional set-up, we might account for that, along not unusual lines, by assuming that the intonation pattern of the first indicative in (11a) and (12a), presupposes the issue raised by the interrogatives in (11a) and (12a). We can look upon the sequences in (b) and (c) as the result of presupposition accommodation.

Compositional coherence is in accordance with the difference between (11) and (12). The sequences (11b) and (12c) are both coherent. Rab is a partial answer to $?xRxb$ and $?xRax$. Given that $?xRxb$ asks for the specification of the (whole) denotation of the property $\lambda xRxb$, the answer that a does have that property, leaves open the issue who else does. And this is precisely the issue that $\forall x(Rax \rightarrow (x = b))$ resolves. Likewise for $\forall x(Rxb \rightarrow (x = a))$ after the partial answer Rab to $?xRax$.

⁹In G&S (1984) this is put to work in an account of conversational implicatures, and the same could be done here. But I leave that for another occasion.

¹⁰English is not the perfect language for this type of example, because of the easy availability of do-support. Lacking do-support, Dutch would be better.

But *not* the other way around. The sequences in (11c) and (12b) are incoherent. That Alf rescued no-one else but Bea, is not directed towards the resolution of the issue who else but Alf rescued Bea, which is the issue that in principle is still open after $?xRxb; Rab$. The last sentences in (11c) and (12b) address issues which after an update of an indifferent state with the first two sentences, the state is most likely indifferent towards. And there will certainly be states in which after an update with the first two sentences the last sentence is informative and consistent. But then it should partially resolve an issue, which it need not do at all.

Compare (11) with (13):

(13) (Who rescued Bea?) Alf rescued Bea. And, actually, no-one else.

Unlike in (11), in (13) the ambiguity of (10) seems to turn up again. Actually, I tend to believe that for (13) the reading in (11c) is more salient than the reading in (11b).

The word 'actually' crucially seems gives rise to the availability of both readings. Apparently, the conversational effect of 'actually', is an indication of the fact that the issue at hand is being overruled. Our simple logic does not have such tools. If it had, we might be able to deal with these cases.

Some other examples. The two sentence sequence in (14a) is just as alright as the three sentence sequence in (11a). And from the unavailability of the reading (11c) for (11a), we might expect that (15a) is hardly acceptable.

(14) a. Who rescued Bea? Only Alf rescued Bea.

b. $?xRxb; Rab \wedge \forall x(Rxb \rightarrow (x = a))$

(15) a. ??Who rescued Bea? Alf rescued only Bea.

b. $?xRxb; Rab \wedge \forall x(Rax \rightarrow (x = b))$

But in this case, the absolute notion of coherence does not give a result which can account for the marginal status of (15a). According to that notion, the sequence in (15b) is coherent. The fact that the two last sentences in (11a) are compressed into a single utterance makes a difference. Compositional coherence goes utterance by utterance. And the indicative sentence in (14) as a whole counts as at least partially resolving the issue raised by the interrogative, since it properly implies that Alf was someone who rescued Bea. But that it properly implies that, is precisely the trouble. There is an obvious alternative answer, just saying that Alf was (the) one who rescued Bea, which would result in a *more coherent* discourse, since it is *more relevant*. And the comparative notion accounts precisely for that. The answer 'Alf rescued Bea' counts as more coherent. The information that Alf rescued no-one else, may of course be given, but only after the issue who rescued Bea has been resolved as good as it can. The following examples also give an illustration of that:

(16) Did Alf rescue Bea? Yes he did. And (in fact) he rescued only Bea.

(17) ??Did Alf rescue Bea? Alf rescued only Bea.

The last two sentences of (16), and the last sentence in (17) provide the same information. Still, the discourse (16), where we first just resolve the issue raised by the interrogative, and then go on to provide some extra information that is not asked for as such, is alright. But if we make the answer as such over-informative, as in (17), by putting the extra information already in it, the resulting discourse is much less acceptable. The comparative notion of coherence accounts for this. The utterance that directly follows the question in (16) is more relevant to it, than that in (17). That obvious alternative is preferred.

10 Specificity

In Section 3, Definition 7, we introduced a notion of *specific assignment*, $\sigma[x] = \{\langle i[x/d], j[x/d] \rangle \mid \langle i, j \rangle \in \sigma \ \& \ d \in D\}$, that we did not put to use. If we would put it to use, we would do so in the

definition of existential quantification as $\sigma[\exists x\phi] = \sigma[x][\phi]$. But this requires a direct statement of the update interpretation of the indicative part of the language relative to structured states, instead of the step-wise procedure we followed above. There is no problem in doing so:¹¹

Definition 26 (Dynamic Update Semantics L_2) Let σ be a structured state, $\phi \in L_2$. The update of σ with ϕ , $\sigma[\phi]$, is recursively defined as follows:

$$\sigma[R\vec{t}] = \{\iota \in \sigma \mid \iota_1(\vec{t}) \in i_1(R) \ \& \ \iota_2(\vec{t}) \in i_2(R)\}$$

$$\sigma[\neg\phi] = \{\iota \in \sigma \mid \iota_1, \iota_2 \not\prec \sqrt{\sigma[\phi]}\}$$

$$\sigma[\phi \wedge \psi] = \sigma[\phi][\psi]$$

$$\sigma[\exists x\phi] = \sigma[x][\phi]$$

$$\sigma[?\vec{x}\phi] = \{\iota \in \sigma \mid \sigma[\lambda\vec{x}\phi](\iota_1) = \sigma[\lambda\vec{x}\phi](\iota_2)\}$$

where $\sigma[\lambda\vec{x}\phi]$ is the function f from $\sqrt{\sigma}$ into D^n such that:

$$\forall i \in \sqrt{\sigma}: f(i) = \{\vec{e} \in D^n \mid i[\vec{x}/\vec{e}] \prec \sqrt{\sigma[\vec{x}][\phi]}\}$$

Except for the interpretation of existential quantification nothing has contentwise changed. But the change in the effect of quantification is significant: $\exists x\phi \models ?y(y = x)$. Raising issues is no longer the privilege of interrogatives. However, this holds only under the dynamic interpretation, statically, if we take $\sigma[\exists x\phi] = \{\iota \in \sigma \mid \iota \prec \sigma[x][\phi]\}$, the specific assignment has no effect. In general, where the dynamic effects are blocked, the issue raised by an existential quantifier is also annihilated.

By the way, we don't *have* to define the quantifier this way, if we don't feel like it. We could also define $\sigma[x] = \{\langle i[x/d], j[x/e] \rangle \mid \langle i, j \rangle \in \sigma \ \& \ d, e \in D\}$. If we use that in the definition above, everything remains as it was.

But the logical and empirical intuition is, that the quantifier which raises an issue, belongs more naturally to the system, and that we should investigate what it can do in an analysis of the specific use of indefinites. And, although many, including myself, have argued specificity to be a pragmatic phenomenon, it would suddenly turn out to be a dynamic semantic affair. But that switch, although it went unmentioned, happened throughout the story I told here.

References

- Aloni, M. (1997). Quantification in Dynamic Semantics. In *Proceedings of the Eleventh Amsterdam Colloquium*, P. Dekker, M. Stokhof & Y. Venema eds pages 73–78. ILLC, University of Amsterdam.
- Gerbrandy, J. (1997). Questions of Identity. In *Proceedings of the Eleventh Amsterdam Colloquium*, P. Dekker, M. Stokhof & Y. Venema eds pages 133–138. ILLC, University of Amsterdam.
- Groenendijk, J., & M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Dissertation. University of Amsterdam.
- Groenendijk, J., & M. Stokhof (1997). Questions. In *Handbook of Logic and Language*, J. van Benthem & A. ter Meulen eds, pages 1055–1124. Elsevier.
- Groenendijk, J., M. Stokhof, & F. Veltman (1996). Coreference and Modality. In *The Handbook of Contemporary Semantic Theory*, S. Lappin ed., pages 179–213. Blackwell.
- Hulstijn, J. (1997). Structured Information States. - Raising and Resolving Issues -. In *Proceedings MunDial 97*, A. Benz & G. Jäger eds, pages 99–117. Universität München.
- Jäger, G. (1995). Only Updates. In *Proceedings of the Tenth Amsterdam Colloquium*, P. Dekker & M. Stokhof eds pages 387–405. ILLC, University of Amsterdam.

¹¹Although syntactically excluded, you can look at what $\neg?\vec{x}\phi$, and $?x?y\phi$ would mean: $\forall\sigma:\sigma[\neg?\vec{x}\phi] = \emptyset$ & $\sigma[?x?y\phi] = \sqrt{\sigma}^2$. Good they are excluded!

Theory of Mutuality (Syntactic Skeleton)

—Extended Abstract—

Wolfgang Heydrich
University of Hamburg / University of Bielefeld
Abendrothsweg 23
D-20251 Hamburg
heydrich@lili.uni-bielefeld.de

Abstract

Lewis (1969) has analyzed common knowledge as mutual reason to believe. I shortly discuss his proposals in order to reconstruct the account in the framework of relevance logic (system R with quantifiers for individuals—persons and situations—and propositions). I formulate three postulates concerning the primitives of Lewis' analysis. A notion of '*p*'s being public information in group *G*' (in symbols: ' P_{GP} ') arises which licenses reason to believe of arbitrary higher (finite and transfinite) degree. Being able to prove that P_{GP} entails $P_G P_{GP}$, I wonder which postulates could be added to secure the converse.

1 Lewis' Proposal

According to Lewis (1969), "it is *common knowledge* in a population *P* that ____ if and only if some state of affairs *A* holds such that: •

- (1) *Accessibility* Everyone in *P* has reason to believe that *A* holds.
- (2) *Perspicuity* *A* indicates to everyone in *P* that everyone in *P* has reason to believe that *A* holds.
- (3) *Significance* *A* indicates to everyone in *P* that ____." (p.56)

Here, the notion of indication is understood such that

"*A* indicates to someone *x* that ____ if and only if, if *x* had reason to believe that *A* held, *x* would thereby [*sic!*] have reason to believe that ____." (p.52s.)

Interestingly, Lewis formulates his account in terms of *reason to believe* (instead of *belief*) and *indication of propositional content* (instead of *induction of belief*). The explicandum is concerned with features of groups of agents ("populations"). These features are not to be reduced to psychological properties of persons.

In ascribing reason to believe to a person we are not necessarily (or exclusively) concerned with her inner (mental) state. Perhaps, external (environmental) features are more important. Reason to believe concerns the information *available* for an agent in a situation (not the information actually *processed* by her). Reasons to believe are rooted in the informational supply of external situations. Having reason to believe that *p* does not imply believing that *p*; and believing that *p* does not imply having reason to believe that *p*. Nevertheless both notions are closely linked by the constraint of rationality: In the absence of reason to believe the contrary (i.e. *non-p*), an agent might be called rational with respect to *p*, just in case her having reason to believe that *p* is sufficient for her believing that *p*.

Although *having reason to believe* is in itself not a notion of *psychological* epistemology, it is epistemologically highly relevant. One might think of it as a notion of *logical* epistemology.

Notice that, according to one (albeit not unproblematic) explication of knowledge (“justified true belief”), it is just this notion which adds up with the features of *truth* and *belief* to constitute the explicans.

Lewis’ account is intended to license nested reason to believe. Given certain ancillary premisses, common knowledge is—at least in his sense—to imply *mutual* reason to believe: nested reason to believe of arbitrary degree. Lewis’ demonstration of this point, however, is somewhat loose and informal. One of the tasks of this paper is to provide a more explicit reconstruction.

Lewis’ notion of *common knowledge* (like (*mutual*) *reason to believe*) is an environmental concept (not a mental one): it characterizes groups in the informational set-up of a common situation. This very feature makes it a promising starting-point for an *externalist* theory of *common ground*, *presuppositions* and related notions.

There are, however, several questions which need clarification in order to give a formal reconstruction of Lewis’ ideas.

The first point is perhaps merely terminological. Calling Lewis’ explicandum “common knowledge” does not seem quite felicitous. If common knowledge is a kind of knowledge (as one is led to assume), one would expect that *p*’s being common knowledge in some population implies *p*. But this does not seem to be the case as far as Lewis’ account goes. Instead of *p*’s being *common knowledge* within population *P*, I will speak about its being *public information* in group *G* (in symbols: ‘ P_{GP} ’). P_{GP} should *not* imply *p*.

Second, Lewis’ notion of *common knowledge* is relativized to populations. Analogously, my notion of *public information* is tied to a parameter for groups of persons.—It does not bother me here, what groups are. The only feature that interests me is that groups have persons as members (persons are conceived of as a sort of individuals). I am writing “ $x \in G$ ” to express that person *x* is a member of group *G*.

Third, Lewis refers to *states of affairs*. Obviously, identity criteria for this kind of entities need clarification.—I am treating them as situations (another sort of *individuals*). This eliminates the need of special identity conditions.

Fourth, we need an account of Lewis’ primitives of *holding* and *having reason to believe*.—I am rendering the first as a one-place predicate of situations (in symbols: ‘ Hs ’, where *s* is a situation) and the second as an operator sending persons and propositions to propositions (in symbols: ‘ $R_x p$ ’, where *x* is a person and *p* is a proposition). I will give a simple and very rudimentary axiomatic characterization of these primitives, which is not intended as a comprehensive analysis. It will suffice, though, to derive the central result of this paper: that public information implies nested reason to believe of arbitrary (finite and transfinite) degree.

Fifth, condition (2) in Lewis’ explication of common knowledge exhibits an ambiguity between *de re* and *de dicto*. Given my rudimentary characterization of the primitives involved, neither reading implies the other.—I will substitute Lewis’ condition of *perspicuity* by a condition of *transparency* which (together with accessibility) is strong enough to imply the *de re* version of perspicuity.

Sixth, the logic of the conditional used needs specification. Obviously, given Lewis’ definition of indication, material implication will not do, since some state of affairs may hold and still fail to indicate to me that *p*, in case I have independent reason to believe that *p*. Strict implication and Lewis’ (1973) counterfactual conditional, however, do not fare much better. Someone may have by necessity reason to believe in some truth, and still some (thematically unrelated) state of affairs may hold, without indicating this special truth to him. These considerations motivate appeal to a relevant conditional. Note that defining *s*’s indicating *p* to *x* (along the lines of Lewis’ proposal) by the constraint ‘ $R_x Hs \rightarrow R_x p$ ’ involves us with (at least) two sources of intensionality: the *R*-operator is obviously intensional and the arrow must be intensional too. The question arises of how to distribute the overall intensionality involved in the notion of *indication* in an optimal way.—The proposal of this paper is to understand the arrow in the sense of the relevant system *R* (cf. Anderson and Belnap 1973; 1992), since this allows for quite a simple characterization of the *R*-operator.

2 Background Theory

Let me be more explicit about the background theory of my reconstruction.

Language

Let $x, x_1, x_2, \dots, y, z, \dots, s, t, \dots$ be individual variables ($x, x_1, x_2, \dots, y, z, \dots$ for persons, s, t, \dots for situations), and p, q, r, h, \dots propositional variables. Let “ G ” be a parameter for groups of persons, “ H ” a one-place predicate of situations, and “ R ” an operator sending pairs of persons and propositions to propositions.

The set of formulas is the smallest set containing for each variable σ for situations, ξ for persons, τ for individuals (persons or situations) and π for propositions (i) ‘ $H\sigma$ ’, ‘ $\xi \in G$ ’, and ‘ π ’ as well as (ii) ‘ $\neg\beta$ ’, ‘ $(\alpha \wedge \beta)$ ’, ‘ $(\alpha \rightarrow \beta)$ ’, ‘ $R_x\alpha$ ’, ‘ $\forall\tau\alpha$ ’, and ‘ $\forall\pi\alpha$ ’ if it contains α and β .

“ Hs ” may be read “ s obtains”. “ $x \in G$ ” may be read “ x is a member of G ”. “ $R_x p$ ” may be read “ x has reason to believe that p ”.

I will often suppress outermost (and otherwise redundant) parentheses and write ‘ $\alpha \leftrightarrow \beta$ ’, ‘ $\alpha \vee \beta$ ’, ‘ $\alpha \supset \beta$ ’, ‘ $\alpha \equiv \beta$ ’, ‘ $\exists\tau\alpha$ ’, and ‘ $\exists\pi\alpha$ ’ for ‘ $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$ ’, ‘ $\neg(\neg\alpha \wedge \neg\beta)$ ’, ‘ $\neg\alpha \vee \beta$ ’, ‘ $(\alpha \supset \beta) \wedge (\beta \supset \alpha)$ ’, ‘ $\neg\forall\tau\neg\alpha$ ’, and ‘ $\neg\forall\pi\neg\alpha$ ’ respectively. Additionally, I will use ‘ $\forall\xi \in G\alpha$ ’ for ‘ $\forall\xi(\xi \in G \supset \alpha)$ ’ and ‘ $\exists\xi \in G\alpha$ ’ for ‘ $\exists\xi(\xi \in G \wedge \alpha)$ ’ as well as ‘ $\forall\xi_1, \dots, \xi_n \in G\alpha$ ’ for ‘ $\forall\xi_1 \in G \dots \forall\xi_n \in G\alpha$ ’ and ‘ $\exists\xi_1, \dots, \xi_n \in G\alpha$ ’ for ‘ $\exists\xi_1 \in G \dots \exists\xi_n \in G\alpha$ ’ (where $0 \leq n$ and ξ_1, \dots, ξ_n are variables for persons).

Axioms and Rules

The axioms and rules I am assuming within the background theory (QR) correspond to relevance logic (system R) with propositional and individual quantifiers.¹ Let α and β be formulas and χ an arbitrary variable (either propositional or individual). Let ‘ $\alpha[\gamma/\chi]$ ’ be the result of proper substituting γ for χ in α for arbitrary expressions γ .

- AX: $\vdash \alpha \rightarrow ((\alpha \rightarrow \beta) \rightarrow \beta)$
A1: $\vdash (\alpha \rightarrow \beta) \rightarrow ((\gamma \rightarrow \alpha) \rightarrow (\gamma \rightarrow \beta))$
A2: $\vdash (\alpha \rightarrow (\alpha \rightarrow \beta)) \rightarrow (\alpha \rightarrow \beta)$
A3: $\vdash (\alpha \rightarrow \neg\beta) \rightarrow (\beta \rightarrow \neg\alpha)$
A4: $\vdash \neg\neg\alpha \rightarrow \alpha$
A5: $\vdash (\alpha \rightarrow \neg\alpha) \rightarrow \neg\alpha$
B1: $\vdash (\alpha \wedge \beta) \rightarrow \alpha$
B2: $\vdash (\alpha \wedge \beta) \rightarrow \beta$
B3: $\vdash ((\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma)) \rightarrow (\alpha \rightarrow (\beta \wedge \gamma))$
B7: $\vdash (\alpha \wedge (\beta \vee \gamma)) \rightarrow ((\alpha \wedge \beta) \vee \gamma)$
C1: $\vdash \forall\chi(\alpha \rightarrow \beta) \rightarrow (\forall\chi\alpha \rightarrow \forall\chi\beta)$
C2: $\vdash (\forall\chi\alpha \wedge \forall\chi\beta) \rightarrow \forall\chi(\alpha \wedge \beta)$
C3: $\vdash \forall\chi\alpha \rightarrow \alpha[\gamma/\chi]$ (where *either* γ is a formula and χ a propositional variable *or* both are individual variables)
C4: $\vdash \alpha \rightarrow \forall\chi\alpha$ (χ is not free in α)
C5: $\vdash \forall\chi(\alpha \vee \beta) \rightarrow (\alpha \vee \forall\chi\beta)$ (χ is not free in α)
MP: If $\vdash \alpha \rightarrow \beta$ and $\vdash \alpha$, then $\vdash \beta$.
&: If $\vdash \alpha$ and $\vdash \beta$ then $\vdash \alpha \wedge \beta$.
Q: If α is an axiom, so is $\forall\chi\alpha$.

Since all the axioms and rules in this list are valid classically (under the interpretation of the arrow as material implication), they determine a subsystem of classical logic. Actually, if we add

¹Cf. Anderson and Belnap (1973), Anderson and Belnap (1992), the numbering among the propositional axioms follows Heydrich (1995). B4–B6 are left out, because we have defined disjunction in terms of conjunction and negation.

A0: $\vdash \alpha \rightarrow (\beta \rightarrow \alpha)$

(or substitute A0 for AX), we get all the classical laws. However, QR need not be interpreted as a restriction of classical logic, but may also be seen as an extension: in the fragment of arrow-free formulas both systems coincide, and for classical logic, $\{\neg, \wedge, \vee\}$ is a functionally complete subset of logical operators. Hence, QR is adding something substantial to the theory of truth-functions and quantification: the idea of a relevant constraint between propositions.

To get a feeling of what this amounts to, it is useful to have a look at some classically valid formulas which are *not* theorems of QR:

- (1) $(\alpha \supset \beta) \supset (\alpha \rightarrow \beta)$
- (2) a. $(\neg\alpha \wedge \alpha) \rightarrow \beta$
- (2) b. $\neg(\alpha \rightarrow \alpha) \rightarrow \beta$
- (2) c. $\neg(\alpha \rightarrow \alpha) \rightarrow \neg\alpha$
- (3) a. $\beta \rightarrow (\neg\alpha \vee \alpha)$
- (3) b. $\beta \rightarrow (\alpha \rightarrow \alpha)$
- (3) c. $\alpha \rightarrow (\alpha \rightarrow \alpha)$
- (4) $(\alpha \rightarrow \beta) \vee (\beta \rightarrow \alpha)$
- (5) $(\neg\alpha \wedge (\alpha \vee \beta)) \rightarrow \beta$

The invalidity of (1) shows that ‘ \rightarrow ’ (conditional constraint) is stronger than ‘ \supset ’ (material implication): β might be true without being necessary for α , and α might be false without being sufficient for β . (The converse of (1) is valid in QR even with the arrow as main operator.) According to the invalidity of (2)(a), (b) ((3)(a), (b)) there are propositions which are not necessary for inconsistent (not sufficient for valid) propositions. (2)(c) is a special case of (2)(b), but still not valid: in general, the (inconsistent) assumption of a proposition’s not being necessary for itself is not sufficient for the proposition’s being false. (2)(c) is, by contraposition, equivalent² with (3)(c), which is an instance both of A0 and of (3)(b). According to its invalidity, some propositions do not entail the (valid) assumption of being necessary for themselves. According to the invalidity of (4) there are pairs of propositions such that neither of both is sufficient for the other. Rejecting (5) (the disjunctive syllogism) as a valid schema is the core idea of relevance logic. Note that the antecedent of (5) is equivalent² with the disjunction $(\neg\alpha \wedge \alpha) \vee (\neg\alpha \wedge \beta)$ such that (5) implies (2)(a), the paradigm case of *ex falso quodlibet*.

3 Relevant Reconstruction of Lewis’ Account

Following Lewis (1969) proposals, we define p ’s being public information in G by the existence of a *transparent* situation s (in symbols: ‘ Ts ’) which is *accessible* for (the members of) G (in symbols: ‘ A_Gs ’) and *indicates* p to all of them (in symbols: ‘ $s \models_G p$ ’).

- DEF1: $Ts \quad := \quad Hs \rightarrow (Hs \rightarrow Hs)$
- DEF2: $A_Gs \quad := \quad \forall_{x \in G} R_x Hs$
- DEF3: $s \models_G p \quad := \quad \forall_{x \in G} (R_x Hs \rightarrow R_x p)$
- DEF4: $P_G p \quad := \quad \exists s (Ts \wedge A_Gs \wedge s \models_G p)$

Notice that our rendering of transparency diverges considerably from Lewis’ original formulation of perspicuity.

Understood classically (with \rightarrow as material implication), Ts would be a valid proposition. (The same is true, if we interpret the arrow as strict implication or in the sense of Lewis’ (1973) counterfactual conditional.) In QR, however, Ts —an instance of (3)(c)—is contingent.

We will show in a moment (cf. TH 3 below) that, given some straightforward postulates, the conjunction of Ts and A_Gs implies Lewis’ condition for perspicuity in its *de re* version.

We add three postulates to the axioms and rules specified in section 2. Let α and β be arbitrary formulas:

²Not only in the \equiv -sense, but also in the \leftrightarrow -sense.

- P1: $\vdash(\alpha \rightarrow \beta) \rightarrow (R_x\alpha \rightarrow R_x\beta)$
P2: $\vdash\forall_{x \in G} P_G\alpha \leftrightarrow P_G\forall_{x \in G}\alpha$
P3': $\vdash P_G\alpha \wedge P_G\beta \rightarrow P_G(\alpha \wedge \beta)$

P1 expresses that relevant constraints *carry* reason to believe: they entail that reasons to believe their antecedens are sufficient reasons to believe their consequens. P2 stipulates that public information (within group G) does not exhibit a *de re/de dicto*-ambiguity in the domain of members of G . Finally, P3 claims that public information accumulates conjunctively in one and the same group. (In section 5, we will propose extensions of this system by (i) generalizing P3' and (ii) adding a special inference rule (Γ) or an additional postulate (P4).)

Our first theorem states that relevant constraints carry public information³:

- TH 1: $\vdash(p \rightarrow q) \rightarrow (P_Gp \rightarrow P_Gq)$
Proof: Assume $p \rightarrow q$. Hence $R_xp \rightarrow R_xq$ (P1); hence $s \models_G p \rightarrow s \models_G q$ (DEF 3); hence $P_Gp \rightarrow P_Gq$ (DEF 1,2,4). Qed.

Transparent situations are constrained to indicate their transparency to the members of arbitrary groups:

- TH 2: $\vdash Ts \rightarrow s \models_G Ts$
Proof: Assume Ts . Hence $Hs \rightarrow (Hs \rightarrow (Hs \rightarrow Hs))$ (DEF 1). Hence $\forall_{x \in G}(R_xHs \rightarrow R_xTs)$ (P1, DEF 1). Hence $s \models_G Ts$ (DEF 3). Qed.

The next theorem states that a situation's *de re*-perspicuity follows from its transparency and accessibility:

- TH 3: $\vdash(Ts \wedge A_Gs) \rightarrow \forall_{x \in G} s \models_G R_xHs$
Proof: Assume Ts and A_Gs . Hence $\forall_{x \in G} R_xHs$ (DEF 2) and $R_xHs \rightarrow (Hs \rightarrow R_xHs)$ (DEF 1, P1). Hence $\forall_{x,y \in G}(R_yHs \rightarrow R_yR_xHs)$ (P1), and $\forall_{x \in G} s \models_G R_xHs$ (DEF 3). Qed.

Transparent situations which indicate a proposition to every member in group G are constrained also to indicate, for all x in G , their indicating it to x to everyone in G .

- TH 4: $\vdash(Ts \wedge s \models_G p) \rightarrow \forall_{x \in G} s \models_G (R_xHs \rightarrow R_xp)$
Proof: Assume Ts and $s \models_G p$. Hence $\forall_{x \in G}(R_xHs \rightarrow R_xp)$ (DEF 3). Also:
 $Hs \rightarrow (R_xHs \rightarrow R_xHs)$ (DEF 1, P1), hence $(R_xHs \rightarrow R_xp) \rightarrow (Hs \rightarrow (R_xHs \rightarrow R_xp))$.
Hence $\forall_{x \in G}(Hs \rightarrow (R_xHs \rightarrow R_xp))$ and $\forall_{x,y \in G}(R_yHs \rightarrow R_y(R_xHs \rightarrow R_xp))$ (P1).
Hence $\forall_{x \in G} s \models_G (R_xHs \rightarrow R_xp)$ (DEF 3). Qed.

Now, we can show that the concept of being public information in a group iterates:

- TH 5: $\vdash P_Gp \rightarrow P_G P_Gp$
Proof: Assume P_Gp . So $\exists s(Ts \wedge A_Gs \wedge s \models_G Ts \wedge \forall_{x \in G} s \models_G R_xHs \wedge \forall_{x \in G} s \models_G (R_xHs \rightarrow R_xp))$ (DEF 4, TH 2,3,4). Hence $\exists t(P_GTt \wedge \forall_{x \in G} P_GR_xHt \wedge \forall_{x \in G} P_G(R_xHt \rightarrow R_xp))$ (DEF 4) and $\exists t(P_GTt \wedge P_GA_Gt \wedge P_Gt \models_G p)$ (P2). Hence $\exists t P_G(Tt \wedge A_Gt \wedge t \models_G p)$ (P3') and $P_G\exists t(Tt \wedge A_Gt \wedge t \models_G p)$ (TH 1). Hence $P_G P_Gp$ (DEF 4). Qed.

4 Public Information and Mutual Reason to Believe

In order to show that our notion of public information captures information which is fully transparent and freely accessible (in the relevant group), we would like to show that it implies nested reason to believe of arbitrary (finite and transfinite) degree. The basis idea is that nested reason to believe may be ordered cumulatively.

³Reference to the background theory QR (system R of quantified relevance logic) will be suppressed in the sequel.

Let us define p 's 1^{st} -degree accessibility in G such that every member of G has reason to believe that p :

$$\text{DEF5: } R_G^1 p \quad := \quad \forall_{x \in G} R_x p$$

For $d > 0$, then, $d + 1^{st}$ -accessibility should imply d^{th} -degree accessibility and the successor of d^{th} -degree accessibility. ($R_G^{d+1} p \leftrightarrow (R_G^d p \wedge S_G R_G^d p)$) The question is, how to define the successor-function S_G . Clearly, there are several possibilities. As for $S_G R_G^1 p$, we seem to have at least two plausible options: a *de re* nesting ($\forall_{y \in G} R_y R_x p$, i.e. $\forall_{y, x \in G} R_y R_x p$) and a *de dicto* nesting ($\forall_{y \in G} R_y R_G^1 p$, i.e. $\forall_{y \in G} R_y \forall_{x \in G} R_x p$). As long as we do not assume either to imply the other, we better opt for their conjunction. Hence, we have:

$$\begin{aligned} R_G^2 p &\leftrightarrow (R_G^1 p \wedge S_G R_G^1 p) \text{ and} \\ S_G R_G^1 p &\leftrightarrow (\forall_{y, x \in G} R_y R_x p \wedge \forall_{y \in G} R_y \forall_{x \in G} R_x p). \end{aligned}$$

We will generally define a proposition's successor (with respect to G) as the conjunction of all the nestings (with respect to G) of its necessary conditions.

The nestings (with respect to G) of $\forall_{y, x \in G} R_y R_x p$ (the first conjunct in the analysis of $S_G R_G^1 p$ —hence one of the necessary conditions of $R_G^2 p$) are $\forall_{z, y, x \in G} R_z R_y R_x p$, $\forall_{z, y \in G} R_z \forall_{x \in G} R_y R_x p$ and $\forall_{z \in G} R_z \forall_{y, x \in G} R_y R_x p$; the nestings of $\forall_{y \in G} R_y \forall_{x \in G} R_x p$ (the second conjunct) are $\forall_{z, y \in G} R_z R_y \forall_{x \in G} R_x p$ and $\forall_{z \in G} R_z \forall_{y \in G} R_y \forall_{x \in G} R_x p$. Hence, we get for the third degree:

$$\begin{aligned} R_G^3 p &\leftrightarrow (R_G^2 p \wedge S_G R_G^2 p) \text{ and} \\ S_G R_G^2 p &\leftrightarrow (S_G R_G^1 p \wedge \forall_{z, y, x \in G} R_z R_y R_x p \wedge \forall_{z, y \in G} R_z \forall_{x \in G} R_y R_x p \wedge \forall_{z \in G} R_z \forall_{y, x \in G} R_y R_x p \wedge \\ &\quad \forall_{z, y \in G} R_z R_y \forall_{x \in G} R_x p \wedge \forall_{z \in G} R_z \forall_{y \in G} R_y \forall_{x \in G} R_x p) \end{aligned}$$

Etc., etc. for all degrees $0 < d < \omega$. (Syntactically, one may construct a nesting of a formula by looking for a prefix of the form $\forall_{\xi_1, \dots, \xi_n \in G}$, where $n \geq 0$. Note that each formula has at least one (possibly empty) prefix of this form. Simply substitute $\forall_{\xi_1, \dots, \xi_n \in G} R_\xi$ for the prefix, where the variable ξ (for persons) does not occur free in the formula.)

We capture these notions by the following definitions.

$$\text{DEF6: } N_G(p, q) \quad := \quad \exists r((q \leftrightarrow \forall_{x_1, \dots, x_n \in G} r) \wedge (p \leftrightarrow \forall_{x_1, \dots, x_n \in G} R_x r)) \text{ (for some } n \geq 0)$$

$$\text{DEF7: } S_G p \quad := \quad \forall q \forall r(((p \rightarrow r) \wedge N_G(q, r)) \rightarrow q)$$

$$\text{DEF8: } R_G^{d+1} p \quad := \quad R_G^d p \wedge S_G R_G^d p \text{ (for } d > 0)$$

DEF 6 is a definitional schema. Note that DEF 7 makes use of quantification over propositions. It is, so to speak, stating simultaneously all propositions q which are nestings of necessary conditions of p . This is tantamount to the infinite conjunction of these propositions. Note further:

$$\text{TH 6: } \vdash S_G p \rightarrow R_G^1 p$$

Proof: Assume $S_G p$, i.e. $\forall q \forall r(((p \rightarrow r) \wedge N_G(q, r)) \rightarrow q)$ (DEF 7).

Hence $((p \rightarrow p) \wedge N_G(\forall_{x \in G} R_x p, p)) \rightarrow \forall_{x \in G} R_x p$. Since $p \rightarrow p$ and $N_G(\forall_{x \in G} R_x p, p)$,
 $\forall_{x \in G} R_x p$,
 i.e. $R_G^1 p$ (DEF 5). Qed.

We want to go on by defining $R_G^\omega p$ as tantamount to the infinite conjunction of all propositions $R_G^d p$ (with $0 < d < \omega$). Let us call each conjunction q of one or more propositions in the series: $\alpha[p]$, $\alpha[\alpha[p]]$, $\alpha[\alpha[\alpha[p]]]$,... an *expansion of p with respect to α* . We define:

$$\text{DEF9: } p <_\alpha q \quad := \quad \forall r(((r \rightarrow \alpha) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow \alpha[h/p]))) \rightarrow (r \rightarrow q))$$

We can prove:

$$\text{TH 7: } \vdash p <_{S_G p} R_G^1 p$$

Proof: Assume $r \rightarrow S_G p$. Hence $r \rightarrow R_G^1 p$ (TH 6). Hence $\forall r(((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow R_G^1 p))$, i.e. $p <_{S_G p} R_G^1 p$ (DEF 9). Qed.

TH 8: $\vdash p <_{SGP} q \rightarrow p <_{SGP} S_G q$

Proof: Assume $p <_{SGP} q$, i.e. $\forall r((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow q)$ (DEF 9). Now assume $(r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))$.

Hence $r \rightarrow q$ and $(r \rightarrow q) \rightarrow (r \rightarrow S_G q)$. Hence $r \rightarrow S_G q$.

Hence $\forall r((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow S_G q)$,

i.e. $p <_{SGP} S_G q$ (DEF 9). Qed.

TH 9: $\vdash (p <_{SGP} q \wedge p <_{SGP} u) \rightarrow p <_{SGP} (q \wedge u)$

Proof: Assume $p <_{SGP} q \wedge p <_{SGP} r$,

i.e. $\forall r((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow q) \wedge \forall r((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow u)$.

Hence $\forall r((r \rightarrow S_G p) \wedge \forall h((r \rightarrow h) \rightarrow (r \rightarrow S_G h))) \rightarrow (r \rightarrow (q \wedge u))$.

Hence $p <_{SGP} (q \wedge u)$ (DEF 9). Qed.

Consequently, $p <_{SGP} R_G^d p$ for all degrees d ($0 < d < \omega$).

We may now introduce, for each context α , a *limit-operator* L_α and define R_G^ω as informational accessibility in G up to the smallest transfinite degree:

DEF10: $L_\alpha p := \forall q((p <_\alpha q) \rightarrow q)$

DEF11: $R_G^\omega p := L_{SGP} p$

$L_\alpha p$ states simultaneously (i.e. it is tantamount to the possibly infinite conjunction of) all expansions of p with respect to α . Hence, $R_G^\omega p$ is tantamount to the infinite conjunction of all propositions of form $R_G^d p$ (for $0 < d < \omega$).

Although R_G^ω implies informational accessibility up to infinitely many increasing degrees, it is not yet the notion of information (being fully transparent and freely accessible in G) we need. $R_G^\omega p$ does not imply $S_G R_G^\omega p$, hence it is implied by, but does not imply p 's accessibility up to the next degree $R_G^\omega p \wedge S_G R_G^\omega p$, i.e. $R_G^{\omega+1} p$. Notice that, given DEF 8, we actually have already defined an infinity of further *distinct* successive degrees of accessibility: For all $0 \leq d$: $R_G^{\omega+(d+1)} p \leftrightarrow (R_G^{\omega+d} \wedge S_G R_G^{\omega+d} p)$.

One might ask, however, whether we should not *require* that $R_G^\omega p$ imply $S_G R_G^\omega p$ such that $R_G^\omega p$ and $R_G^{\omega+1} p$ turn out to be equivalent (and all transfinite degrees reduce to ω). Is there really an informational difference (*modulo* \leftrightarrow) between mutual reason to believe (in the sense of R_G^ω) and reason to believe that there is mutual reason to believe ($S_G R_G^\omega$)? A non-arbitrary way to establish the equivalence would be to stipulate

$$(6) \quad \forall p(\alpha \rightarrow S_G p) \rightarrow S_G \forall p(\alpha \rightarrow p)$$

The antecedent of (6) states (roughly) that each proposition of kind α is constrained such that everybody in G has reason to believe it. The consequence of (6) states (roughly, again) that everybody in G has reason to believe that each proposition of kind α is constrained to be true. This is essentially the difference between conditionally constrained reason to believe and having reason to believe a conditional constraint.

Suppose that α determines the kind of propositions that state for some even natural number greater than 2, that it is the sum of two primes. Suppose further that everyone in G disposes of effective procedures to determine the truth of such a proposition. And suppose that Goldbach's conjecture is correct. Then, the antecedens of (6) seems to be true: All propositions determined by kind α are constrained in a way that people in G have reason to believe them. But in view of the absence of a general proof for the conjecture, people in G may very well lack reason to believe that all propositions determined by α are constrained to be true.

Having introduced $R_G^\omega, R_G^{\omega+1}, R_G^{\omega+2}, \dots$, one might go on along these lines and define p 's informational accessibility in G to still greater degrees like $R_G^{2\omega} p$ ($:= R_G^\omega R_G^\omega p$) or $R_G^{\omega^2} p$ ($:= L_{R_G^\omega p} p$), etc. What we are really interested in, however, is full transparency and free accessibility in a sense entailing informational accessibility up to *arbitrary* (finite and transfinite) degree. This is what our notion of public information is to secure. Consider the next three theorems:

TH 10: $\vdash P_{GP} \rightarrow S_{GP}$

Proof: Assume P_{GP} . Now assume $(p \rightarrow q) \wedge N_G(r, q)$. Then P_{Gq} (TH 1) and for some $n \geq 0$ and some proposition u : $q \leftrightarrow \forall x_n, \dots, x_1 \in G u$ and $r \leftrightarrow \forall x_n, \dots, x_1 \in G R_x u$ (DEF 6). Hence $\forall x_n, \dots, x_1 \in G P_{Gu}$ (P 2). Hence $\forall x_n, \dots, x_1 \in G \exists s \forall x \in G (R_x Hs \wedge (R_x Hs \rightarrow R_x u))$ (DEF 2;3,4). Hence $\forall x_n, \dots, x_1 \in G R_x u$, i.e. r . Hence $((p \rightarrow q) \wedge N_G(r, q)) \rightarrow r$. Hence S_{GP} (DEF 7). Qed.

TH 11: $\vdash (P_{GP} \rightarrow q) \rightarrow (P_{GP} \rightarrow S_{Gq})$

Proof: Assume $P_{GP} \rightarrow q$. Hence $P_{GP} \rightarrow P_{Gq}$ (TH 1,5). Assume P_{GP} . Then P_{Gq} . Hence S_{Gq} (TH 10). Hence $P_{GP} \rightarrow S_{Gq}$. Qed.

TH 12: $\vdash \forall p (P_{GP} \rightarrow \alpha) \rightarrow (P_{GP} \rightarrow L_\alpha p)$

Proof: Assume (a) $\forall p (P_{GP} \rightarrow \alpha)$. Further assume (b) P_{GP} , and assume (c) $p <_\alpha q$, i.e. $\forall r (((r \rightarrow \alpha) \wedge \forall h ((r \rightarrow h) \rightarrow (r \rightarrow \alpha[h/p]))) \rightarrow (r \rightarrow q))$ (DEF 9). We derive q from these assumptions. Because of (c): $((P_{GP} \rightarrow \alpha) \wedge \forall h ((P_{GP} \rightarrow h) \rightarrow (P_{GP} \rightarrow \alpha[h/p]))) \rightarrow (P_{GP} \rightarrow q)$. Assume $P_{GP} \rightarrow h$. Hence $P_{GP} \rightarrow P_{Gh}$. (TH 1,5). Now assume P_{GP} , hence P_{Gh} and because of (a): $\alpha[h/p]$. Hence $P_{GP} \rightarrow \alpha[h/p]$. Hence $\forall h ((P_{GP} \rightarrow h) \rightarrow (P_{GP} \rightarrow \alpha[h/p]))$ and, because of (a), $P_{GP} \rightarrow \alpha$. Hence $P_{GP} \rightarrow q$ and, because of (b), q . Hence $(p <_\alpha q) \rightarrow q$ and $L_\alpha p$ (DEF 10). Hence $P_{GP} \rightarrow L_\alpha p$. Qed.

p 's being public information (in G) is sufficient for its 1st-degree accessibility (in G) (TH 6,10). Additionally, it is sufficient for all of its necessary conditions' successors (with respect to G) (TH 11). Finally, it is sufficient for the infinite conjunction of all expansions of p with respect to kind α , provided each public information in G is constrained to be of kind α (TH 12).

These features of public information guarantee that it licenses nested reason to believe of arbitrary (finite and transfinite) degree.

5 Further Postulates.

There are several postulates which might be considered for a more comprehensive analysis of the notion of having reason to believe. One possibility that comes to mind concerns accumulation of reason to believe:

$$(7) \quad \vdash (R_x p \wedge R_x q) \rightarrow R_x (p \wedge q)$$

However⁴, it seems—given the evidence available—quite possible for someone like Holmes to have reason to believe that the butler is the murderer and to have reason to believe that the gardener is the murderer without having reason to believe that the butler is the gardener.

In addition (or independently), one might want to rule out reason to believe inconsistent propositions by assuming

$$(8) \quad \vdash R_x \neg p \rightarrow \neg R_x p.$$

According to our background theory: $\vdash (\alpha \wedge \neg \alpha) \rightarrow \neg(\alpha \wedge \neg \alpha)$ and $\vdash (\alpha \rightarrow \neg \alpha) \rightarrow \neg \alpha$. Hence—with (8) and (P1)—we get $\vdash \neg R_x (p \wedge \neg p)$. This again, seems to be very strong. It would preclude inconsistencies *among* or conjunctive accumulation of pieces of public information⁵. (Note that we have $\vdash (P_{GP} \wedge P_{G\neg p}) \rightarrow P_G (p \wedge \neg p)$ (P 3') and $\vdash P_G (p \wedge \neg p) \rightarrow R_G^1 (p \wedge \neg p)$ (TH 6)).

Another postulate that suggests itself is:

$$(9) \quad \vdash R_x (p \rightarrow q) \rightarrow (R_x p \rightarrow R_x q)$$

I have nothing to say against (9)—²except, perhaps, that I do not need it in this paper. (Below, I will make use of a weaker variant (necessary condition) of (9), namely (P4).

The following four potential postulates are interesting because they allow to simplify our treatment of the successor function (cf. DEF 7, 8):

⁴As Peter Kühnlein has pointed out to me.

⁵As Barry Smith has pointed out to me.

- (10) $\vdash R_x(p \supset q) \rightarrow (R_x p \supset R_x q)$
(11) $\vdash \forall_x R_y \alpha \rightarrow R_y \forall_x \alpha$
(12) $\vdash x \in G \rightarrow (y \in G \supset R_y x \in G)$
(13) $\vdash x \notin G \rightarrow (y \in G \supset R_y x \notin G)$

Note that the converse of (11) follows from (P1). What (12) and (13) claim is (roughly) that a person's membership in G (non-membership in G) is sufficient reason for any member of G to recognize her as a member of G (non-member of G). This, of course, is a very strong assumption about group membership.

Given (10) and (12), we can prove $\vdash R_G^1 \forall_{x \in GP} p \rightarrow \forall_{x \in G} R_G^1 p^6$; *de dicto* nestings turn out to imply *de re* nestings, and we can define $R_G^{d+1} p$ simply by $R_G^d p \wedge R_G^1 \forall_{x \in G} R_G^d p$ (for $d > 0$). Given (11) and (13), we can prove $\vdash \forall_{x \in G} R_G^1 p \rightarrow R_G^1 \forall_{x \in GP} p^7$; *de re* nestings turn out to imply *de dicto* nestings, and we can define $R_G^{d+1} p$ simply by $R_G^d p \wedge \forall_{x \in G} R_G^1 R_G^d p$ (for $d > 0$).

An interesting desideratum concerning our notion of public information is the equivalence of P_{GP} and $P_G P_{GP}$. Note that (TH 5) already gives us one half of this equivalence. To derive the other direction, we first generalize (P3') for arbitrary (also infinite) conjunction:

$$P3: \quad \vdash \forall p(\alpha \rightarrow P_{GP} p) \rightarrow P_G \forall p(\alpha \rightarrow p)$$

P3' turns out to be a special case:

$$P3': \quad \vdash P_{GP} p \wedge P_{Gq} \rightarrow P_G(p \wedge q)$$

Proof: Assume $P_{GP} p \wedge P_{Gq}$. Since $(P_{GP} p \wedge P_{Gq}) \rightarrow (((p \rightarrow r) \vee (q \rightarrow r)) \rightarrow P_G r)$ (TH 1), $\forall r(((p \rightarrow r) \vee (q \rightarrow r)) \rightarrow P_G r)$; hence $P_G \forall r(((p \rightarrow r) \vee (q \rightarrow r)) \rightarrow r)$ (P3). Since $\forall r(((p \rightarrow r) \vee (q \rightarrow r)) \rightarrow r) \rightarrow (p \wedge q)$, $P_G(p \wedge q)$ (TH 1). Qed.

We can prove:

$$TH 13: \quad \vdash P_G(P_{GP} \rightarrow p)$$

Proof: $\vdash \forall p(P_{GP} \rightarrow P_{GP})$. Hence $\vdash P_G \forall p(P_{GP} \rightarrow p)$ (P3) and $\vdash P_G(P_{GP} \rightarrow p)$ (TH 1). Qed.

Now, we have two options. According to the first option we content ourselves with the weak sense of (material) equivalence of P_{GP} and $P_G P_{GP}$: We add Ackermann's rule (γ) to our postulates P1–P3⁸, or equivalently:

$$\Gamma \quad \text{If } \vdash \alpha \supset \beta \text{ and } \vdash \alpha, \text{ then } \vdash \beta.$$

In view of TH 5, it is sufficient to show

$$TH 14: \quad \vdash P_G P_{GP} \supset P_{GP}$$

Proof: Assume $P_G(P_{GP} \rightarrow p) \wedge P_G P_{GP}$. Hence $P_G(P_{GP} \wedge (P_{GP} \rightarrow p))$ (TH 1, P3') and P_{GP} (TH 1). Hence $\vdash (P_G(P_{GP} \rightarrow p) \wedge P_G P_{GP}) \rightarrow P_{GP}$ and $\vdash P_G(P_{GP} \rightarrow p) \supset (P_G P_{GP} \supset P_{GP})$. Hence $\vdash P_G P_{GP} \supset P_{GP}$ (TH13, Γ). Qed.

The other option is to assume:

$$P4: \quad \vdash P_G(p \rightarrow q) \rightarrow (P_{GP} \rightarrow P_{Gq})$$

This leads to the following strengthening of TH 14, hence to the equivalence of $P_G P_{GP}$ and P_{GP} in the strong sense of relevance logic (\leftrightarrow):

$$TH 15: \quad \vdash P_G P_{GP} \rightarrow P_{GP}$$

Proof: $\vdash P_G(P_{GP} \rightarrow p)$ (TH 13), hence $\vdash P_G P_{GP} \rightarrow P_{GP}$ (P4). Qed.

⁶Proof: Assume $R_G^1 \forall_{x \in GP} p$. Hence, $y \in G \supset R_y(x \in G \supset p)$ (P1). Hence, $(y \in G \supset R_y x \in G) \supset (y \in G \supset R_y p)$ (10). Hence, $x \in G \supset (y \in G \supset R_y p)$ (12). Hence, $R_G^1 \forall_{x \in GP} p \rightarrow \forall_{x \in G} R_G^1 p$ (DEF 5). Qed.

⁷Proof: Assume $\neg R_G^1 \forall_{x \in GP} p$. Hence, $\exists y \in G \exists x \neg R_y(x \notin G \vee p)$ (11) and $\exists y \in G \exists x(\neg R_y x \in G \wedge \neg R_y p)$ (P1). Hence, $\exists y, x \in G \neg R_y p$ (13). Hence $\forall_{x \in G} R_G^1 p \rightarrow R_G^1 \forall_{x \in GP} p$ (DEF 5). Qed.

⁸It is well known that rule (γ) (hence Γ) is admissible in (quantified) relevance logic (cf. Anderson & Belnap (1992), §42). But, of course, this does not mean that (γ) remains admissible (redundant) for QR + P1–P3.

I am closing this exposition by asking two questions: Which (concise and intuitively plausible) assumptions formulated exclusively in terms of the Lewisian primitives R , H and G imply P2–P4 in the framework of QR? What kind of modal structure do our assumptions induce for the notion of public information. (TH 5 and TH 14/15 are reminiscent of S4, P3' and P4 of K. The left-to-right-direction of P2 looks like the Barcan Formula restricted to G . Finally, there is some similarity between P1 and Lemmon's 'if $\vdash \alpha \supset \beta$, then $\vdash L\alpha \supset L\beta$ '⁹. But we do not (and do not want to) have 'if $\vdash \alpha$, then $\vdash P_G\alpha$ ', ' $\vdash P_G\alpha \supset \alpha$ ', or ' $\vdash P_G\neg p \supset \neg P_G p$ '.

References

- Anderson, A.R. and N.D. Belnap. 1973. *Entailment. The Logic of Relevance and Necessity.*, volume I. Princeton University Press.
- Anderson, A.R. and N.D. Belnap. 1992. *Entailment. The Logic of Relevance and Necessity.*, volume II. Princeton University Press.
- Heydrich, W. 1995. *Relevanzlogik und Situationssemantik.* de Gruyter.
- Hughes, G.E. and M.J. Cresswell. 1968. *A New Introduction to Modal Logic.* Routledge.
- Lewis, D.K. 1969. *Convention—A Philosophical Study.* Harvard UP.
- Lewis, D.K. 1973. *Counterfactuals.* Harvard University Press.

⁹Cf. Hughes and Cresswell (1996)

A DRT-Style Modelling of Agents' Mental States in Construction Dialogue

Wolfgang Heydrich and Peter Kühnlein and Hannes Rieser
Collaborative Research Center SFB 360 Situated Artificial Agents

Abstract

Elaborating and extending the well-known DRT-formalism, Hans Kamp has proposed a format of articulated DRSs in order to represent mental states (Kamp, 1990). Assuming acquaintance with these proposals, we shortly discuss some aspects of their philosophical background and argue that they provide a promising starting-point for the systematic investigation of dialogues. In the main part of this paper, we apply Kamp's method to a sample turn exchange from a construction dialogue (from the Bielefeld SFB setting).

1 Kamp's Perspective: Mental States in Dialogues

Agents develop beliefs, assumptions and other mental attitudes with respect to situations. An agent seeing an object before him will, e.g., believe that this object is before him. His belief, in turn, may form a presupposition of an assertive speech act, being about this very object. In some cases, a belief in the presence of this object, may be a cause of a speech act directed towards another agent in the situation, e.g., if the first agent wants the second one to pick up the object and perhaps to carry it away. Now this intention of the first agent's will, as a rule, end up in an utterance, such as: "Please, take this object away", to which the second agent may respond: "Yes", if he is willing to take it away. Situations like these seem to be part and parcel of our daily undertakings. Hence, serious description of natural dialogue seems to necessitate that we consider the mental states of the speakers involved.

However, describing dialogues from the perspective of the mental states involved is no easy task. Paradigms of dynamic semantics, DRT, Situation Semantics and Dynamic Predicate Logic, first class candidates for modelling semantic or pragmatic processes, have three obstacles to overcome in order to deal with dialogues.

- The first one is the problem of adapting the respective paradigm to the description of dialogue. Surely, the paradigms have been invented to describe "monological" discourse without reference to speakers or utterance tokens. If you want to handle discourse, you have to describe the contributions of the participants as well as how these depend on previous contributions of the speakers involved.
- The second problem is the description of mental states. Even if matters of representation and interpretation were discussed in approaches to dynamic semantics (see e.g. Kamp & Reyle 1993' *Preface* and *Preliminaries*) and Kamp 1981), none of them was especially designed for capturing mental states, beliefs and the like, and *a fortiori* the dependence of an agent's mental state on the history of both his own mental states and the mental states of his co-participants.
- The third obstacle is to explain how mental states are related to overt behaviour, which is bound up with questions about the generation of mental states, the temporal and causal ordering among them etc.

An approach which tries to overcome these obstacles, especially obstacle two, is Hans Kamp's seminal paper *Prolegomena to a Structural Account of Belief and Other Attitudes* (Kamp, 1990) (hereafter abbr. as *Prolegomena*). Following Kamp's suggestion, we utilize extended DRT-tools to represent psychological (attitudinal) states and describe the exchange of two turns by two agents in a task-oriented dialogue. Our aim is to investigate the scope of the descriptonal apparatus developed by Hans Kamp.

2 What Are Structural Accounts Good For?

Before doing so, we pause and ask why Kamp could be interested in developing a semantics of psychological states. Which overall philosophical positions are confronted by his new account?

(a) There is a controversial discussion of how mental states should be individuated. Its starting point is Putnam (1975)'s thought experiment forming the core of an argument for so-called "wide content" of mental states.

The upshot of this experiment is that molecularly identical twins living on earth and twin-earth (which is an exact replica of earth except that there is XYZ wherever H₂O is on earth) do not have identical contents of mental states despite their being in identical physical conditions: The contents of *mental* states do not solely depend on intrinsic physical states. Mental states have to be individuated on the basis of environments, too. This is the core claim of every *anti-individualist* position.¹

There are opponents to this view, of course, not the least known being Fodor (1991). He argues that Putnam's twins exhibit identical behaviour and (behaviour being caused by mental states) can be in identical mental states. In effect, he defends the view that mental states have to be individuated by "narrow" content.²

If Kamp succeeds in giving a primarily *structural* description of attitudes, then it is possible both to give sense to the claim that two individuals have identical mental states independently of differences of the latter's truth conditions and to the claim, that they have different mental states despite the identity of their truth conditions. The whole problem boils down to that of one's being able to give a structural description according to different criteria of equivalence.

The view on the identity of mental states is no longer a question of *which states should be counted as identical* but more a question of *which criteria for the selection of an equivalence class are appropriate*.

Note that representations of mental state tokens are mapped onto truth-values by way of instantiation of parameters. Taking into account varying ways of anchoring of one and the same structure leaves us with wide content. Abstracting away from anchoring provides narrow content. This, at least, seems to be Kamp's way of defending some structural account of mental states against both parties in the discussion. In other words: Kamp is *neutral* towards both the individualist and the anti-individualist position.

(b) It is Kamp's position that no successful description of dialogues can avail itself of mental states. But *what* is explained with reference to them?

In the following passage we sketch the rôle of Kamp's *articulated* DRSs in the explanation of action. We relate these DRSs to causes or reasons usually postulated in action theory.

Most theories of action identify reasons for doing something with psychological states of the agents. *Belief-desire*-explanations, as attempts to explain action by reference to mental states are called, seem to be widely accepted among theorists, despite weaknesses that hedging by *ceteris paribus*-conditions reveals.

¹Burge (1979) has advanced other thought experiments to the same effect, thereby broadening the scope of *anti-individualist* arguments to include more of the social dimension of thought.

²It is important not to confuse the narrow content account with a claim of true solipsism. What a defender of a narrow content account has to claim is not that there are no relations from mental states to the external world but only that those relations should not count as relevant for the *type*-individuation of states.

Of course, there are several problems connected with every account of action that refers to mental states or events: Firstly, prototypical action consists, at least in part, in motion; clearly, *motion* is a purely physical notion. Hence, it is necessary to explain how mental events effect physical events. Secondly, one may hold that mental states are not regularly to be considered as reasons for doing something and this view seems compatible with the claim that mental states are causally relevant for behaviour.³ What makes this account attractive are the following three facts: First, it appears to support the view that the true relevance of mental states for action depends on their *causal* explanatory status. Secondly, it is an *externalist* position towards reasons for action in quite the same way as Lewis' (1969) account is externalistic towards reasons to believe: Reasons are in the world. And thirdly, it allows talk of acting *without reason* in cases in which reality does not support the reasoning done by the agent.

Kamp, however, seems to underwrite some variant of *belief-desire*-theory of action. Additionally he even states that he would try to formulate a working foundation for some theory of practical reasoning to be developed.

Anyway, Kamp does not seem to claim that action can be explained only with reference to an underlying process of practical reasoning. On the contrary, he only discusses cases in which it is not possible to explain differences in behaviour solely from considerations about the factual circumstances.

Hence, it is possible to understand the descriptions of mental states given by the structural account as descriptions of *causes* for action.

Once again, it seems that the Kamp-style structural account is neutral as to the discussion about the rôle of mental states and can be reformulated in either ways.

3 Applying the Method

The Experimental Setting

There was a number of (psycho-)linguistic experiments performed at the SFB360, "Situating Artificial Agents", University of Bielefeld. The experimental conditions were as follows: Two subjects had to cooperate to build complex objects out of wooden blocks; hence, the series of experiments is called *blocks world experiments*. One of the subjects (called Instructor throughout this paper) was asked to give instructions to the other (Constructor), thereby bringing Const into a position to build some such object. While Inst saw a ready-built object, Const, who was separated from Inst by a screen, had no other information available than the instructions uttered by Inst. The dialogues between Inst and Const in this setting are called *construction dialogues*. A schematic depiction of the setting is given in figure 1.

The sessions were filmed on video and transcriptions were made. One of the construction dialogues begins in the following way: Inst utters sentence (1), Const responds with the utterance of (2) as follows:

- (1) Also unten steht ein langgezogenes grünes Rechteck hochkant
Well on ground stands a longish green rectangle upright
There is a longish green rectangle standing upright on the ground
- (2) Hochkant, ja
Upright, yes
Upright, yes

³An example for behaviour which we would intuitively not claim to be explainable primarily by reference to mental states could be: *S* drives down an alley and arrives at a red traffic light. The reason why *S* stops, then, is not that she has the belief that the traffic light is red, but that the traffic light is red. Consider a corresponding case, where *S* stops because she believes the traffic light to be red, while it really is green, causes an accident and is judged guilty because she stops *without reason*. What one believes sometimes doesn't seem to matter, in this respect.

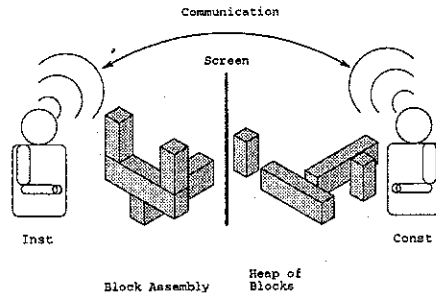


Figure 1: Setting for the blocks worlds experiments

Utterances (1) and (2) together form our turn exchange in question. (With a weakening explained in subsection 2.) What we will do in subsection 2 is give a model of the respective mental states of Inst's and Const's while they are engaged in this turn exchange.

Discourse and the Flow of Information

In order to integrate a Kamp-style analysis of mental states of agents into the systematic investigation of dialogues it seems to be appropriate to start with some explanatory schemata designed to capture relevant features and procedural patterns of interaction and communication.

We think about dialogue as a cooperative undertaking of agents engaged in developing and transforming their common situation, as a kind of interaction which combines *communicative* and *constructional moves*. Our agents bear two kinds of relations to the stages of their developing situation. They may either *pick up information* (flow of information from situation to agent) or *act* (flow of information from agent to situation). Forms of picking up information are *perception* (PERC) and *control* (CONT). *Acting* may be *verbal* (VERB) or *constructional* (CONS). We analyse controlled action as a pair of action and control. A constructional move of controlled action may be represented graphically like this:

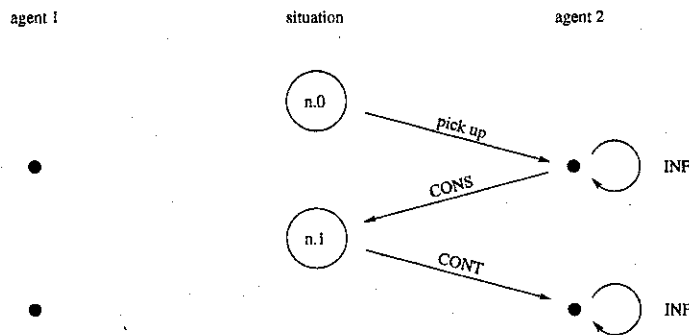


Figure 2: Constructional step by agent 2

Here, $n.0$ and $n.1$ represent stages of the developing situation. The dots on the right indicate an agent. The circled arrows symbolize internal inferential processes (INF) of update.

A characteristic feature of our experimental setting is the separating screen between the participants. It prevents the agents from perceiving their respective coagents manipulating (or even checking) the blocks on their side of the screen. Hence, constructional moves are not open for information pick up in an unrestricted way. In general, we distinguish stages of the developing situation which are *freely accessible*, i.e. open for information pick up by every participant from

those which are not. Situation stage $n+1$ in figure 2 is not freely accessible since it is open for information pick up by agent 2, but not by agent 1. (No arrow leads from stage $n+1$ to the dot on the left in figure 2.)

Verbal contributions to the informational supply of a situation are open for information pick up. Very often, however, they require verbal feedback, which functions as some sort of ratification. Only after such a ratification the agents may take it for granted that the original contribution has become shared information among them. We assume that the communicative moves in our setting consist of verbal contributions by one of the agents which are, at least by default, manifestly followed by certain verbal reactions (feedback) of the coagent.

Communicative moves involve both agents. In contrast to the constructional moves along the lines of figure 2, they constitute cooperative activity. But, of course, the rôles of agent and coagent are quite different.

Our characterization of communicative moves is largely neutral with respect to matters of function, intention or illocutionary force. One purpose, which is important in our setting, is *reporting* the result of a constructional step. A related intention consists in just getting some descriptive information across. Of course, such a purpose might be coupled with additional intentions, e.g. the intention of convincing the addressee that the configuration's having feature so-and-so is the result of some action that had been required before and has just been carried out successfully. We are not dealing with specific intentional profiles of this sort in this paper. Instead, we restrict ourselves to the communicative transfer of descriptive information.

A combination of several moves of verbal and constructional action gives us a schema of the cooperative pattern of *requiring, accomplishing and reporting* constructional steps:

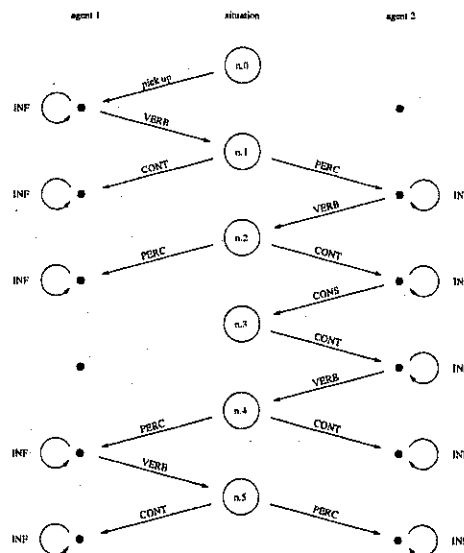


Figure 3: Discourse Model: Requiring, accomplishing, and reporting constructional moves

This schema provides a basic pattern for construction dialogues. According to it, agent 1 (= Inst) and agent 2 (= Const) may be thought of as organizing their interaction in cycles. Each cycle starts with Inst's picking up some feature in the configuration of blocks (on her side of the screen) and it ends with it being shared information among Inst and Const that the feature in question is realized by the configuration of blocks on both sides of the screen.

Construction dialogue

In the reconstruction of the dialogue we make some strong idealizations. To begin with, natural language data in general as well as the transcripts of the blocks world experiments show that agents often do not wait until their co-agents stop speaking before they start to react. Rather, they anticipate what the other will say and respond accordingly.

Consequently, there are no sharply defined points in time at which a situation can be said to be established. Metaphorically speaking, there is a "grey area" over which the situation is "smeared". We take it that this is due to incremental processing that takes place, but have to offer no account of such processing. What we suppose is that, if such an account were given, it would have to render points in time such that they could be paired, e.g. functionally, with our idealized situations.

Another consequence is, that we may not take the existence of separate processes of action and monitoring for granted. Repairs and self-repairs during the time of utterance can only be explained if one assumes that both agents track their behaviour permanently. This means that there should be another "grey area" in our scheme: One that ranges over arrows connecting mental states with situations and mental states vertically.

Additionally, we assume *tacit default* back-channelling behaviour on the side of our agents. This is suggested by our account of the dialogue structure. Note that "Upright, yes" (cf. sentence (2)) can be interpreted in at least two ways. We can take it as either (a) the constructor's indication of his being attentive or (b) as his reply meant as indication of his action carried out. Perhaps (2) is underspecified with respect to both these readings.

A final point is that we leave out the problem of metonymy here. Clearly, the expression *rectangle* used in (1) refers to a block in the setting. As we do not intend to give an account of metonymically used expressions here, we simply substitute *rectangle* by *block* and thereby get the utterance

- (3) Also unten steht ein langgezogener grüner Block hochkant
Well on ground stands a longish green block upright
There is a longish green block standing upright on the ground

Utterance (3) will be the utterance which, in combination with (2), really is our linguistic basis.

Inst's mental state immediately before her first utterance

The first relevant state worthwhile to model is Inst's state before she utters the first sentence in the dialogue. In this situation, it is assumed, Inst has already perceived the objects in front of her, analyzed the relations between them and, in accordance with her plan, chosen one as a starting point of her description based on some focusing procedure not discussed here, but see Clermont et al. (1995).

The DRS reflecting (the relevant part of) Inst's mental state looks like this: To avoid ambiguity between the usage of *Bel* as a mode indicator and as an operator (in $\langle Bel, K \rangle$ and $Bel(a, p)$), we decided to simply use *B* as mode indicator for a belief state. Anyway, it should be clear what is meant when we write $\langle B, K \rangle$.

The discourse referents and conditions in this DRS are read as follows: *h* (for *here*) and *n* (for *now*) are indexical discourse referents, pointing to the place and time of the respective speaker. There is no hint of Kamp's as to the use of *h*, but its construction is analogous to that of his *n*. The indexical discourse referents guarantee that the mental state described by the DRS indeed refers to (parts of) Inst's own situation. Note that this does not mean that we are dealing with *de se*-attitudes here, although we will soon have such cases. What is achieved is that the DRS is, so to speak, parameterized by Inst's spatio-temporal location. This account of *n* and *h* also underlies the inclusion in condition $s_1 \subseteq h$, which is intended to mean that s_1 is located at Inst's own location; cf. Kamp & Reyle (1993), p. 511. For an explanation of s_1 see below.

Discourse referent *x* is taken to be anchored to the block under consideration by Inst; this is expressed by the notation $\langle x, x \rangle$ outside the DRS. It is especially important to note the construction of the duplex condition in this DRS as well as the solution for rendering the VP construction.

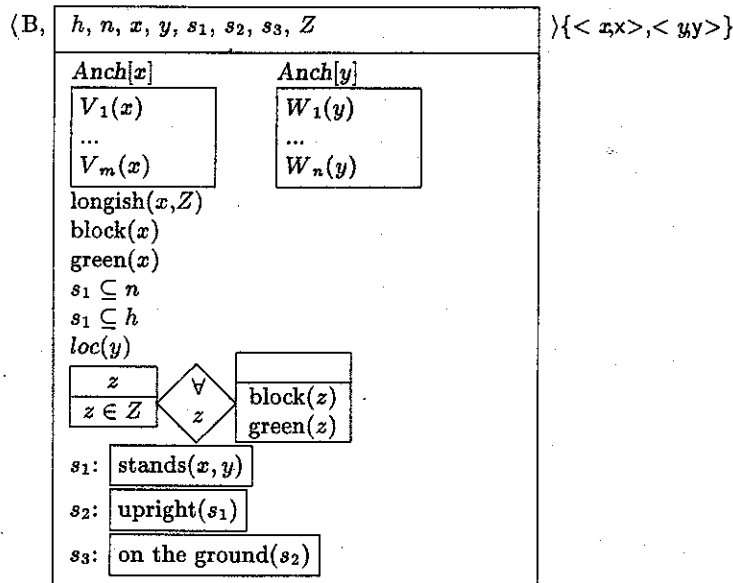


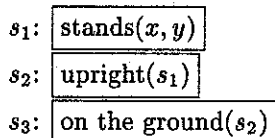
Figure 4: First Inst's DRS

The duplex condition



together with the entry *longish*(x, Z) has to be taken as saying that x denotes a block which is longish with respect to a set of other green blocks. Here, Z is introduced to refer to that set of blocks. (But, note, not to the set of *all* green blocks.) There are, of course, different analyses possible for the expression "longish".

The VP construction *on ground stands ... upright* is given the following treatment: Clearly, object x , the block, is standing at some location y . Its standing there is a situation described by s_1 : stands(x, y). This information can be made more precise by adding to it that it is an upright-standing-situation, which in turn is expressed by s_2 : upright(s_1). Still, we do not have complete information. Hence, we may add that this standing upright happens on the ground, i.e. s_3 : on the ground(s_2). As Kamp does not give any clue concerning the question of handling constructions like these, we chose to give a fine-grained analysis here which is expressed by



(Note that y is used to refer to the location of the block, which is expressed by an external anchor. The attitudinal state of Inst's is *truly de re*, therefore, as far as the block and the location are concerned⁴.)

Given the interpretation of $s_1 \subseteq h$, it is obviously the case that y is part of Inst's own situation.

⁴This should explain the use of the remaining discourse referents, namely s_1, s_2 , and s_3 : They are intended to introduce Inst's situation. It is not completely satisfactory to choose this notation, as nothing of exactly this kind is provided by Kamp himself. But there are striking parallels in (Kamp & Reyle, 1993, p.511), for the present case, and as the problem of giving a DRT-analysis of VPs is not at the core of our investigation and should be treated separately, we adopted this notation.

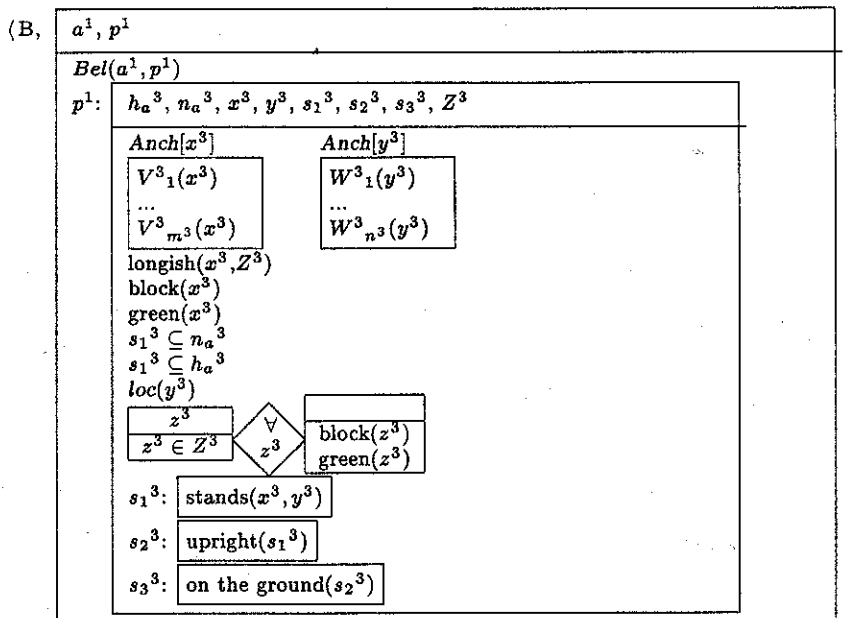


Figure 5: First version of first Const's DRS

Const's mental state immediately after Inst's first utterance

We now turn to Const's mental state caused by Inst's utterance. Inst's directive must have been based on the blockworld-situation on Inst's side, which Const has to arrange in a parallel fashion on his side. Both of them are fully aware of this fact. Hence, Const can build up a belief about a belief of Inst's concerning the information Inst's situation yields to him.

Now it seems to us that we encounter a particular difficulty, if we want to fix how Inst's belief, embedded in a belief of Const's, should be expressed. Const can proceed in either of two ways. He can attribute to Inst a belief in roughly the following way: "That guy Inst over there has the following belief: *Here, on my side, now stands a green longish block upright on the ground*". Clearly, this belief is derived from Inst's utterance, and moreover, the content of the belief is seen from the perspective of Inst's, very much so as if Const were in Inst's position, sitting on the other side etc.. A different, albeit equally reasonable method to fix Inst's belief as seen by Const is the following: "That guy over there has the belief that *just before on his side there stood a green longish block upright on the ground*". The first version in quotational form has occurrences of the indexicals *here, on my side* and *now*. Since it seems to be tied to Inst's self, we call it the *indexical-reading* of Inst's belief embedded in Const's belief. In order to represent co-temporality in Const's embedded belief we therefore use *n* (now), which is, however, Inst's *n*, and *h* (here) for Inst's location. a^1 serves as a discourse referent for "the guy over there". We do not use *i* in this case, since this is reserved to first-person reference. Following Kamp's suggestions, *i, n*, and *h* come with their own (implicit) formal conditions. Hence, what emerges (see figure 5) is that the anchoring of *n* and *h* must be wholly left to Inst. Const cannot care about that.

A second reading could use demonstrative devices instead, *t* for "there" and *b* for "just before". It would be descriptive and pose no special problems.

Const's mental state immediately before the construction step

After Inst's utterance Const, according to figure 3 above, could have said something, referring back to Inst's utterance. This would have changed the situation generated by Inst's utterance to a next situation to be in turn perceived by Inst who could have changed it by verbal means

etc. However, this does not happen. Hence, we may safely assume that Const thinks that the situation is under control and does not require any additional effort from his side. So Const will establish a complex attitudinal state comprising at least the following:

- (1) the content of what he believes to be the case on Inst's side, which must be formulated deictically,
- (2) a belief he attributes to Inst concerning Inst's situation, formulated indexically, and
- (3) an indication of his commitment that his belief and Inst's belief concern the same subject matter, provided via internal links

(see figure 6). Two points require some debate. The first concerning (b) is whether we better should have ascribed the belief about the situation on Inst's side to Const already in the last section. However, as explained in subsection 2, there is no principled way to decide on the basis of our model and the discourse data *when* the respective inferences are drawn: This was just the reason why we spoke about "grey areas". In any case, the belief must be fixed before the *Shared-predicate* is introduced. The second point is whether we have to assume

- (4) conditions involving a *Shared-predicate* in Inst's as well as in Const's attitudinal state:

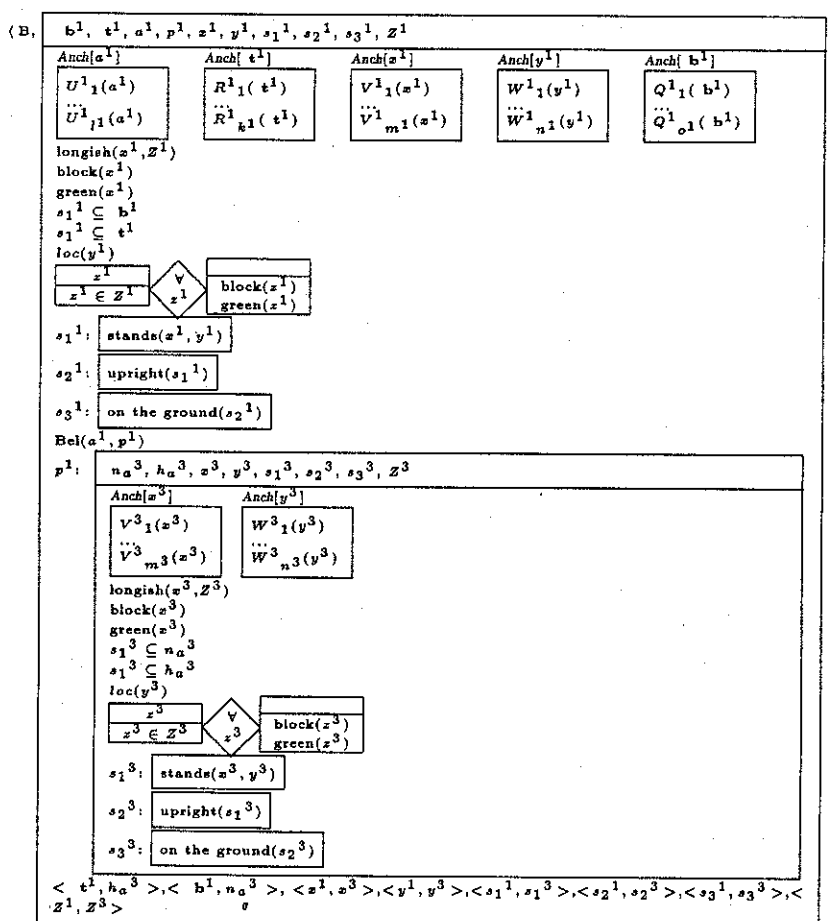


Figure 6: First version of Const's mental state after Inst's utterance

In figure 6 we use the deictic variant of mental state representation in order to capture (a). Remember that **b** is used as a discourse referent for "just before". The introduction of discourse

referents is as follows: In Const's belief concerning Inst's side we have anchors concerning Inst (*Anch*[a]), the time interval *t* (*Anch*[t]) and place *b* (*Anch*[b]) involved as well as for the objects *x* and *y*. In Const's belief attributed to Inst we have indexical discourse referents for Inst's situation location h_a^3 , which is different from Const's (t^1), and anchors for the object *x* and *x*'s location *y* within h_a^3 . The anchors depend on the introduction of these entities by Inst. a^1 refers deictically to Inst, t^1 to his location and similarly for the other discourse referents. Due to the screen, there is no direct relation between Const and the entities on Inst's side.

Finally, (c) is brought about by linking the content of Const's belief and his assumed content of Inst's belief internally as follows:

$\langle t^1, h_a^3 \rangle, \langle b^1, n_a^3 \rangle, \langle x^1, x^3 \rangle, \langle y^1, y^3 \rangle, \langle s_1^1, s_1^3 \rangle, \langle s_2^1, s_2^3 \rangle, \langle s_3^1, s_3^3 \rangle, \langle Z^1, Z^3 \rangle$

The question pertaining to (d) is whether we need more than the things indicated in figure 6 on the side of Const in order to explain why he acts, thereby changing his situation. We now turn to this problem.

Stressing his own contribution to the mutuality topic, Kamp (*Prolegomena* 5.2, p. 79) sets forth the following arguments in favour of a tighter connection between a speaker's and his addressee's mental states:

- (i) *Not only do some of B's [the recipient's videlicet the constructor's] attitudes become dependent on those of A, some of A's become conversely dependent on attitudes of B* (*Prolegomena*, p. 78).
- (ii) [Thus] *verification of A's attitudes cannot be detached from verification of those of B nor vice versa* (*Prolegomena*, p. 79).
- (iii) *What is crucial for sharing of discourse referents is that the participants take themselves to be sharing those discourse referents* (*Prolegomena*, p. 81).
- (iv) There cannot be established external links between agents' mental states without matching shared attitudinal contents.

Considering (i) we can say that although Const's complex attitudinal state depends on Inst's, the converse is not true, since Inst must develop his belief without in any way depending on Const's. The main source of Inst's attitudinal state is his perception of his own situation. Also, concerning (ii) we observe that Inst can verify his attitude without considering Const's. However, there is the following argument, which perhaps provides some reason for introducing a *Shared*-predicate in order to characterize Inst's as well as Const's mental state. In the empirical data, Const does not react verbally to Inst's directive by e.g. way of a repair or a side sequence⁵. We take this to indicate that Inst and Const believe their attitudes to converge on the same objects. This motivates a *Shared*-predicate on the side of both participants. Hence we extend the attitudinal state represented in figure 6 concerning Const by the *Shared*-predicate and the corresponding list of internal links below (Fig. 7).

$\langle t^1, h_a^5 \rangle, \langle b^1, n_a^5 \rangle, \langle x^1, x^5 \rangle, \langle y^1, y^5 \rangle, \langle s_1^1, s_1^5 \rangle, \langle s_2^1, s_2^5 \rangle, \langle s_3^1, s_3^5 \rangle, \langle Z^1, Z^5 \rangle$

Figure 7: Internal links, cf. figure 8

The old links remain. In addition, t^1 etc. are linked with the corresponding discourse referents in the *Shared*-predicate.

The corresponding full-fledged attitudinal state of Inst's is not shown here. Due to two *Shared*-predicates we could, however, capture (iii) above, hence external links according to (iv) could also be introduced. From the external point of view we know that Inst and Const have different external anchors for their discourse referents⁶. So the question arises whether we can set external links in these cases. Note that external linking in NO WAY entails external anchoring, so

⁵According to figure 3, however, we assume that he could have issued a repair or a side sequence.

⁶We suppose that internal as well as external linking are both symmetric and transitive and if *x* is internally linked with *y* and *y* is externally linked with *z*, then *x* is externally linked with *z*.

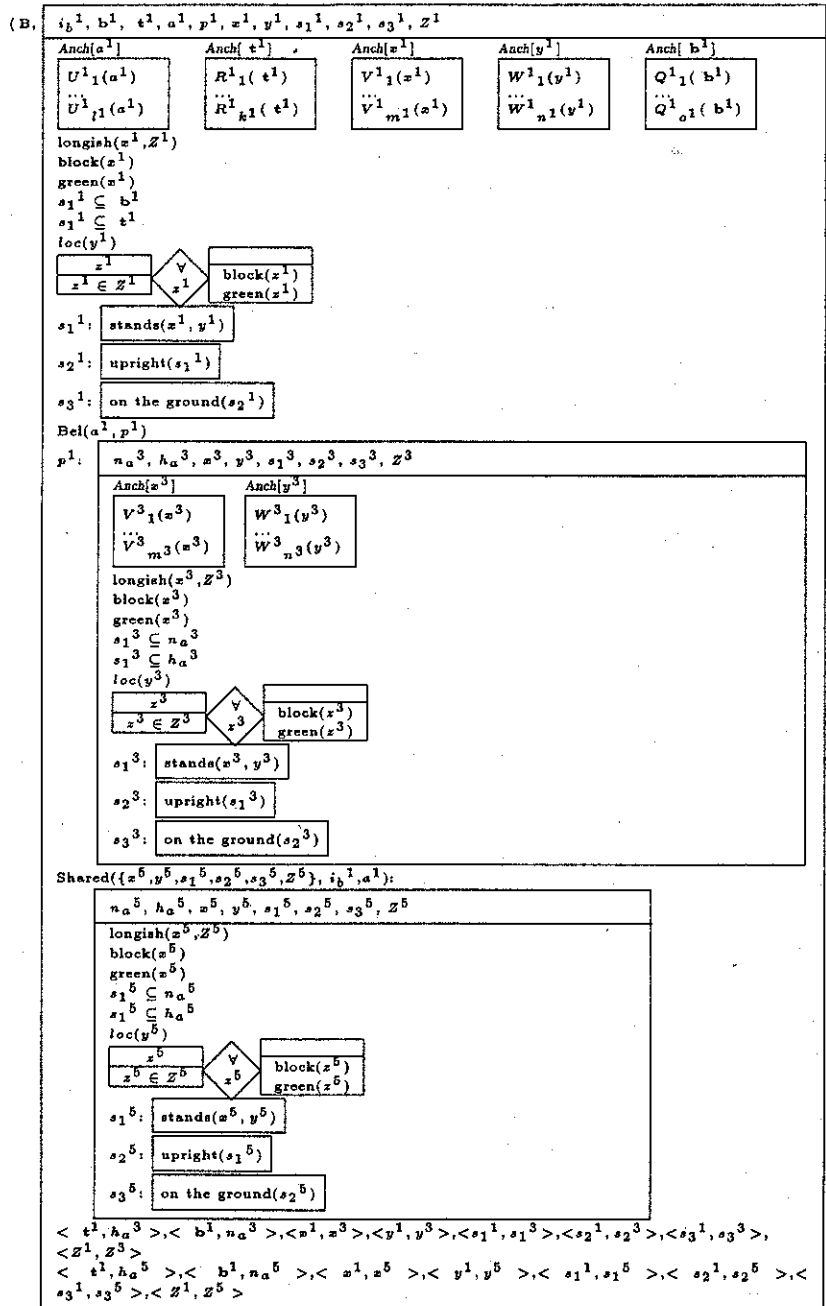


Figure 8: Final version of Const's mental state after Inst's utterance

we can apply external links thereby indicating that the agents think they have developed beliefs about the same entities.

Const's mental state immediately after the construction step

After Const has identified and selected the block, he will have a *de re*-belief concerning this very block. The introduction of discourse referents is much the same here as in the case of Inst's initial perception and ensuing belief (see figure 4). Observe that the discourse referent introducing the

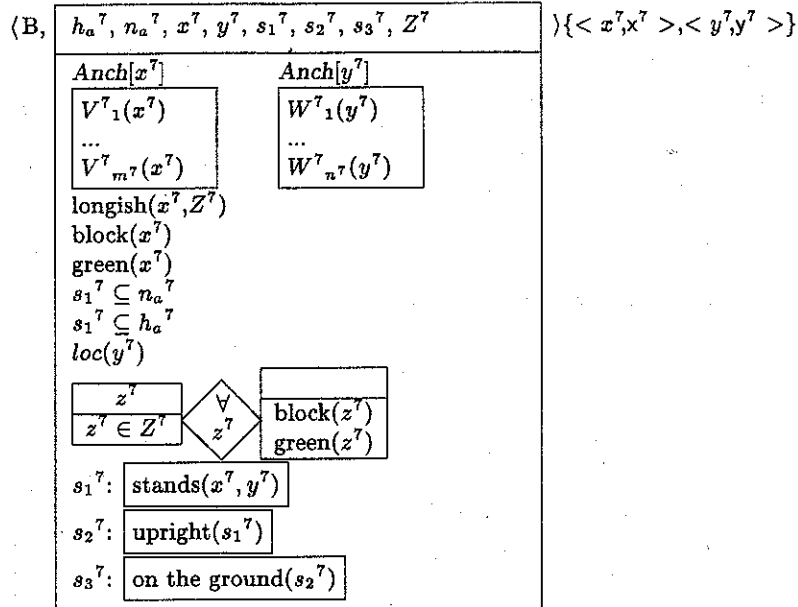


Figure 9: Const's mental state after the construction step

block, x^T , and the discourse referent standing for its location, y^T , both are externally anchored: { < x^T, x^T >, < y^T, y^T > }.

Const's mental state hence seems to be as depicted in figure 9. Still, a problem remains having to do with the succession of mental states, their individuation etc.: Why do we open up a new mental state for Const? If mental states roughly behave as DRS-methodology suggests, Const's old belief as shown in figure 8 provides the very context for the belief acquired after his act. By this reasoning we arrive at the somewhat larger belief represented in figure 10 with the indices of discourse referents appropriately changed, of course. No special problems arise for mental state representation, since we have new objects leading to new discourse referents and no new cases of internal linking.

Inst's mental state immediately after Inst's first utterance

After Const has done what he was expected to do, he replies with *upright, yes*, indicating that he has created a situation similar to the one on Inst's side. Hence, Inst will, due to Const's reply, form a belief about the object on Const's side, furthermore a belief about Const's belief and there also will be a *Shared*-predicate indicating that Inst believes that his belief and the embedded belief concerning Const's belief concern the same things. We leave out a description of this newly acquired mental state of Inst's.

But, of course, as it was the case with the mental state rendered in the DRS of figure 9, the earlier beliefs of Inst's form again the background against which this newly acquired belief is built. So, there could be a complex state integrating the newly acquired belief of Inst's.

In addition, because of Inst's non-intervention, we can establish a symmetric belief on Const's side. As a consequence of both *Shared*-predicates we can then also use external links mediating between Const's belief and Inst's belief.

A puzzle remains

What we have achieved by our procedure so far is: We have internally linked the mental states of each agent involved in the construction business with the mental state he/she attributes to his/her

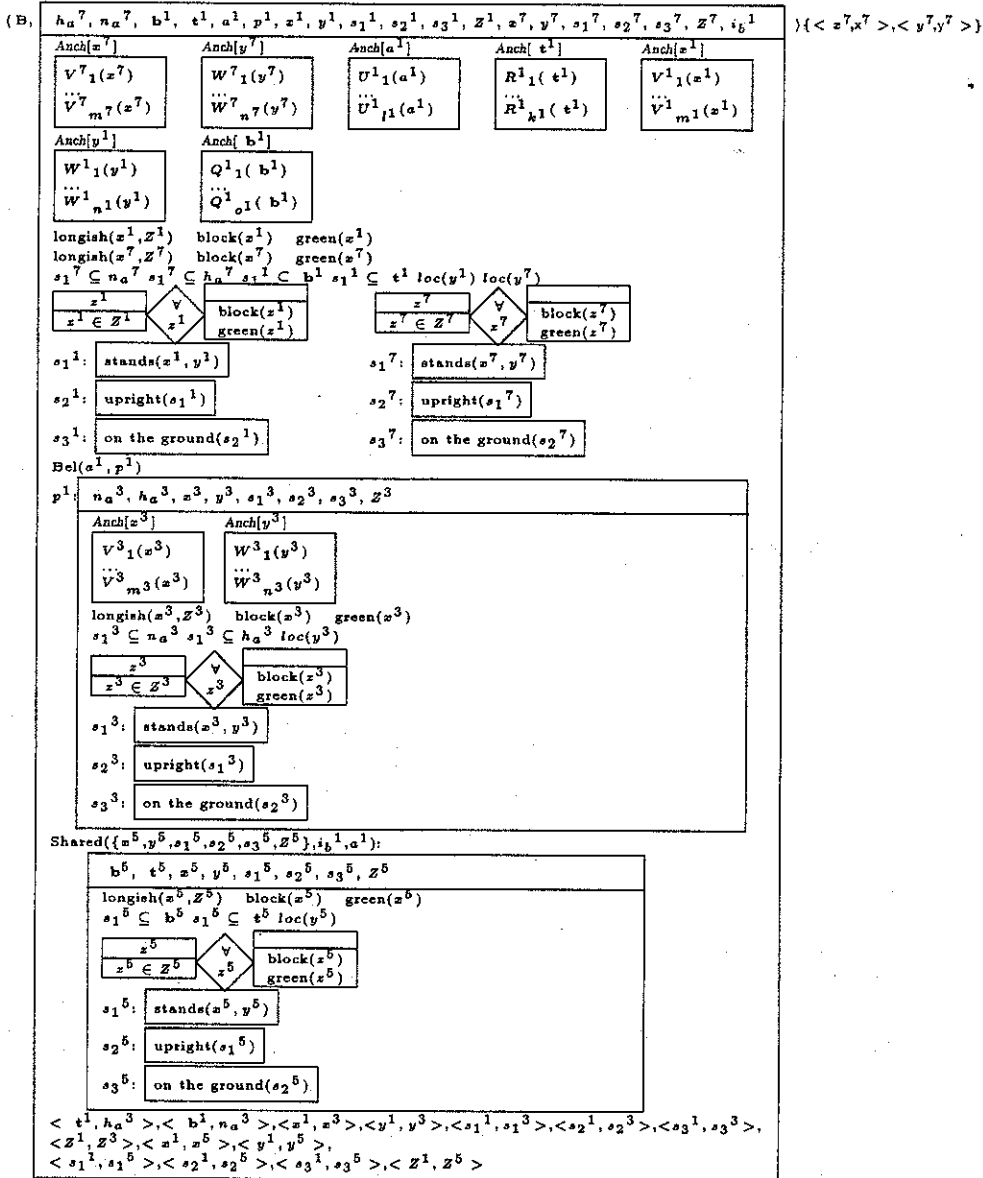


Figure 10: Complex belief of Const's after the construction step

co-agent. However, we have not yet modelled the impression the agents presumably have, namely that they have brought about the "same situation" on both sides of the screen. This seems to hinge rather on the conditions tied up with discourse referents than on the discourse referents themselves.

Although the *Prolegomena* (p. 81) stress the importance of the content shared, there is no systematic means for expressing what happens in the context of our example: There Inst and Const connect different local orientations with the phrase 'standing upright'. For Inst, the green block is standing upright if it is posited on the smaller side which, standardly speaking, is more like lying since the surface touching the ground is longer than the surface in the vertical dimension. For Const standing of the block requires it to be posited on the smallest side, with height extending width. We could represent this situation by introducing discourse referents for properties and anchoring them to different external relations on both sides. In the case of Inst we might have

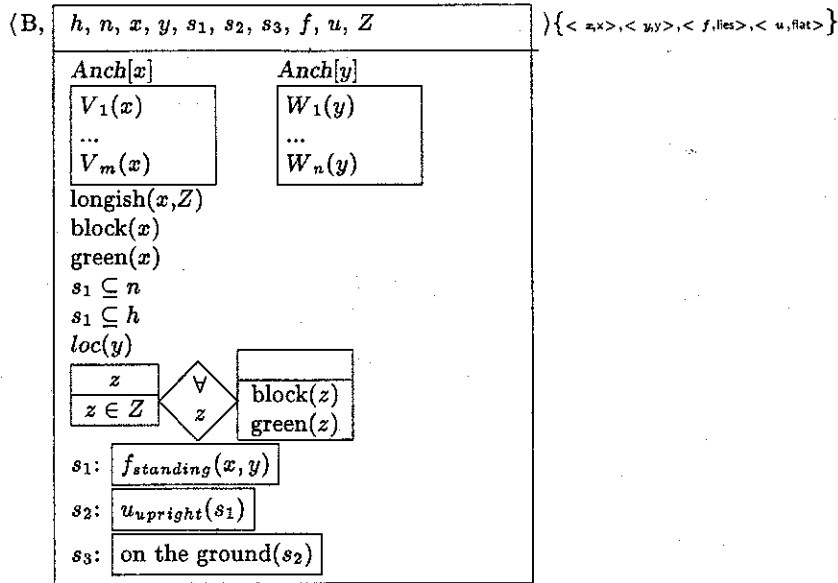


Figure 11: Inst's DRS from figure 4, modified to match the problem of anchoring to properties

for a start: Inst's DRS displayed in figure 11 is paralleled by a DRS on Const's side. Const, however anchors discourse referents and relational parameters differently:
 { < x^i, x^i >, < y^i, y^i >, < $f^i, stands$ >, < $u^i, uprt$ > }

References

- Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, 4:73–121.
- Clermont, T., Meier, C., M., P., Prestin, E., Rieser, H., Ritter, H., and Velichkovsky, B. (1995). Augenbewegung, Fokus und Referenz. Technical Report 95/8, SFB 360.
- Fodor, J. A. (1991). A Modal Argument for Narrow Content. *Journal of Philosophy*, 88:5–26.
- Kamp, H. (1981). A theory of truth and semantic representation. In Jeroen Groenendijk et. al., eds., *Formal methods in the study of language*. Mathematical Centre, Amsterdam.
- Kamp, H. (1990). Prolegomena to a Structural Account of Belief and Other Attitudes. In Anderson, C. A. and Owens, J., editors, *Propositional Attitudes: The Role of Content in Logic, Language and Mind*. CSLI.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic—Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Kluwer.
- Lewis, D. K. (1969). *Convention—A Philosophical Study*. Harvard UP.
- Putnam, H. (1975). *Mind, Language, and Reality: Philosophical Papers*, volume 2, chapter The Meaning of 'Meaning'. Cambridge UP.

Questions Under Discussion and Dialogue Moves*

Staffan Larsson
Research Assistant, PhD Student
Dept. of Linguistics, Göteborg University
sl@ling.gu.se

Abstract

The QUD provides a partially ordered representation of several types of questions which can be addressed at a certain stage of a dialogue. A question can, of course, be seen a kind speech act or dialogue move. Dialogue moves are often defined in terms of the BDI model, which provides e.g. mechanisms for interpretation of indirect questions. However, the generality of the BDI model makes it rather complex, and it is not specifically designed for managing dialogue in the way that the QUD model is. By combining both representations and relate them, a model with the advantages of both models may hopefully be constructed. Dialogue moves can be formalized in terms of the BDI model and still interact with the QUD.

1 Introduction

In this paper I will compare the model of dialogue involving Questions Under Discussion (the QUD model) presented in Ginzburg (1996, 1997) and the classical AI model of mental states involving belief, desire, intentions and possibly obligations (the BDI model) (see e.g. Allen 1995; Poesio & Traum 1997). Both these theories provide ways to represent the common ground between participants in a dialogue, and thus implicitly provide a way to represent the effects of dialogue moves. It will be shown that each approach contains something that the other doesn't and that it is desirable to combine both representations and to relate them in various ways. This combination can hopefully provide a simple and efficient structure used in processing dialogue utterances (the QUD), which is related to a more general, and therefore more complex, reasoning mechanisms and representations of a variety of mental attitudes (the BDI model).

2 The BDI Model of Questions

In the BDI model, questions can be treated as dialogue moves (or speech acts) which establish (i.e. make mutually known) a desire for certain information¹. For example, the effect of A saying "Is it raining?" or "What time is it?" can be formalized respectively as:

$$\begin{aligned} &des(A, knowif(A, raining)) \\ &des(A, knowref(A, \lambda X(current-time(X)))). \end{aligned}$$

Below, such desires will be referred to as *info-goals*. The operators *knowif* and *knowref* are defined in Allen (1995) as follows²:

*Work on this paper has been supported by S-DIME (Swedish dialogue move engine), NUTEK/HSFR Language Technology project F305/97, (<http://www.ling.gu.se/~sl/sdime.html>). Thanks to Robin Cooper for helpful discussions.

¹This is slightly different from Allen's model, where the goal of a question is that the addressee form an intention answer the question. The model used in this paper is more general, since it does not specify any particular action that the addressee should perform or intend to perform in order to achieve the desired goal.

²While basically plausible, the definition of *knowref* needs to be modified to handle (at least) sets of individuals satisfying a certain description rather than only single individuals.

- (1) $knowref(A, \lambda x P_x) \equiv \exists! y (bel(A, ((\lambda x P_x) y))) \equiv \exists! y (bel(A, P_y))$
 $knowif(A, P) \equiv bel(A, P) \vee bel(A, \neg P)$

Using these operators, the intended effects of dialogue moves can be represented, e.g. as below (A is the speaker, B is the hearer, P is a proposition, $\lambda x P_x$ a proposition-abstract and α is an action):

(2) Move	Intended effect
$Assert(A, B, P)$	$mbel(A, B, P)$
$Request(A, B, \alpha)$	$mbel(A, B, des(A, int(B, \alpha)))$
$Query_yn(A, B, P)$	$mbel(A, B, des(A, knowif(A, Q)))$
$Query_wh(A, B, P_x)$	$mbel(A, B, des(A, knowref(A, \lambda x P_x)))$

In (2), we assume that the intended effect of dialogue moves is to make certain facts mutually known. In the case of questions, the intended effect is to make it mutually known that the querier wants to obtain some piece of information. The operator $mbel$ represents mutual belief, which is definable in terms beliefs³. The common ground between participants in a dialogue contains the facts that they mutually believe to be true.

The BDI model is usually couched in some plan-recognition framework, where interpretation of speech acts can be indirect as well as direct. Indirect interpretation usually involves inferring some underlying goal which motivated the speaker to ask the question. The plan recognition approach usually involves the assumption that goals can be hierarchically ordered, e.g. as goal-trees where there is a top goal and usually some leaf goals, interconnected by intermediate goals.

An important point here is that questions, as well as answers, can be indirect. For example, a *wh*-question (or a "request for a referent") can be conveyed using either an imperative sentence ("Tell me the time"), a yes/no interrogative ("Do you know the time?") sentence or a *wh*-interrogative sentence ("What's the time?"). The different ways to realize a certain speech act can be represented e.g. as alternative decompositions of the speech act. The example in (3) is taken from Allen (1995)⁴:

- (3) The Illocutionary Act **RequestRef**:
Decomp1: $WhQuestion(Speaker, Hearer, P_x)$
Decomp2: $Interrog(Sepaker, Hearer, KnowRef(Hearer, P_x))$
Decomp3: $Imper(Speaker, Hearer, InformRef(Hearer, Speaker, P_x))$

An answer can be related to a question by fulfilling the goal established by the question, i.e. satisfying the desire that was made mutually known by asking the question. There may also be other ways of satisfying such desires, e.g. by nonverbal actions. This relation may be called (goal-directed) *relevance*⁵.

- (4) A: Do you know the time?
 B: 12 o'clock

Here, B does not interpret A's question literally, or B would have answered "yes" instead. The inferred underlying goal is A's wanting to know the time, and B's answer is relevant to this goal.

Info-goals can be established in various ways, e.g. by indirectly informing about them as illustrated in (5).

- (5) (Example taken from Ginzburg (1997))
 A: Who is coming tonight?
 B: Why do you ask?
 A: Well after the last party and my antics here I'm anxious.
 B: Oh well, no cause for worry: Jack and Melissa won't be coming.

³Typically, $mbel(A, B, P)$ is analyzed as an infinite conjunction $bel(A, P) \& bel(B, P) \& bel(A, bel(B, P)) \& bel(B, bel(A, P)) \& bel(A, bel(B, bel(A, P))) \dots$

⁴Some notational simplifications have been made.

⁵See Allwood (1995). An account of this concept of relevance and its use in inferring implicatures from Grice's Maxim of Relation can be found in Larsson (1996).

A's first utterance establishes A's info-goal g to know who is coming. A's second utterance serves (indirectly) to inform B of a superordinate info-goal g' : roughly, to know whether or not A has reason to worry.

3 The QUD Model of Questions

Starting from earlier attempts to define the semantics of questions, Ginzburgs defines within a situation-theoretic framework a number of relations that can hold between an answer-SOA⁶ (τ), a question q and a mental state ms of the dialogue participant (DP) who asked q :

- τ potentially-resolves q
- τ strongly-positively-resolves q
- τ negatively-resolves q
- τ resolves q relative to ms
- τ is about q with respect to ms
- τ is goal-fulfilling with respect to ms
- τ partially-resolves q

In Ginzburg's theory, wh-questions are represented as $(s?\mu)$ for SOA-abstract μ and yes/no-questions as $(s?\sigma)$ for SOA σ . For example, the question "Does Jill like Jack?" can be represented as $(s?\langle \text{LIKE, liker:jill, likee:jack} \rangle)$, while "Who does Jill like?" corresponds to $(s?\lambda x(\text{LIKE, liker:jill, likee:x}))$.

All these relations are defined partly in terms of the semantic relation between τ and q . For example, for a yes-no question σ , an answer τ is *about* σ iff whenever $s \models \tau$ it is also the case that $s \models \sigma \vee \bar{\sigma}$. The important thing to notice here is that this definition is given in terms of SOA's (or infons) which reflect the semantic content of the question and answer.

Many relations also have a context-sensitive pragmatic component. For example, a response τ to a query μ is *goal-fulfilling* if (1) τ is about μ and (2) τ conveys information which fulfills the queriers goal relative to the inferential capabilities of the queriers mental state ms , formally $\tau \Rightarrow_{ms} \text{goal-SOA}(ms)$ ⁷.

In Ginzburg (1996), Ginzburg structures a participant's version of the common ground into three separate fields: Facts, Questions Under Discussion (QUD) and Latest-move. The QUD is intended to model an aspect of conversation as the setting up of possible questions to discuss and the subsequent resolving of some of these questions. At any time, a speaker may choose to add something to the QUD, or to address one of the questions in the QUD.

There are several ways in which the QUD can be updated:

- A question is accepted for discussion.
- A question q is rejected; in this case, the question "whether DP will discuss q " is added.
- An assertion p is made; the question "whether p " is added.
- An utterance u of a sentence S with meaning μ is made; the "meaning question" $q(u, S) = (u?\mu)$, paraphrasable as "what values are assigned to the contextual parameters of μ ", is added.

⁶SOA stands for State Of Affairs, and is also known as *infor*. In situation theory, a proposition consists of a situation and a SOA such that the situation supports the SOA, formally $p = (s \models \sigma)$ for a proposition p , a situation s and a SOA σ . For an introduction to situation theory, see e.g. Cooper 1992.

⁷The goal-SOA for a question is, roughly, an exhaustive answer to a question in a given situation. See Ginzburg (1997) for details.

The effect of a question on the context is to (a) “significantly restrict the space of felicitous follow-up assertions or queries”, and (b) “to license an elliptical form which (overtly) conveys only the focus component of the response”.

4 Comparing QUD and BDI

I would like to argue that while the QUD and BDI models have a partial overlap of coverage when it comes to representing the common ground and accounting for the relation between answers and questions, they also each make some important contributions, which mean that it may be fruitful to combine both representations and relate them to each other.

Correlations Between QUD and BDI

Info-goals and Goal-content There seems to be one element in the QUD theory which matches desires for information: the goal-content of a mental state. Thus we get a hypothesis of the relation between info-goals and Ginzburg’s goal-content (assuming that σ is a SOA and μ is a SOA-abstract):

- (6) If q is a yes/no-question, i.e. $q = (s?\sigma)$, then
 $goal\text{-}content(ms_A) = q$ iff $des(A, knowif(A, q))$
 If q is a wh-question, i.e. $q = (s?\mu)$, then
 $goal\text{-}content(ms_A) = q$ iff $des(A, knowref(A, q))$

Some restrictions are imposed on goal content, most notably that an agent can have only one goal at any given time. This restriction is seen by Ginzburg as a simplification with no special theoretical import, which leaves open the possibility of a reformulation of the goal-content predicate which allows for several goals. As a start, we can see $goal\text{-}content(ms)$ as a set, and write $P \in goal\text{-}content(ms)$ instead of $goal\text{-}content(ms) = P$.

Sensitivity to Mental State As we have seen, Ginzburg emphasises that many relations between questions and answers partly depend on the beliefs and goals of the querier. It seems clear that this observation can also be accounted for in the BDI model, since both goals and beliefs are represented and involved in general reasoning mechanisms which include inference as well as planning and plan recognition.

Private and Public Info-goals In Ginzburg’s model, goal-content is a part of the private mental state of a DP. For an answer to be goal-fulfilling the querier must be able to achieve his/her possibly private goal(s) using possibly private beliefs. In the BDI model of questions, however, the point of a question is to make an info-goal public (i.e. mutually known) so that the other participant may address it.

This may appear confusing, but the problem dissolves if we assume that all public goals are also (presumed to be) private goals. A goal which is addressed in dialogue is likely to be public, but this is not strictly necessary for an answer to fulfill it. For example, the DP giving the answer may make an educated guess concerning the querier’s goal, or he may just randomly fulfill it.

Advantages of BDI

Interpretation of Indirect Questions If we take both QUD and desires-for-information as the entities assumed to “drive” a dialogue, i.e. to motivate further utterances, a naive hypothesis about the relation between the QUD and the BDI models might be this:

- (7) If q is a yes/no-question, i.e. $q = (s?\sigma)$, then
 $q \in QUD$ iff both DPs mutually believe that there is a DP A
such that $des(A, knowif(A, q))$
If q is a wh-question, i.e. $q = (s?\mu)$, then
 $q \in QUD$ iff both DPs mutually believe that there is a DP A
such that $des(A, knowref(A, q))$

However, this is not true in general, since the QUD is derived from the literal interpretation of a question while desires for information can be inferred indirectly (using e.g. plan recognition) from several types of behaviour, including orders, questions, assertions or nonverbal actions (see Section 2). How different answers relate to an indirect wh-question conveyed by an overt yn-question is demonstrated in (8):

- (8) A: Do you know the time? (Q)
B: It's 12 o'clock. (R1)
B: Yes. (R2)
B: Yes, it's 12 o'clock. (R3)
B: It's 12 o'clock, yes. (R4)

Let's assume (as one normally would) that A asks q in order to find out what the time is, rather than to find out whether or not B knows the time. In this case, R1 and R3, but not R2, fulfill A 's goal of knowing the time. Note, however, that according to the QUD theory, R2 is (directly) about q , while R1 is not about q . The reason for this is, roughly, that q is a yes-no question while the answer is an answer to a wh-question⁸. We will return to R4 below.

In some cases, however, the QUD exactly matches the goal content: "when the question asked transparently reflects the speaker's goals", i.e. when a question is given a literal interpretation and is not related to any super-ordinate goal. If a dialogue system is never or rarely confronted with indirect questions, an account of such questions will of course not be needed.

Unanswered but Irrelevant Questions Consider the following example:

- (9) A and B are waiting outside a locked door in school, trying to get in. A has a cellular phone.
A: What's the janitor's number? I can call him and ask for help.
 C enters and opens the door, and A and B happily enter and never discuss the janitor's phone number again.

When A asks about the number to the janitor, this question is added to the QUD. When the door is opened by C , the question becomes irrelevant despite the fact that it hasn't received any answer. In this case, it seems plausible that the QUD should be downdated when the super-ordinate goal of getting past the locked door is solved. A possible solution to this is to always check for each QUD whether its super-ordinate goal (if there is one) has already been solved.

Dialogue Move Taxonomies The BDI model is intended as a general framework for representing mental states, and thus allows many different kinds of dialogue moves to be represented in terms of changes to mental states. Exactly which moves to include in a dialogue move taxonomy, what to call them and how to define them has been subject to much debate, but most taxonomies include at least assertions, questions, and requests and often also offers, suggestions and response moves such as answers and rejections. The QUD model as formulated in Ginzburg (1996) includes only assertions (including answers) and questions, and the issue of how to include other moves in the model is left open.

⁸The definition of aboutness being closed under implication does not help here; the fact that it's 12 o'clock (the SOA conveyed by R1) does not implicate that B knows the time, even though the fact that B answered "12 o'clock" does.

Advantages of QUD

The Ordering of QUD vs. the Structure of Goals The questions in the QUD are partially ordered in a way which is intended to reflect their relative foregroundedness, i.e. the degree to which they are perceived as being in the focus of the current dialogue segment. This ordering is achieved mainly through the QUD update rules, but can also be negotiated. In contrast, the goals (mutually known desires) in the BDI model are ordered in tree structures, where a goal typically has some subgoals and perhaps a superordinate goal. Both of these structures are useful in dialogue; the former, to provide the participants with the currently foregrounded questions; the latter, to provide relations between questions which can be used e.g. to remove questions which have been made irrelevant.

QUDs which do not correspond to info-goals It should be noted that the BDI model of questions as presented above only covers one kind of question that can enter the QUD: those that arise when a question is asked and accepted for discussion. Questions which arise by a DP rejecting a question ("whether DP will discuss q "), by a DP making an assertion ("whether q ") and so-called meaning questions do not seem to correspond to desire for information, but rather seem to be imposed on the participants whether they like it or not. Of course, we cannot rule out the possibility that these phenomena may be explained within the BDI framework, but so far this has not been done (to my knowledge).

Info-goals which do not correspond to QUDs Another reason for rejecting the definition in (7) is that not all info-goals correspond to questions under discussion. Specifically, there may be info-goals which are mutually known by the participants from earlier encounters, but which have not been raised in the current discussion. The QUD offers a way to distinguish between "old" questions and questions which are under discussion.

The Notion of Resolvedness Within a situation-theoretic framework, Ginzburg provides a detailed investigation into different ways in which an answer can be said to *resolve* a question. Resolvedness (including partial resolvedness) can be expected to play a major part when deciding whether or not to downdate the QUD (and downdate the BDI model with the corresponding info-goals).

In the BDI model, an answer fulfills an info-goal only if the desired state represented by the *knowif* and *knowref* operators is achieved. Ginzburg's notions of e.g. partial resolvedness seems to provide a more diverse and realistic model of how the QUD (and BDI info-goals) can be downdated even in situations where an exhaustive answer is not given.

5 Combining QUD and BDI

One way of providing a link between the QUD and the BDI model is to provide additional update- and downdate rules for the QUD. Since the two representations share some content (info-goals), it is important that these parallel representations are consistent with each other. The rules given below are intended as rough suggestions.

Updating the QUD with Indirect Questions

To account for indirect as well as direct questions, a QUD update rule along the following lines could be useful:

- (10) Provided the latest move resulted in the BDI model being updated with info-goal $des(A, knowif(A, q))$ or $des(A, knowref(A, q))$, update QUD so q becomes topmost.

If a question is uttered (whether direct or indirect), it will be interpreted as a desire for information and added to the QUD. Note that this does not, however, license us to say that every desire for

information is part of the QUD, since the participants may mutually know a variety of info-goals which have perhaps been discussed previously but which have not been raised in the current discussion.

It would be natural to extend Ginzburg's taxonomy of relations between questions and answers with what may be referred to as *relevance*. We could say that τ is relevant to q if $\tau \Rightarrow_{ms} \text{goal-SOA}(ms)$. This requires us to add some link between q and $\text{goal-content}(ms)$, the goal(s) of the querier. A plausible candidate seems to be to say that the goal-content must be inferred from q . The plan-recognition process used in the BDI model to interpret indirect speech acts seems to be a possible implementation of this inference process.

Downdating the QUD Based on Irrelevance

As we saw in (9), if an info-goal (or indeed any goal) is fulfilled, all its subgoals become irrelevant, and they should thus be removed from the QUD. To account for this, we may give a QUD downdate rule along the lines of (11)⁹:

- (11) If a goal g' is fulfilled, then for all all info-subgoals g of g' , remove the corresponding question q from QUD (if it's there).

This gives us an explanation to (9) above. When A asks about the number to the janitor, that question q is added to QUD. The plan-recognition mechanism of the BDI model will also infer that the reason that A wants the janitors number is so the janitor can help A and B to achieve the superordinate goal g' to get in. When C opens the door and lets A and B in, g' is fulfilled and since q corresponds to a subgoal of g' , q is removed from QUD.

We can also give an account of what happens in (5), repeated below:

- (12) A: Who is coming tonight?
 B: Why do you ask?
 A: Well after the last party and my antics here I'm anxious.
 B: Oh well, no cause for worry: Jack and Melissa won't be coming.

As stated above, A's first utterance establishes A's info-goal g to know who is coming (corresponding to a question q). A's second utterance serves (indirectly) to inform B of a superordinate info-goal g' to know whether or not to worry. B's final reply fulfills g' . Since the reason for establishing g is really to get an answer to g' , g is made irrelevant once g' is fulfilled¹⁰. This gives a motivation for B's second utterance (τ): even if τ does not resolving q , it does fulfill g' and thus serves to remove q from QUD.

Downdating the QUD Based on Answers to Indirect Questions

Indirect questions can be treated in a similar way. The removal of the indirect question licenses removing the corresponding direct question, if such exists. The linking between an indirect question and the utterance from which it was inferred is provided by the plan-recognition component of the BDI model.

- (13) If a question q' which was inferred from a direct question q is removed from QUD, remove both q' and q from QUD.

We can now give an account of (8) above, repeated below as (14) for convenience.

⁹As indicated above, for each question q the corresponding info-goal g is $\text{des}(A, \text{knowif}(A, q))$ if q is a yes/no-question and $\text{des}(A, \text{knowref}(A, q))$ if q is a wh-question.

¹⁰The rest of B's utterance ("Jack and Melissa won't be coming") can be seen as an explanation and motivation for that fact that A shouldn't worry.

- (14) A: Do you know the time? (Q)
 B: It's 12 o'clock. (R1)
 B: Yes. (R2)
 B: Yes, it's twelve o'clock (R3)
 B: It's 12 o'clock. Yes. (R4)

In (14), it seems that while R1 fulfills A's goal and is in this sense felicitous, R3 seems to be even more so. This can be explained by the fact that not only does R3 fulfill A's goal, it is also about q . In fact, only R3 is properly goal-resolving in Ginzburg's terminology.

How can this be explained by the up- and downdating principles presented above? When A utters q , the first thing that happens is that QUD is updated with q (provided the question is accepted). Also, our additional QUD update rule predicts that QUD will also be updated with the indirect interpretation q' corresponding to "what's the time", provided the BDI model is updated with the info-goal $g' = des(A, knowref(A, q'))$. This, of course, requires that the indirect question can be recognized by the plan-recognition mechanism of the BDI model. Now, if B responds with R1, q' will be answered and removed from QUD, which allows q to be removed from the QUD.

If B answers R2, only q will be removed from the QUD, and the answer is thus not felicitous. Of course, if the question was literally interpreted, it will be felicitous since only q need to be removed. If B answers R3, q will first be removed from QUD since it is answered by "yes", and q' will subsequently be removed when it is answered by "it's 12 o'clock".

By contrast, R4 is a very strange answer. This can be explained by the fact that the first part of R4 answers q' and thus removes both q and q' from QUD. When the second part ("yes") is uttered, the question it was (presumably) intended to answer is no longer in the QUD.

Allen notes that "... even if the literal reading is eliminated and an indirect reading is taken, the original form of the utterance affects what answers are allowed", and that "this suggests that at least the form of the literal meaning must be retained somehow in the final analysis". Our current updating rules allow both direct and indirect questions to be stored in QUD.

Defining Dialogue Moves

As noted above, dialogue moves (or speech acts) have been formally defined in terms of the BDI model. If we combine the QUD model with the BDI model, do we need to reformulate these definitions (specifically, the definitions of questions) in terms of QUD? It seems that in Ginzburg's model, the role of definitions of dialogue moves is taken by rules for updating and downdating the DP's gameboard (consisting of QUD, FACTS and LATEST-MOVE, see Ginzburg 1996).

An important difference between the QUD and BDI models is that the QUD update and downdate rules are formulated from the hearer's viewpoint (how to react to an utterance) while the "intended effect" (how the speaker intends the hearer to react) used to define dialogue moves in the BDI model is taken from the speaker's viewpoint. The update and downdate rules can perhaps be seen as heuristics for managing dialogue, replacing the general BDI reasoning mechanism by default (i.e. as long as it seems to be working). This means that traditional BDI move definitions can be retained, and the consequences of moves on the QUD is managed by update and downdate rules which in some cases are triggered by changes in the BDI model.

6 Conclusion

The QUD provides a simple unified representations of questions which can be addressed at a certain stage of a dialogue, including direct questions as well as meaning-questions and certain presuppositions. The representation is partially ordered in a way intended to reflect the relative importance and foregroundedness of the questions under discussion at each stage of the dialogue. Thus, we see that the QUD model provides something not accounted for in the standard BDI model; however, it does not include an account of indirect questions.

The BDI-model provides general representations and reasoning mechanisms providing, among other things, interpretation of indirect questions. However, its generality makes it complex, and it does not include (as far as this author is aware) models of e.g. meaning-questions, nor an ordering of info-goals similar to that of the QUD model.

Ginzburg's notions of e.g. resolvedness and partial resolvedness may provide more flexible accounts of QUD and info-goal downdating in cases where answers are not exhaustive. If this part of Ginzburg's theory is to be integrated with the BDI model, it appears that the BDI model needs to be reformulated within a situation-theoretic framework.

By combining both representations and relating them using a set of QUD update and downdate rules in addition to those in Ginzburg (1996), we can construct a model with the advantages of both models. Since QUD is seen as partly derived from the BDI model (particularly, from info-goals), dialogue moves can be formalized in terms of the BDI model and still interact with the QUD.

References

- Allen, James F. (1995): *Natural Language Processing*. 2nd ed. Redwood, CA: The Benjamin/Cummings Publishing Company Inc.
- Allwood, Jens (1995): *An Activity Based Approach to Pragmatics*. In *Gothenburg Papers in Theoretical Linguistics 76*, Dept. of linguistics, University of Gteborg. Forthcoming in Bunt & Black (eds.) *Approaches to Pragmatics*.
- Cooper, Robin (1992): *A Working Person's Guide to Situation Theory*. In *Topics in Semantic Interpretation*, ed. by Steffen Leo Hansen and Finn Soerensen, Samfundslitteratur, Frederiksberg, Denmark
- Ginzburg, J. (1996b): *Interrogatives: Questions, Facts and Dialogue*, in *The Handbook of Contemporary Semantic Theory*, ed. by Shalom Lappin, Blackwell, Oxford.
- Ginzburg, J. (1997): *Resolving Questions I & II*. *Linguistics and Philosophy* 17 & 18.
- Larsson, S. (1996): *Computing Implicature: The Case of Relevance*. M.A. Thesis, Dept. of Linguistics, Gteborg University.
- Poesio, M. and Traum, D. (1997): *Conversational Actions and Discourse Situations*. *Computational Intelligence*, v.13, n.3, 1997.

Formal Design, Verification and Simulation of Multi-Modal Dialogues*

Ian Lewin
SRI International
Cambridge Computer Science Research Centre,
23 Millers Yard, Mill Lane,
Cambridge, CB2 1RQ, U.K.
ian@cam.sri.com

Abstract

We have designed and implemented a dialogue management design tool for use in dialogue design as a component of user-interface design in multi-modal applications. The tool provides: a formal language (typed feature structures) for describing states and events; a simple rule formalism for specifying dialogues; an automatic dialogue-property checking module; a dialogue-simulator for interactive testing of designs. It is designed to be compatible with GUIDE (Mellor et al., 1995) – a hardware/software environment permitting arbitrary coupling of input/output devices to facilitate experimentation in multi-modal design – but this does not restrict its generality.

1 Dialogue Design in HCI

Dialogue is a notion that researchers in Human Computer Interaction are naturally interested in because they are keen to be able to *design* dialogues with certain properties in mind, notably usability (howsoever conceived). At least some such properties can, it is hoped, be expressed and discussed more rigorously through formalization. Also, dialogue properties may be explored through experimentation with implemented formalizations or through formal analysis.

HCI researchers have been naturally led to questions such as the following. Can dialogues be specified formally? If so, what sort of language is required to specify them? Are the formalizations readily understandable by humans? Once developed, are they easy to maintain and update? Can we prove useful properties of them? Can there be a generic “dialogue manager” for exploring designs interactively? Can the designs be used at least semi-automatically in deriving implementations suitable for end users? Are there generic principles of dialogue design? One can of course discern identical concerns in linguistics varying from those whose interest is primarily philosophical and descriptive to those whose interest is building reconfigurable “plug-in” dialogue components for systems with speech interfaces. Some of the answers that have been proffered also provide for a fascinating comparison with those of linguistic research.

A dialogue, in the HCI analysis, is something that takes place in the User Interface, which mediates between the user and the application. Dialogues are typically constructed out of “utterances” which are described at a highly surface level, for example, *MouseUp*, *MouseDown* and *Return* as delivered by mouse and keyboard devices and *Stop* and *Go* as delivered by a speech recognizer. Dialogue structure determines, or reflects, *what* can be said, *by who* and *when*. For example, one may not be able to select a file from the file menu until the file menu itself has been selected. Together with this notion of dialogue structure is often also one of *dialogue state*, where

*The investigation which is the subject of this report was carried out under the terms of Contract CSM3/002 for the Manager, Systems Engineering and Information Group, DERA Malvern

such a state is thought of as determining what utterances can be made next (and by who). That is, one cannot currently select a file from the file menu because the current dialogue state does not permit it. A dialogue state might typically be described by something like "the cursor is positioned over the slider control and 'dragging' is not true". One moves from one dialogue state to another by making an utterance. It is important that these states are generally *not* conceived of as being states of the user or the application (or indeed both), but as states of the User Interface. It is recognized that the distinction of dialogue from the application may sometimes appear arbitrary and certainly may not always be straightforward or even possible (see further below). However, it is generally thought to be worthwhile attempting. Since utterances are generally intended to effect some action in the computer system as well as updating the dialogue state, they are also usually given an "application semantics", for example a mapping from the utterance Stop to an application function which stops a currently running video.

The clearest expression of a separable dialogue component is found in the "Seeheim Model" for user interfaces (Pffaf, 1985; Green, 1986). The general applicability of the model is not unquestioned (see, for example, Cockton 1986; Took 1990), but many researchers have adopted it. The model first distinguishes the user, the application and the user-interface and then, within the user-interface, it distinguishes a dialogue control component, a presentation component and an application interface model. The dialogue control component governs, for example, turn-taking and the range of "utterances" that can be made next given the current dialogue state. The presentation component governs the "look and feel" of the dialogue including interaction with input-output devices; and the application interface model maps the user interface onto the underlying application.

In HCI, the techniques to address our earlier list of questions have often come from formal and computational linguistics, as well as computer science. The models that have been suggested address them to different degrees, but the most important consideration appears generally to have been to have an implementation that a dialogue designer could experiment with interactively. There is, for example, a long history of using state transition networks for describing dialogues (an early example being Parnas 1969). The attraction of networks has lain partly in the simplicity of models for executing them, partly in the possibility of graphically presenting network diagrams (useful in the process of design), and occasionally, for the less expressive versions, in the possibility of formally analysing them. Unsurprisingly to linguists, they became criticised for being large and unwieldy for complex interfaces (Green, 1986; Myers, 1993; Abowd et al., 1995), and for being insufficiently expressive. Examples were presented which finite state and recursive networks could not describe intuitively or at all (Green, 1986; Alexander, 1990). The desire for expressiveness and a simple execution model also led to the use of augmented transition networks (Kamran, 1985; Cockton, 1990). Notations such as BNF and CFGs have also been investigated, but again it appears that the desire always to have fast implementations led researchers to restricted forms of these notations (Green, 1986). Somewhat surprisingly, the use of feature grammars appears never to have been considered although there are signs of partial re-invention (see further below).

From computer science and artificial intelligence have come other models including production systems, Event-Driven Programming models and more declarative notations such as Communicating Sequential Processes (Olsen, 1990; Green, 1986; Alexander, 1990). (The latter again promising executability as well as analyzability).

2 Design of a dialogue design tool

Our dialogue notation is adapted from recent linguistic literature and is intended to be both general (covering linguistic and non-linguistic cases) as well as being an analyzable and interactively testable notation.

One principal requirement for our work was that it be consistent with GUIDE (Mellor et al., 1995), a hardware/software environment permitting arbitrary coupling of input/output devices to facilitate experimentation in multi-modal design. In principle, dialogue designers are to be encouraged to explore designs using a variety of different input-output modalities. The different

designs will, one hopes, be measurable for different qualities. For example, one might hope to show that, subject to an error rate of less than $x\%$, using a speech recognizer will lead to faster dialogues than using a hierarchically structured pull-down-menu system for selecting alternatives. It was this requirement that led us particularly to the use of feature-value structures and partial specification for the specification of events. By judicious use of partial specification, one can separate out those features of a dialogue design which can be specified independently of the modality of the events which make it up. For example, if a design can permit speaking or typing "go" then one might write a modality independent rule by invoking the partial description

$$\left[\begin{array}{l} \text{device} = X \\ \text{content} = \text{go} \end{array} \right]$$

In line with much earlier work (as described above), we also think of dialogue events as being essentially *dialogue state transformers* and we choose to provide typed feature structures for the description of dialogue states also. So, the above example "the cursor is positioned over the slider control and 'dragging' is not true" might be described by a term such as

$$\left[\begin{array}{l} \text{cursorposition} = X \\ \text{sliderposition} = X \\ \text{drag} = \text{false} \end{array} \right]$$

Again, the possibility of partial specification permits a dialogue designer to abstract over features of dialogue state.

We provide only two non-logical constants in the typed feature formalism: state and event. That is, dialogues will always concern states and events but the choice of state and event features and their possible values is left entirely up to the dialogue designer. Feature terms are permitted to nest (that is, the value of a feature may be something which itself is described by something with features and values) but recursion is not permitted. There will always be a finite set of events and states. This is to facilitate the automatic analysis of designs.

In order to specify the dialogue structure, we provide a very simple rule-based notation of the form

Event : From-state => To-state

where **From-state** and **To-state** are the state pre- and post- conditions of the rule and **Event** is a description of an event which causes the rule to fire. That is, informally, *if* an event described by **Event** occurs, then *if* the dialogue is in a state described by **From-state** then the dialogue can move to a state described by **To-state**. The formalism permits non-determinism. The consequent of a rule may be partially specified – in which case, if a rule fires, then the result is *one* of the states so described. It can be useful for a designer to be able to specify that the result of some action is simply any one of a number of possibilities. Precisely which possibility is selected and how it is selected may not matter so far as the dialogue is concerned. It is also possible to have several rules for the same event with different but overlapping preconditions – in which case, if that event occurs, it is non-deterministic which rule will fire. (This nondeterminism can also arise if rules have overlapping event descriptions).

Our dialogue design formalism is made up of components familiar to computational linguists but its use is novel to HCI. Olsen (1990) apparently first uses an implicit feature value notation precisely so that the preconditions of dialogue rules can be satisfied by many dialogue states in an underlying finite state machine. (He appears keen that his notation be finite-state equivalent although it is not explained why this is desirable.) The values of Olsen's features are always atomic and never variables. Also, events are described only by simple atoms. Olsen is particularly concerned about executing his dialogues (and executing them quickly) and so he specifies a particular execution strategy (a production system with a rule-ordering for conflict resolution) which therefore becomes a determinant of the set of possible dialogues. Olsen also adopts a defaulting value interpretation for his rule operators: if a feature is unspecified in the consequent of a rule then the value remains unchanged rather than being interpreted nondeterministically.

Abowd et al. (1995) usefully ignore the execution strategy and describe what they call a "tabular interface" to Olsen's production systems. Each row i of the table corresponds to a rule. The first column $j = 0$ contains a dialogue event name and each of the remainder columns $j > 0$ represents one feature of the dialogue state and contains a pair of values: the precondition and postcondition value for that feature. So, if all the preconditions in row i are true and the event named in cell $(i, 0)$ occurs then the new dialogue state is described by all the postconditions in row i . Unspecified preconditions are satisfied by all possible states and unspecified postconditions are again given a default interpretation (the value remains unchanged from whatever it was before).

Although we provide a more linguistically expressive language than that of Abowd et al. (1995), we directly follow them in providing an automatic translation of designs into the SMV model description language (McMillan, 1992b). SMV - Symbolic Model Verifier - is a public domain (research use restricted) program for specifying and efficiently proving properties of finite state networks containing large numbers of states. This enables one to provide a number of ready-made theorems and theorem-templates which a designer can fill-in with event or state descriptions as necessary and simply click on in order to have the system prove whether those theorems hold true of the current design. Examples of such theorems include: Is a certain state reachable from another state? In 1,2,5,10,50 steps? Is that state always reachable, no matter what happens? Is it inevitable? Can each rule in the design be fired at some point in a dialogue? Can a rule, once fired, be fired again? Is there a state where nothing can happen?

To provide design prototyping, we further provide an automatic translation of designs into Visual Basic CallBacks (procedures paired with events that are executed whenever those events are recognized). The user provides a set of translation templates which map events in design notation into known Visual Basic CallBacks. The system then provides a simple simulation environment which displays a description of the current dialogue state and updates it accordingly when the relevant CallBacks are processed.

3 Dialogue Design Formalism

A dialogue design consists of a type hierarchy, a set of feature constraints, a start state description and a set of rules. The type hierarchy and feature constraints are adapted from Carpenter (1992).

Type Hierarchy

The type hierarchy is a tree of types with a natural supertype-subtype ordering. The topmost node is *top*, i.e. everything is a subtype of type *top*, and *top* has (at least) the two daughters *state* and *event*. The dialogue designer is free to add to this type hierarchy as he wishes, for example, subdividing states and events into different types of state and event and adding new sorts of object below *top*. (Such new sorts will be the values of features of states and events).

Feature Constraints

The feature constraints form a set of declarations of the form

`type1.feature isof type2`

and state that *feature* is an *appropriate* feature for the type *type₁* and that, in that type, *feature* takes values of type *type₂*. That is, any object of type *type₁* possesses a feature *feature* which takes values in type *type₂*. As in Carpenter's work, a feature which is appropriate for a type is also appropriate for all its subtypes and, by default, its possible values in those subtypes are the same as for the supertype although one can further restrict them to subtypes of those types by adding a further feature constraint. The designer is free to add whatever features he likes and restrict the ranges of their possible values as he likes subject to two conditions. Two types can share a feature only in virtue of being instances of some common type which is declared to have that feature itself (in Carpenter's terminology, features are always *introduced* at precisely one type as one searches

breadthfirst down the type hierarchy). Finiteness is ensured by forbidding recursion: no feature appropriate for a type $type_1$ may take a value of $type_1$ itself nor any subtype or supertype of $type_1$, nor indeed any other type with a feature whose value is (or depends upon, through a chain of features and values) of type $type_1$ or a supertype or subtype of $type_1$.

The type hierarchy and the feature constraints together define the set of possible states and events – the design space for a given problem. In Figure 3, there are fourteen possible events. There are only two types of event of which the system events possess one simple boolean feature important (consequently there are two possible system events) and user events possess two features device and contents taking three and four possible values respectively (consequently there are twelve possible user events). This particular type hierarchy permits a number of, arguably, unwanted events, for example, a speech-recognizer delivering a click event. Whether they are wanted or not is of course a question, possibly a highly suggestive one, for the dialogue designer. They can be removed by adding more feature constraints and more finely classifying the type hierarchy. For example, we can identify speech events as a subtype of user events and add two feature constraints that state that the device feature in speech events always has a speechrecognizer value and that the content feature has some other range of values (*not* including click).

Type Hierarchy

```
top(state(menupost,nomenu),
    positions(offscreen,onscreen(one,two,three)),
    flagtypes,
    flagcolours(red,black),
    flagpositions(up,down),
    event(system,user),
    devices(mouse,keyboard,speechrecognizer),
    boolean(true,false),
    contents(click,movt,return,read))
```

Feature Constraints

```
state.flag isof flagtypes
flagtypes.position isof flagpositions
flagtypes.colour isof flagcolours

state.cursor isof onscreen
state.mailbutton isof positions
state.readbutton isof positions

menupost.mailbutton isof one
menupost.readbutton isof two
nomenu.mailbutton isof one
nomenu.readbutton isof offscreen

user.device isof devices
user.content isof contents
system.important isof boolean
```

Figure 1: A Type Hierarchy and Feature Constraints

The type hierarchy and feature constraints also provide the non-logical constants for the feature-value language used to describe states and events. A feature-value equation is written *feature = term*. A term is a list of feature-value equations or a variable, each optionally preceded

by a type. Figure 2 shows 6 terms that all denote that subset of states with red flags.

$$\begin{array}{l}
 \text{state: } \left[\begin{array}{l} \text{flag} = \left[\begin{array}{l} \text{colour} = \text{red} \\ \text{position} = \text{X} \end{array} \right] \\ \text{cursor} = \text{Y} \end{array} \right] \quad \left[\text{flag} = \left[\begin{array}{l} \text{colour} = \text{red} \\ \text{position} = \text{flagpositions:[]} \end{array} \right] \right] \\
 \\
 \text{state: } \left[\begin{array}{l} \text{flag} = \left[\text{colour} = \text{red} \right] \\ \text{cursor} = \text{onscreen:Y} \end{array} \right] \quad \left[\text{flag} = \left[\begin{array}{l} \text{colour} = \text{red} \\ \text{position} = \text{flagpositions:X} \end{array} \right] \right] \\
 \\
 \text{state: } \left[\begin{array}{l} \text{flag} = \left[\text{colour} = \text{red} \right] \\ \text{cursor} = \text{Y} \end{array} \right] \quad \left[\text{flag} = \left[\text{colour} = \text{red} \right] \right]
 \end{array}$$

Figure 2: Co-denoting, partially specified terms

Start State Description

The start state description is a feature-value term that denotes all the possible starting states of the dialogue.

Rules

A rule consists of a rule-name, an event description, and two state descriptions (a precondition and a postcondition) in the form

`rule(name, event, state-precondition, state-postcondition)`

The meaning of the rule is that, if an event *e* occurs which is described by *event* and the dialogue is in a state described by *state-precondition* then the dialogue moves to a state described by *state-postcondition*. Rule 6 in Figure 3 shows a rule that states that whenever a user event occurs (through any device) whose content is *read*, then the flag becomes positioned down and is coloured black (other things remaining unchanged).

4 An Example Design

In order to illustrate the notation, we describe a very simple example of a multi-modal email interface.

Consider a simple extension to a standard window-based email notification program. On the screen at all times, is an icon of a mailbox with a flag. The flag signals up if new mail has arrived and down otherwise. To make things more interesting, we will suppose that the mailbox has some means of determining whether the new mail is important or not. The flag is red if there is new important mail and black if the new mail is not important. (The flag is just black when there is no new mail). The user can start the mail-reading program either by clicking on the icon, which will bring up a menu from which he should select the *readbutton* item, or by saying *read* to a speech recognizer, or by typing *read* at the keyboard. Starting the mail-reading program will cause the mail flag to signal down.

The Type Hierarchy and Feature Constraints for this example are those shown earlier in figure 3. We have already discussed the events for this example. States have four features *flag*, *cursor*, *mailbutton* and *readbutton*. The value of the *flag* feature determines the appearance of the flag. In the example Type Hierarchy, flags have two possible colours and two possible positions. We noted above that there is actually no scenario in which the flag can be both red and down. One can encode in this in the Type Hierarchy by using a method similar to the one we suggested earlier for ruling out *click* events that emanate from the speech recognizer. Alternatively, one can leave them in as possible states and then use the Dialogue Property Checking module to check that no such state can actually be reached in the current design. (Such a check is illustrated in figure

3). The value of `cursor` determines where the cursor is on the screen. In our example, we are only interested in whether the cursor is positioned over our mailbox icon, over the `readbutton` menu item or somewhere else. Consequently, we just identify three onscreen positions. The fourth position `offscreen` is not a valid value for the cursor but it is a valid value for the `readbutton` menu item. That is, when that menu item is not posted on the screen it is considered to be posted offscreen. In order to incorporate this information, we subclassify states into two types: `menupost` where the mail icon is posted on the screen and so is the `readbutton` menu item; and `nomenu` where the mail icon remains posted but the `readbutton` menu item is not.

We identify our dialogue start state as follows

$$\text{nomenu: } \left[\begin{array}{l} \text{flag} = \left[\begin{array}{l} \text{position} = \text{down} \\ \text{colour} = \text{black} \end{array} \right] \\ \text{cursor} = \text{three} \end{array} \right]$$

Then, we describe the dialogue through the following six rules

Rule 1 (Cursor Movement)

$$\text{rule}(1, \text{event: } \left[\begin{array}{l} \text{device} = \text{mouse} \\ \text{content} = \text{movt} \end{array} \right], \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \\ \text{flag} = Z \end{array} \right], \text{state: } \left[\begin{array}{l} \text{cursor} = W \\ \text{readbutton} = Y \\ \text{flag} = Z \end{array} \right])$$

Cursor movement is a particular mouse event which results in the cursor position changing but nothing else. The movement is modelled nondeterministically since we have no interest in precisely where the cursor goes, merely that it does go. (Note that we do not bother to mention the `mailbutton` feature since it cannot vary - it is present merely to enable us to determine when the cursor is over it).

Rule 2 (Click on Mail Icon)

$$\text{rule}(2, \text{event: } \left[\begin{array}{l} \text{device} = \text{mouse} \\ \text{content} = \text{click} \end{array} \right], \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{mailbutton} = X \\ \text{flag} = Y \end{array} \right], \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = \text{two} \\ \text{flag} = Y \end{array} \right])$$

If the mouse is clicked when the cursor is over the mailbox icon, then the mail `readbutton` menu item is posted. (Nothing else changes).

Rule 3 (Click on readbutton menu item)

$$\text{rule}(3, \left[\begin{array}{l} \text{event: } \left[\begin{array}{l} \text{device} = \text{mouse} \\ \text{content} = \text{click} \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = X \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = \text{offscreen} \\ \text{flag} = \left[\begin{array}{l} \text{position} = \text{down} \\ \text{colour} = \text{black} \end{array} \right] \end{array} \right] \end{array} \right])$$

If the mouse is clicked over the mail `readbutton` menu item, then that menu item is removed from the screen and the flag becomes black and signals down.

Rule 4 (Important Mail arrives)

$$\text{rule}(4, \left[\begin{array}{l} \text{system: } \left[\text{important} = \text{true} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \\ \text{flag} = \left[\begin{array}{l} \text{position} = \text{up} \\ \text{colour} = \text{red} \end{array} \right] \end{array} \right] \end{array} \right])$$

The arrival of important mail leads to a red flag signalling up.

Rule 5 (Unimportant Mail arrives)

$$\text{rule(5, } \left. \begin{array}{l} \text{system: } \left[\begin{array}{l} \text{important} = \text{false} \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \\ \text{colour} = Z \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \\ \text{flag} = \left[\begin{array}{l} \text{position} = \text{up} \\ \text{colour} = Z \end{array} \right] \end{array} \right] \end{array} \right))$$

If unimportant mail arrives, then the flag signals up but nothing else changes. (Consequently, if unimportant mail should arrive after important mail which has not been read, the flag will still signal red).

Rule 6 (Read Mail)

$$\text{rule(6, } \left. \begin{array}{l} \text{event: } \left[\begin{array}{l} \text{content} = \text{read} \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \end{array} \right], \\ \text{state: } \left[\begin{array}{l} \text{cursor} = X \\ \text{readbutton} = Y \\ \text{flag} = \left[\begin{array}{l} \text{position} = \text{down} \\ \text{colour} = \text{black} \end{array} \right] \end{array} \right] \end{array} \right))$$

If mail is read (via any device), then the flag signals down and is coloured black.

5 Dialogue Property Checker

Dialogue Property Checking is carried out through a simple graphical interface to SMV. Figure 3 shows an example screen.

Dialogue designs are loaded into the left-hand window (where they can also be edited) and can then be compiled into a finite state model as described by the SMV input language McMillan (1992a). Properties of this model can then be verified by clicking on the buttons shown to the right of design itself. In the example, the *Deadlock-free* property has already been tested and the tick shown immediately to the right indicates that this property holds of the given design. The SMV model checker is invoked to perform the proof. The properties shown are associated with suitable formulae in Computational Tree Logic, a subset of modal branching time logic, and these formulae are then evaluated by the model checker against the supplied model. (Many of the formulae we use are taken from Abowd et al. (1995)).

Individual state and event descriptions can also be entered in the two small windows labelled *State* or *Event*. The contents of these windows form the inputs to the template-based theorems which are listed above them. In the example, the *State* or *Event* <1> window describes a state whose flag is red and down. The fourth property listed is <1> *reachable* which is marked with a cross. This means, therefore, that in the current design one cannot reach a state where the flag is red and down.

Evidently the expressive capabilities of Computational Tree Logic determine what properties of dialogues one *can* test via this general approach. The syntax of CTL can be given by

1. Every atomic proposition is a CTL formula
2. If f and g are formulae, then so are $\neg f$, $(f \wedge g)$, AXf , EXf , $A(fUg)$, $E(fUg)$

where the branching time operators AXf means "in every next state f ", EXf means "in some next state f ", $A(fUg)$ means "on all paths of states f remains true (at least) until some state i at which g is true", $E(fUg)$ means "there is a path on which f remains true (at least) until some state i is reached at which g is true". Other operators such as AGf ("it will always be the case along every path that f ") can be defined in terms of these ($AGf =_{def} \neg E(\text{true}U\neg f)$).

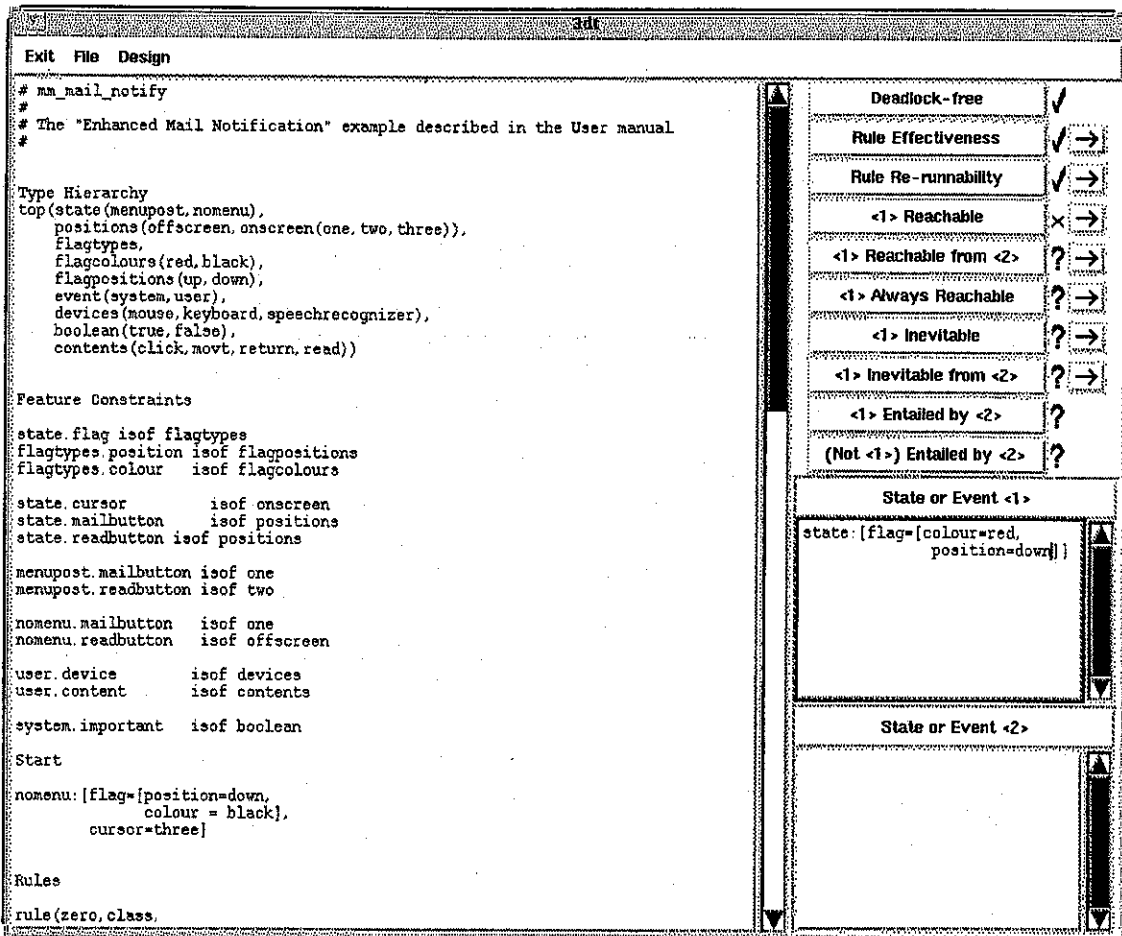


Figure 3: Dialogue Design Tool Snapshot showing a portion of a design in the left hand window, and, on the right, that the first three theorems hold true, the third does not, and the remainder are untested

It is an open question whether the properties that can be expressed through this logic will prove useful or intuitive for dialogue designers. Although it appears natural to hide the details of formulae of Computational Tree Logic from dialogue designers, it is sometimes only through appreciation of them that one can really understand what property is actually being tested. Interestingly, Abowd et al. (1995) make a deliberate attempt to explain the dialogue properties they test without using CTL syntax (“[the] idiosyncracies do not enhance the understanding”) but they themselves misinterpret a formula (from Wang and Abowd 1994). Given a rule for an event e , with preconditions p , they interpret $AG((e \wedge p) \rightarrow EX(E(trueUp)))$ as reversibility which is glossed as “it is always true that if e is performed in a certain state, then that original state can be reached in the subsequent state by some action”. However, all that is actually guaranteed is that the preconditions for the rule can again be satisfied after the rule has fired – which does not mean one will necessarily be back in the very same state. This property is the one labelled Rule Re-runability in figure 3.

6 Dialogue Simulation

In order to provide interactive simulations of dialogue designs, the dialogue design tool includes a small dialogue simulator. The simulator is built around the notion of an Event Driven programming cycle because the tool is designed to be compatible with Guide which is implemented in Visual Basic, which is itself an instance of Event Driven programming. The Event Cycle loops around waiting for events to happen, and, when they do, appropriate pieces of program code ("callbacks") are executed. In the simulator, there is a notion of "current dialogue state" and a compiled set of dialogue rules. Whenever a callback is executed, a conflict-set of possible rules to fire is generated by testing the current dialogue state against the preconditions of all the relevant rules (those associated with the callback). If there is more than one applicable rule (this can happen owing to indeterminacy), then the simulator simply throws up a dialogue box on the screen asking the user to choose which rule to fire. The rule is then fired and the system re-enters the Event Cycle.

A small debugging window shows the current dialogue state (in feature-value notation) to the user. Also, the user has all the standard debugging facilities of Visual Basic at his disposal.

In order to carry out the compilation, the key requirement is to map events in design notation into callbacks. If a dialogue designer has a particular implementation environment in mind (e.g. he is going to use the standard callbacks supplied with Visual Basic plus one or two known add-on packages), then he will have ensured that his design notation reflected the structure of the callbacks that he knows he is going to invoke. In this case, the mapping ought to be trivial. The mapping is defined by a set of translation templates, each of which contains a description of an event in design notation and a callback procedure name (plus any formal parameters associated with it). The compilation program matches each template against all the rules in the design and builds a table of possible design rules to fire when the callback is executed. The body of the callback will first test whether the rule's precondition is met and, if it is, add it to the conflict set. Then it calls a procedure to choose and fire a rule from that set. The callback may also test the value of formal parameters. This is necessary if the callbacks are less fine-grained than the events described in design notation. For example, a design may specify what should happen when key *a* is pressed. Visual Basic does not provide a callback for this particular event, but only for key-presses in general. Which key was pressed has to be determined inside the generic callback. If there are such tests to be performed on the values of formal parameters, then a "fill-in-the-blanks" file is generated by the compilation routine for the user to edit. The file defines a simple tabular mapping between design notation values and implementational values - for example, that *a* corresponds to a value of 97, this being the test that Visual Basic applies for detecting key-presses of *a*.

7 Summary

Research in Human-Computer Interaction has been investigating the notion of dialogue between man and machine with the aim of improving the design of these dialogues. In particular, there is a strand of research which attempts to use Formal Methods to aid dialogue designers. Many of the issues bear worthwhile comparison with investigations (theoretical and practical) in linguistics. We have described one extension to this strand which used ideas and notation from recent developments in linguistics in order to understand and extend the expressivity of the sorts of dialogue rules proposed in HCI. We have implemented our notation in a dialogue design tool which is packaged with a module for automatically proving certain properties of dialogues and for simulating their execution.

The tool has been tested on a number of simple toy examples. We are currently engaged in a more substantial evaluation in a further project exploring the design space for a real-world multi-modal application for our client.

References

- Abowd, G., Wang, H., and Monk, A. (1995). A formal technique for automated dialogue development. In *Proceedings of 1st Symposium on Designing Interactive Systems -DIS85*, Ann Arbor, MI.
- Alexander, H. (1990). Structuring dialogues using csp. In *Formal Methods in Human-Computer Interaction*, Cambridge Series on Human-Computer Interaction, pages 273–295. Cambridge University Press.
- Carpenter, B. (1992). *The Logic of Typed Feature Structures with Applications to Unification Grammars, Logic Programs and Constraint Resolution*, volume 32 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.
- Cockton, G. (1986). Where do we draw the line? - derivation and evaluation of user interface software separation rules. In *People and Computers: Designing for Usability*, pages 417–432. Cambridge University Press.
- Cockton, G. (1990). Designing abstractions for communication control. In *Formal Methods in Human-Computer Interaction*, Cambridge Series on Human-Computer Interaction, pages 233–261. Cambridge University Press.
- Green, M. (1986). A survey of three dialogue models. *ACM Transactions on Graphics*, 5(3):244–275.
- Kamran, A. (1985). Issues pertaining to the design of a user interface management system. In Pffaf, G., editor, *User Interface Management Systems*, pages 43–48. Springer Verlag, Berlin.
- McMillan, K. (1992a). The smv system. Unpublished report available with SMV software.
- McMillan, K. (1992b). *Symbolic Model Checking: An Approach to the State Explosion Problem*. PhD thesis, Carnegie Mellon University. Technical Report CMU-CS-92-131.
- Mellor, B., Tomlinson, M., and Coleman, N. (1995). The generic user interface design environment. In *Proceedings of ESCA Workshop on Spoken Dialogue Systems, Denmark*, pages 117–120.
- Myers, B. (1993). State of the art in user interface software tools. In Hartson, H. and Hix, D., editors, *Advances in Human Computer Interaction*, volume 4, pages 110–150. Ablex Publishing, Norwood New Jersey.
- Olsen, D. (1990). Propositional production systems for dialogue description. In *Human Factors in Computing Systems: Proceedings of CHI'90*, pages 57–63. ACM Press.
- Parnas, D. (1969). On the use of transition diagrams in the design of a user interface for an interactive computer program. In *Proceedings 24th ACM Conference*, pages 379–385.
- Pffaf, G. (1985). *User Interface Management Systems*. Springer Verlag, Berlin.
- Took, R. (1990). Putting design into practice. In *Formal Methods in Human-Computer Interaction*, Cambridge Series on Human-Computer Interaction, pages 63–96. Cambridge University Press.
- Wang, H. and Abowd, G. (1994). A tabular interface for automated verification of event-based dialogs. Technical Report CMU-CS-94-189, School of Computer Science, Carnegie Mellon University.

A Generic Framework For Dialogue Game Implementation*

N. Maudet and F. Evrard.

IRIT

ENSEEIH

2 rue C. Camichel, 31071 Toulouse cedex, France
{Nicolas.Maudet,Fabrice.Evrard}@enseeiht.fr

Abstract

Many authors have proposed to use *dialogue games* as a structure to model conventions of language use that exist in dialogue interactions. Dialogue games allow two speakers to dialogue in accordance with rules. The main idea is to prohibit some sequences of 'moves' given previous moves (in the manner of *dialogue grammar*) but also given the players' mental states. Unfortunately, each model is built upon a formalism strictly designed for a specific type of interaction (e.g. it is only concerned with local coherence). It is our purpose in this paper to determine common requirements for those models and to propose a generic framework for dialogue game implementation, allowing games to be part of a larger structure reflecting the global coherence of dialogue.

1 Introduction

The term "language game" initially goes back to Wittgenstein (1957) who proposed to see language as a rule-governed activity. But recent and growing interest in game theory seemed to offer new perspectives for this notion. According to Grunig (1989), the term dialogue game is in reference to both Wittgenstein's work and game theory. So, how should we understand these various approaches that use this term? As noted by Burton and Brna (1996), these approaches can be divided in two.

The first approach, which was initiated by Levin and Moore (1980) and Mann (1988), introduces dialogue games as bilateral structures containing conventional actions of both players for specific interactions - a kind of "shared scripts" -. For instance, once he is committed to the helping game, the "helpee" is expected to explain his problem, then the helper is expected to propose a solution, and so on. Steps of the game are cognitive or material acts. The main drawback of this approach is the lack of flexibility of the structure used, only consisting in a list of fixed actions. In a quite similar approach, Brown (1984) uses the notion of *recovery path* to account for unexpected actions. But despite interesting results (see for instance their account for speech acts indirectness), these approaches have become obsolete with respect to recent researchs in planification, especially since plans can be shared between participants of dialogue (Grosz and Sidner, 1990). Besides, Airenti et al. (1993) use slightly modified shared plans - i.e. *behavior games* - which include conventional actions not logically necessary ("*but rather constitute a conventional and habitual part of the interaction*", p.218). Thus it combines the conventional component of the dialogue games together with the flexibility of the shared plans. In summary, and as noted by Traum and Allen (1994), these approaches are only particular cases of plan-based approaches. The well-known intention recognition problem is replaced by what we should call a game recognition problem.

The second approach is much in line with game theory. Utterances, or dialogue acts, are assumed to be moves in a game. The main idea is then to prohibit some sequences of moves, given

*We would like to thank Christophe Delord and Philippe Muller for their fruitful comments

previous moves (in the manner of *dialogue grammar*) but also given the players' mental states. This approach has given rise to the development of several various games: MacKenzie (1979) elaborates a dialectical system (namely DC) in order to avoid the *begging-question* problem in argumentation dialogue. DC have also been used for interacting with KBS (Bench-Capon et al., 1991) and for Intelligent Tutoring Systems Pilkington et al. (1992). Beun (1991) introduced a game for collaborative question solving and Gordon (1994) proposes a model of civil pleading, *the pleading games*, where plaintiff and defendant confront each other. A good review of these approaches can be found in Moore and Hobbs (1998).

These two approaches are not incompatible: they are rather concerned with different problems. The former deals with finding conventional interactions and accounts for the global structure of dialogue. The latter deals with finding norms of a specific interaction and is only concerned with this local structure of dialogue.

In this paper, we are mainly interested in the second approach. Our starting point is the observation that each model is built upon a formalism strictly designed for a specific type of interaction. A question then follows straightforwardly: does there exist *one* dialogue game theory? Or, in other words: despite their use in very different contexts, are there common principles to these models?

Carlson (1983) has already introduced a dialogue game where the aim of the game is not strictly defined (it goes between information sharing and all-competitive debate). But in our opinion the fact that the goal is under-specified obliges to leave many behaviours unexplained - considering only they are the responsibility of the player strategy -.

MacKenzie's and Beun's games are different since the aim is specified at the beginning of the game. DC deals with argumentation dialogue - one player is trying to prove p whereas the other is trying to prove $\neg p$ -, specially trying to prevent the *begging question* problem (e.g. circular dialogues resulting from the assertion of a challenged statement¹). Beun's game deals with collaborative question solving: the state to achieve is the mutual belief of the answer to the question. The typical situation is when none of the players can alone find the answer to the question, thus they have to exchange information in order to achieve their goal.

The first game is typically conflictual (i.e. the goal of one player is opposite to the other one's), whereas the second is cooperative (i.e. the goal is shared).

Our aim is to show that common principles to these games can be found.

Section 2 first makes explicit common requirements for these models, concerning both the representation and the logical constraints for the gameboard. Notions of moves and rules are then highlighted. Section 3 deals with the ability of these models to account for the global structure of dialogue. Two approaches are introduced, and we finally focus on the notions of "level of commitment" and "basic game", specially relevant for the understanding of less cooperative interactions. We conclude this paper and illustrate the introduced framework with a simple case study. For the sake of simplicity, we restrict in this paper to information exchange games where messages are merely propositional formulas.

2 Requirements For Dialogue Games

Representation

We first define the *game board*. This board is typically divided in two: the common board of the game (which contains public information) and the private boards of both players (which contains information only accessible to one player).

common board It is becoming a common assumption to see dialog as the interaction of conversation on a *common board* or *common ground*. In MacKenzie (1979), the common board is divided in two *commitment stores*, containing both moves and consequences of the moves of the players. Our notion of *common board* is larger since it contains the totally ordered

¹The simplest example of such a dialogue is obviously "Why p ?" "Because p !"

set of *moves* played during the games, the identification of the *game* currently played with its associated *rules*, and *facts* players have come to agree on, or that were already accepted before the beginning of the dialogue. This set of facts can be viewed as a mutual beliefs set ($MB_{A,B}$). In some approaches, each player has his own representation of the *common board*. In our opinion, the only justification for such a choice is the perspective of making two conversants play a different game. In this paper, we do not take this possibility into account - but it should be pointed out that this perspective exists.

private boards DC does not use private beliefs, because the system is only a referee. But, if you want to enable a computational agent to play the game, it is necessary to allow each player to have *private beliefs*, in addition to the *common board*². This involves in particular that the games we study are not *complete information games*.

The question of the required level of nested beliefs still remains here. Considering on the one hand that >2 nested beliefs are unnecessary - see Carletta and Mellish (1995) - and on the other hand that it is necessary to model the other player's mental state, we will just consider beliefs about the world (B_A) and beliefs about the other player's beliefs about the world ($B_A B_B$) in our model. This private board structuration is in accordance with Carlson (1983)'s *private lists*.

The need to reason about *ignorance* is particularly strong (specially when you consider the question problem). Ignorance is formally the belief state where an agent A doesn't know if p , (i.e. $\neg(B_A p \vee B_A \neg p)$). Because it allows for the representation of the other's ignorance in our gameboard without the need for the addition of a $\neg B_A$ column, we will use the shortcut $B_A B_B * p$ for $B_A(\neg(B_B p \vee B_B \neg p))$.

In the following, we use the term *cognitive context* to talk about both players' private boards and mutually agreed facts of the common board.

Intention As we will see below, no intention is considered in the structure of our dialogue game. "The idea would be to limit formally the permissible forms of expressions so that intention in utterance would be unambiguous and the need for complex pragmatic parsing would be bypassed." (Pilkington, 1992). This means that we do not deal with interpretation problems - at least inside of a game -.

For this reason, we consider for instance that the Beun (1996) approach falls out of the scope of our work. Indeed, the link between a question about p and its answer is no more the result of a dialogue rule, but rather of the recognition of the fact that the hearer knows that the speaker *wants* to know whether p .

Logical Considerations

We must now determine the way these boards can be modified and updated, and whether all the boards are subjected to the same logical constraints. We emphasize here two properties that should be taken into account in our dialogue game framework.

1. avoid logical omniscience

The use of the well-known logic of knowledge $KD45$ (or weak $S5$) in our framework may lead to some problems. $KD45$ straightforwardly leads to logically omniscient agents. This property, possibly acceptable in some games - Beun assumes this logic - is particularly unsuited to argumentation games. Indeed, winning such a game is often making explicit contradictions in the other's mental states, MacKenzie was well aware of such a problem, and in MacKenzie (1979) he used the notion of *immediate consequence*³ to account for this. Accordingly and to account for this problem, we will consider that inference is a move in the game⁴ (and more

²See also Pilkington et al. (1992).

³"A statement is an immediate consequence of the set consisting of a conditional of which it is the consequence, and the statement which is the antecedent of that conditional", p.129.

⁴See also Hintikka (1984).

generally that any logical operation is a move, see the following Section for further details) and that players cannot hold directly opposite beliefs. Our private beliefs columns aren't subjected to any other logical constraints.

2. allow for different levels of endorsement

Basic operations on boards are classical addition and subtraction. But can we modify all the boards in the same way? (i.e. do we need different levels of endorsement?). The answer is obviously negative: some beliefs may be easier to remove from the boards than others. A satisfactory approach is to consider the common board as cumulative (subtraction unavailable) and the private boards as non cumulative. One may say that it is not accurate enough because beliefs can have different strenghts inside of a board. Indeed, consider for instance that the effects of an assertion from A to B that p are: $B_B B_A p \wedge B_A B_B p$. These two conclusions have obviously not the same strenght⁵. Nevertheless in our framework, we assume that most of these phenomena can be captured by the rules of dialogue. If you reconsider the previous example, this means that $B_A B_B p$ is weaker only because it can be denied at next turn by B^6 .

Moves

A move is the basic action of a player in a game and has a cost. In DC, available locutions or type of moves, are *statement*, *withdraw*, *question*, *challenge* and *resolution*. In Beun's game, a player is able to *assert*, *question*, *disagree* or express *ignorance*. We call this category of moves the *dialogue moves*, and we define a *turn of the game* as the interval between two dialogue moves. Dialogue moves are played alternately, and the speaker who has to play the next dialogue move is called the *active player*.

dialogue moves modify the cognitive context of the game, by the addition or subtraction of facts to the concerned board. In this respect, this notion can be related to that of *dialogue act*⁷. Formally, a dialogue act is defined as a *communicative function* applied to *propositional content*. For instance, the utterance "The sky is blue" is the dialogue act *assert(p)*, where $p =$ "The sky is blue". We also introduce the *empty move*. This move has no effect on the cognitive context but increment the turn of the game as any dialogue move.

In addition, we consider that an active player can use *logical moves*.

logical moves modify a board by the application of a logical rule. A logical move has no propositional content and does not increment the turn of the game. Several logical moves can thus be defined. For instance, the move *infer* can be seen as one application *modus ponens* to its own board - this captures the immediate consequence notion of Mac Kenzie -. We can even imagine moves that forces the other to apply a logical rule on its board!

Goals

Even if no intention is explicitly represented in a game, players have a goal when they play a game: in Beun's game they try to achieve a state where the answer to an initial question is mutually believed, in DC a player is trying to prove p whereas the other is trying to prove $\neg p$. Goals are captured in our model by the structure of the game. Since we restrict ourselves to information exchange dialogue, these goals are merely mental state the players want to achieve. Moreover, in order to take into account cases where these states cannot be achieved, we define the *failure state*. When reaching this state, players have evidence for their inability to achieve their goals. The game is then stopped.

⁵A very convincing approach of these endorsement problems is given in Walker (1996).

⁶See also the related notion of *de facto* commitment in MacKenzie (1979).

⁷"Dialogue act are the functional units used by speakers to change context" (Bunt, 1996).

Rules

The notion of a rule is closely related to that of a game. But, to the best of our knowledge, no approach gives a clear understanding of the different types of rules involved in dialogue games. Mackenzie only distinguishes *dialogue rules* and *commitment rules*, but it seems intuitively that dialogue rules contain rules of different types: for instance, the R.repstat rule specifies that “No statement may occur if it is a commitment of both speaker and hearer”, and the R.request that “After ‘Is it the case that p?’, the next event must be either ‘P’, ‘not P’ or ‘No commitment P’”. In this respect, Beun’s approach is more accurate: it distinguishes *obligation rules*, *maxims* and gives informally the unique dialogue rule of his game (i.e. questions are followed by direct answers, indirect answers or expression of ignorance).

In our model, we have tried to define a taxonomy of rules, based of the following considerations: are these rules game-dependent and not only agent-dependent? How to make classification? In line with many other approaches, we emphasize two relevant contextual factors: the story of the dialogue (i.e. the sequence of moves) and the context, here reduced to the cognitive context.

In order to give a functional interpretation of these rules, let’s first formally define these notions. We define Ω as the set of possible cognitive contexts. According to Bunt, a move can then be defined as a function $m : \Omega \rightarrow \Omega$. The first rules of the games are then merely the *effects rules* given by the functional definition associated with the available types of moves.

effect rules ($m : \Omega \rightarrow \Omega$) These rules give the effect of a move on the cognitive context. In case of a dialogue move, effects on the hearer are in fact beliefs about the speaker rationality associated to this move, and effects on the speaker are supposition about the perlocutionary effect of the act. But the sequence of move sometimes allows to go beyond these conclusions. For instance, *accept(p)* has the consequence that *p* is included in the *common board*.

We now define M as the set of moves of the game. This set is then a set of functions of type m , which can be divided in M_d and M_l , respectively the set of dialogue moves and the set of logical moves ($M = M_d \cup M_l$).

The history of the dialogue is a totally ordered set $h = \langle m_1, m_2, \dots, m_n \mid m_{i=1..n} \in M_d \rangle$. We call H the set of possible histories of the dialogue. We now recognize three new classes of rules.

dialogue conventions ($F_{dc} : H \rightarrow 2^{M_d}$) These rules specify which moves are acceptable after a given history of dialogue. It should be pointed out that we do not only consider the latest move⁸. In DC for instance, the R.Chall rule - central to prevent *begging-question* - prevents from asserting a previously challenged proposition.

speaker’s rationality ($F_{sr} : 2^{M_d} \otimes \Omega \rightarrow 2^{M_d}$) These rules prevent the speaker from playing some moves given his mental state, assuming a certain degree of rationality. Such rules could have been called “gricean” because they are closely following gricean maxims (Grice, 1975). Carlson claims that the dialogue game should explain gricean maxims rather than register them, but we do not totally agree with this point of view: some rationality assumptions are necessary, otherwise interpretation could never arise.

hearer’s rationality ($F_{hr} : \omega \rightarrow \omega$) These rules are concerned with appropriative rationality. Indeed, the hearer is subjected to rationality principle as much as the speaker is - see e.g. Evrard and Awada (1994) -. In particular a rational hearer should not accept all the speaker’s utterances. A rule of “belief transfer” illustrates this idea: $B_A B_B p \wedge \neg B_A \neg p \rightarrow B_A p$

A Turn In The Game

Obviously, not all these rules are activated at the same time. A turn in a game, as previously defined, depends upon the following functional dependence (see Figure 1):

⁸See Ginzburg (1997).

Let h_1 be the history of the dialogue (where m_1 is the last dialogue move, i.e. $h_1 = \langle -, m_1 \rangle$) and ω_1 the cognitive context at turn 1. δ_i are sets of dialogue moves (i.e. $\in 2^{M_d}$), and f_{dc}, f_{sr}, f_{hr} are functions representing dialogue conventions, speaker's rationality and hearer's rationality rules.

1. $\delta_1 = f_{dc}(h_1)$ is the set of *conventionally acceptable* dialogue moves after h_1 .
2. $\omega_{1'} = m_1(\omega_1)$ is the cognitive context after h_1 .
3. $\omega_{1''} = f_{hr}(\omega_{1'})$ is the cognitive context after hearer's mental state update.
4. $\delta_{1'} = f_{sr}(\delta_1, \omega_{1''})$ is the set of conventionally and rationally acceptable dialogue moves (i.e. *playable* dialogue moves) for the active player.

A move is then constrained by two factors: the conventionally acceptable set of moves and the current cognitive context. In this respect, we claim that our model is both rule-based and constraint-based. This involves that, in the same game, two similar moves played at a different turn can give a different $\delta_{1'}$ set of playable moves.

Let's now consider a simple example to illustrate this process, where we assume that players use the Beun's dialogue rule. A asks a question about p to B. B is then the active player.

- $\omega_1 = \{B_{Ap}, B_{Bq} \rightarrow p\}$ and $h_1 = \langle question(p) \rangle$
- $\delta_1 = \{assert(p), assert(\neg p), ignore(p), assert(x \rightarrow p), question(x)\}$
- $\omega_{1'} = \omega_1 \cup \{B_{B}B_A * p\}$
- $\omega_{1''} = \omega_{1'}$
- $\delta_{1'} = \delta_1 - \{assert(\neg p), ignore(p)\} = \{assert(p), assert(q \rightarrow p)\}$

As you may note, $\delta_{1'}$ is not necessarily a singleton. This means that the player needs another function to decide which move he prefers to play.

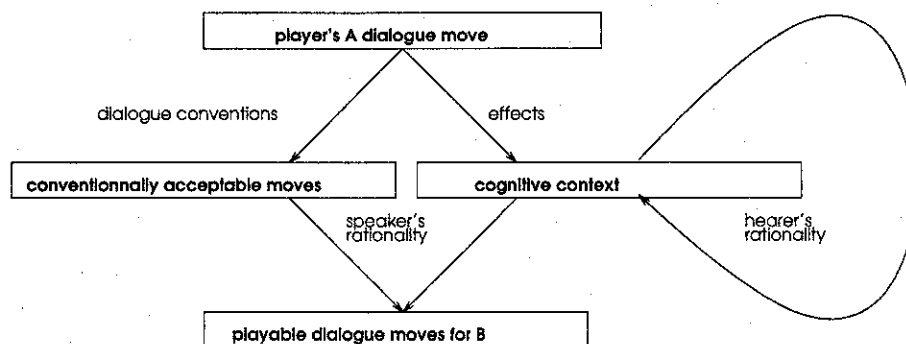


Figure 1: A turn in the game

Strategy

The choice of the move that should be played within possible moves is the *strategy* problem. Strategy is dependent on the agent behaviour and should not take part in the definition of the game. But the border between rationality and strategy is sometimes quite narrow. As Carlson said in Carlson (1984), "Is it a rule of language, or just a good move, to answer a question if one can?"

Nevertheless, we must understand how strategy occurs in the turn process. We assume strategy to be a game-independent function f_{stra} which gives a totally ordered δ_{strat} set of *preferred* moves (including logical moves). This function depends upon goal-dependent heuristics⁹, agent sensibility and cost of the moves¹⁰. Let us return to the description of a turn.

- The chosen move m is the highest element in δ_{strat} such that $m \in \delta_1 \cup M_I$. If this move is a logical move, then return to step 2, else turn ends.

Consider once again the above example. The set of preferred moves could be something like: $\langle assert(p), assert(\neg p), infer, assert(q \rightarrow p), question(q) \rangle$ The speaker will then choose the best ordered playable move. In this case, $assert(p)$ would be the best candidate.

Summary

A dialogue game *game* between A and B is a structure $DG(A, B, game) = \langle S, R, \Phi \rangle$. Φ is the set of available type of moves in the game. S stands for $\langle S_A, S_B, S_F \rangle$ where S_A (resp. S_B) is the state A (resp. B) wants to achieve and S_F is the failure state. R stands for $\langle R_A, R_B \rangle$ where R_A (resp. R_B) is the set of rules associated with A (resp. B). These sets are functionnaly representable as $\langle F_{ef}, f_{dc}, f_{sr}, f_{hr} \rangle$ where F_{ef} is the set of moves functions associated to the elements of δ , and f_{dc}, f_{sr}, f_{hr} are the functions representing dialogue conventions, speaker's rationality and hearer's rationality rules, as previously defined. Two properties can now be defined:

Symmetry We define as *symmetrical*, games in which rules are similar for both players (i.e. $R_A = R_B$)

Cooperativity We define as *cooperative*, games in which the state to achieve is similar for both players (i.e. $S_A = S_B$).

The symmetry property captures the fact that players are not always playing the same rules or committed to the same obligations, even if they are playing the same game (consider for example a justice interaction). The cooperativity property goes without saying. Note that these two properties are completely independent. For instance, Beun's game is symmetrical and cooperative, whereas Mackenzie's is symmetrical and uncooperative.

3 Games And Dialogue Structure

Up to now, we have just considered the dialogue game as a single structure, only concerned with local coherence of dialogue. No explicit information about the dialogue structure appears in these games - recall that the *moves* column is only a totally ordered set without further information -. To our opinion, there are two manners of allowing these models to account for the global structure of dialogue:

- The former is to structuraturé the common board of the game itself, in order to capture aspects of the dialogue structure you want to keep track of. For instance, Ginzburg introduces a "question under discussion" (QUD) partially ordered set - c.f. Ginzburg (1997) - and the *pleading games* need *open*, *conceded* or *denied* statements - c.f. Gordon (1994) -.

⁹In our opinion, *obligation rules* Beun's are of this type.

¹⁰See Pilkington (1992) or Moore (1993) for studies in DC strategies.

- The latter is to consider games as primitives that you can manipulate at a higher intentional level - e.g. Levin and Moore (1980) or Mann (1988) -. The global structure of the dialogue is then defined by the structure on the games - see for instance Burton and Brna (1996) where this structure is a tree of games - . This requires to strictly define notions allowing game's manipulation, like *entry* in a game, *exit* of a game, and so on.

We will focus in this section on two interesting issues of the latter solution: the different *levels of commitment* to a game, and the existence of a *basic game*.

Levels of commitment

The entry in a game involves the entry of the two players: the *actor* of the game and the *partner*¹¹. The actor is the player who initiates the game. For the sake of simplicity, we assume that some dialogue moves of the game, have the further function to be *game moves* of entry. We will note such moves *move+(p)*. For instance, the move *question+(p)* is a question about *p*, but it is first a proposal to enter the question game¹². The partner has then the possibility not to enter the game, as described in the following example:

- (1) A: Will Simon come tonight?
B: I don't want to talk about that.

Else, the partner is *de facto* committed to play the game. But here again we should specify this notion. A distinction is emphasized by several approaches (Airenti et al., 1993; Attardo, 1997; Carlson, 1983), between task level and dialogue level, involving that you can be cooperative at the former level without being cooperative at the latter. In our framework, we use a slightly different notion to express that the partner can play a game without trying to achieve the state conventionally associated to the game. Thus, conventional goal adoption involves speaker's goal adoption only in the case of *cooperative* games. Consider the case where A proposes an argumentation game to B. B is committed to the game when he tries to prove $\neg p$, whereas A tries to prove *p*. The two following examples, where B knows that Simon will come if and only if there is red wine, illustrate this idea:

- (2) A: Will Simon come tonight?
B: I don't know.
- (3) A: Will Simon come tonight?
B: Will there be red wine?
A: Yes.
B: He will come.

In the first example, the player is only committed to rules of the game. He doesn't make his the conventionally associated goal (viz. S_B in our definition). The consequence is that he will not use a strategy to reach S_B , nonetheless he plays with the rules of the game. Note that B doesn't flout rules since he really doesn't know whether Simon will come. In the second example, the player adopts the conventionally associated goal (i.e. he himself tries to achieve the state where it is mutually known whether Simon will come). He then uses a strategy that gives preference to indirect answers regarding over to assertion of ignorance. In summary, we can define different levels of commitment. When a player makes an entry move for *G*, the partner can:

no commit he doesn't want to play the game *G*.

r-commit he player is only committed to the rules R_B of the game *G*.

s-commit he adopts the conventionally associated goal S_B .

¹¹These terms are from Airenti et al. (1993).

¹²See Ginzburg (1997) for similar considerations about QUD.

The difference between the two kinds of commitment is then only sensible at the level strategy. Moreover, the following game's rationality rules are assumed: the actor of the game is s-committed, and s-commitment entails r-commitment.

The Basic Game

A very interesting property of the game structure is that of *nested games*. The notion of nested games capture the fact that you can play two different games G_1 and G_2 at the same time. We note $G_1 \gg G_2$ the fact that G_2 is nested in G_1 . This notion is due to Mann (1988). In our framework, we claim that playing $G_1 \gg G_2$ is playing a resulting game G_{12} whose resulting state is the conjunction of the states, and resulting set of rules is the union of the two sets of rules, with priority to the rules of G_2 in conflicting cases. This notion seems fruitful: in particular it allows to redefine or specify rules in the nested game. Functional definition of the moves is now only game-dependent and can change regarding the context of use. But let us now turn to the *basic game*.

The basic game We assume that primary function of dialogue is that of "information exchange". In the literature Levinson (1979) argues it stands for the *basic* or *unmarked* function of dialogue. Obviously, this is not the only function of dialogue, but the idea is that other functions (i.e. *marked*) derive from it. To capture this idea, we introduce the notion of a basic game. The basic game is supposed to ensure the maximisation of information exchange. Even if this aim is not modelable in term of a mental state to achieve, the notion of cost along with that of "mutual belief increasing" are relevant for its evaluation.

A particular assumption of our model is that players are always s-committed to the *basic game* (i.e. $\forall G, \text{basic-game} \gg G$). But an important point is that, given the definition of nested games given above, some rules of this basic game can be suspended by new rules - see also Leudar and Browning (1988) -. It allows us to consider within our framework uncooperative or at least less cooperative interactions (at the dialogue level) where some maxims of conversation can be flouted.

4 A Case Study

Within the framework previously defined, we represent very simple *basic game*, *question game* and *argumentation game*. In these definitions, $\omega_0 \in \Omega$, $(-) \in H$, M_d is the set of dialogue moves and $\delta_0, \delta_1 \in 2^{M_d}$. Note also that only dialogue moves are specified. For convenience in the effect rules, we note $move_X(p)$ to specify that $move(p)$ has been played by X .

The Basic Game = (maximize information exchange)

dialogue conventions :

$$\begin{aligned} dc((- , assert(p))) &= \\ &\{accept(p), deny + (p), empty, question + (x)\} \\ dc((- , accept(p))) &= \{assert(x), empty, question + (x)\} \\ \text{else } dc((-)) &= M_d \end{aligned}$$

effects :

$$\begin{aligned} assert_A(p)(\omega_0) &= \omega_0 \cup \{B_B B_A p, B_A B_B p\} \\ accept_A(p)(\omega_0) &= \omega_0 \cup \{M B_A B p\} \\ deny_A(p)(\omega_0) &= \omega_0 \cup \{B_B B_A \neg p\} - \{B_B B_A p\} \\ question_A(p)(\omega_0) &= \omega_0 \cup \{B_B B_A * p\} \\ ignore_A(p)(\omega_0) &= \omega_0 \cup \{B_B B_A * p\} \end{aligned}$$

speaker's rationality :

$$\begin{aligned} sr(\delta_0, \omega_0) &= \delta_0 - \delta_1 \text{ where } \delta_1 \text{ is constructed as follows:} \\ \text{if } B_A p \notin \omega_0 &\text{ then } \delta_1 = \delta_1 \cup \{assert(p)\} \\ \text{if } B_A p \in \omega_0 \text{ or } B_A \neg p \in \omega_0 &\text{ then } \delta_1 = \delta_1 \cup \{question(p), ignore(p)\} \\ \text{if } B_A \neg p \notin \omega_0 &\text{ then } \delta_1 = \delta_1 \cup \{deny(p)\} \end{aligned}$$

hearer's rationality :

if $B_A B_B p \in \omega_0$ and $B_A \neg p \notin \omega_0$ then $hr(\omega_0) = \omega_0 \cup \{B_A p\}$
 else $hr(\omega_0) = \omega_0$

The Question Game = $\langle MB_{A,B} p \vee MB_{A,B} \neg p \rangle$

dialogue conventions :

$dc(\langle -, question(p) \rangle) =$
 $\{\langle assert(p), assert(\neg p), ignore(p), question + (x) \rangle\}$

effect :

$assert_A(p)(\omega_0) = \omega_0 \cup \{MB_{A,B} p\}$

The Argumentation Game = $\langle MB_{A,B} p, MB_{A,B} \neg p \rangle$

dialogue conventions :

$dc(\langle -, resolve(p) \rangle) = \{\langle retract(p), retract(\neg p) \rangle\}$

effect :

$retract_A(p)(\omega_0) = \omega_0 - \{B_B B_A p, B_A B_B p\} \cup \{MB_{A,B} p\}$

We illustrate the use of these games with dialogue (4), where context before the beginning of the conversation (Step 0 in Figure 2) is the following: A doesn't know whether Simon will come but he believes that Lucie will come and that it is not true that the presence of Lucie implies that of Simon. On the contrary B believes this last fact, and he also believes that when Lucie comes, she brings red wine. Moreover, A and B agree that when there's red wine, Simon is always there!

- (4) A1: Will Simon be there tonight?
 B2: If Lucie comes, he will come.
 A3: I don't think so.
 B4: When Lucie comes, she brings red wine
 A5: It's true.
 B6: So?
 A7: You're right.
 B8: .
 A9: So Simon will be there.

Steps of the dialogue are described in Figure 2 (note that logical moves do not increment the turn). Inside of the gameboard, "-" stands for negation. We assume that players always s-commit to the games. The rules used are those previously defined, except for the logical moves "I-infer" and "you-infer" that cause *modus ponens* respectively in his own board or in the other's board. Note that turn 1 causes the entry in a question game, and turn 3 in an argumentation game.

5 Conclusion

Mainly inspired by Beun's and Mackenzie's works, we have introduced in this paper a framework within which dialogue games can be represented. A keypoint of our approach is that moves can be either logical or dialogue moves. Logical moves allow to associate a cost to different logical operations, and specially to avoid logical omniscience, particularly unsuited to argumentation games. We have also clearly distinguished the different kinds of rules involved in these games, and we have highlighted their functional dependence. Moreover, we have introduced two manners of extending the games to the global structure of dialogue, and we finally have focused on the notions of commitment to a game and on the existence of a basic game. Putting together these notions could allow to provide a more accurate definition of what "dialogue level cooperation" is.

Easily tractable, dialogue games can provide empirical results about conventional behaviour in dialogue. Combined with intentional models, they can also facilitate the plan recognition process.

	A		CG		B	
	.	B	MOVES	FACTS	A	.
0	t. -(t->s)			t->s		t->s, t->r
1	t. -(t->s)		A(quest)s		*s	t->s, t->r
2	t. -(t->s)	t->s	B(assert)t->s		*s, t->s	t->s, t->r
3	t. -(t->s)	t->s	A(deny)t->s		*s, -(t->s)	t->s, t->r
4	t. -(t->s) t->r	t->s, t->r	B(assert)t->r		*s, -(t->s) t->r	t->s, t->r
5	t. -(t->s)	t->s	A(accept)t->r	t->r	*s, -(t->s)	t->s
	t. -(t->s) t->s	t->s	B(t-infer)		*s, -(t->s)	t->s
	t. -(t->s) t->s	t->s	B(you-infer)		*s, -(t->s)	t->s
6	t	t->s	B(resol)t->s		*s, -(t->s)	t->s
7	t		A(ret)-(t->s)	t->s	*s	
8	t		B(empty)		*s	
	t		A(t-infer)		*s	
9	t		A(assert)s	s		

basic game

question game

argumentation game

Figure 2: Dialogue 4

A strong assumption of our work is that players are always committed to the same game. In forthcoming works, we would like to take into account cases where this assumption doesn't hold, with a special interest in game's adjustment process.

References

- Airenti, G., Bara, B., and Colombetti, M. (1993). Conversation and behaviour games in the pragmatics of dialogue. *Cognitive science*, 17.
- Attardo, S. (1997). Locutionary and perlocutionary cooperation: the perlocutionary cooperative principle. *Journal of pragmatics*, 27:753-779.
- Bench-Capon, T., Dunne, P. E., and Leng, P. H. (1991). Interacting with knowledge-based systems through dialogue games. In *11th International Conference on Expert Systems and Applications Avignon*, pages 123-140.
- Beun, R.-J. (1991). A framework for cooperative dialogue. In *Proceedings of the Venaco Workshop on multimodal dialogue*.
- Beun, R.-J. (1996). Speech act generation in cooperative dialogue. In *Proceedings of the Twente Workshop on language technology*.
- Brown, G. P. (1984). Linguistic and situational context in a model of task-oriented dialogue. In Vaina, L. and Hintikka, J., editors, *Cognitive constraints on communication*. Dordrecht Reidel Publishing Company, Holland.
- Bunt, H. (1996). Dynamic interpretation and dialogue theory. In Taylor, M. M., Bouwhuis, D. G., and Neels, F., editors, *The structure of multimodal dialogue*. John Benjamin, Amsterdam.
- Burton, M. and Brna (1996). Clarissa: an exploration of collaboration through agent-based dialogue games. In *Proceedings of the EuroAIED*, Lisbon.

- Carletta, J. T. and Mellish, C. (1995). Requirements for belief models in cooperative dialogue. Technical Report HCRC/RP-66, Human communication research center, University of Edinburgh.
- Carlson, L. (1983). *Dialogue games: an approach to discourse analysis*. D. Reidel Publishing company.
- Carlson, L. (1984). Focus and dialogue games. In Vaina, L. and Hintikka, J., editors, *Cognitive constraints on communication*, pages 295–333. Dordrecht Reidel Publishing Company, Holland.
- Evrard, F. and Awada, A. (1994). Copilote électronique pour la conduite automobile. In Pavard, B., editor, *Systemes cooperatifs: de la modelisation a la conception*, pages 309–349. Octares Editions, Toulouse.
- Ginzburg, J. (1997). On some semantic consequences of turn taking. In *Proceedings of the MunDial97 Workshop on formal semantics and pragmatics of dialogue*, University of Munich.
- Gordon, T. F. (1994). The pleadings game: an exercise in computational dialectics. *Artificial intelligence and law*, 2(4):239–292.
- Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics 3: Speech Acts*, pages 64–75. Academic Press, New York.
- Grosz, B. G. and Sidner, C. L. (1990). Plans for discourse. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in communication*, pages 417–444. MIT Press, Cambridge.
- Grunig, R. (1989). Dynamique du dialogue et logique. *Verbum*, 12(2):151–158.
- Hintikka, J. (1984). rules, utilities, and strategies in dialogical games. In Vaina, L. and Hintikka, J., editors, *Cognitive constraints on communication*, pages 277–294. Dordrecht Reidel Publishing Company, Holland.
- Leudar, I. and Browning, P. (1988). Meaning, maxims of communication and language game. *Language and communication*, 8:1–16.
- Levin, J. and Moore, J. (1980). Dialogue-games: meta-communication structure for natural language interaction. *Cognitive science*, 1(4):395–420.
- Levinson, S. C. (1979). Activity types and language. *Linguistics*, 17.
- MacKenzie, J. (1979). Question-begging in non-cumulative systems. *Journal of philosophical logic*, 8:117–133.
- Mann, W. C. (1988). Dialogue games: conventions of human interaction. *Argumentation*, 2(4):511–532.
- Moore, D. J. (1993). *Dialogue game theory for intelligent tutoring systems*. PhD thesis, Leeds Metropolitan University, England.
- Moore, D. J. and Hobbs, D. J. (1998). Computational use of philosophical dialogue theories. *Journal of Informal Logic*, 18(2).
- Pilkington, R. (1992). *Intelligent help, Communicating with knowledge based systems*. Paul Chapman Publishing Ltd.
- Pilkington, R. M., Hartley, J. R., Hintze, D., and Moore, D. J. (1992). Learning to argue and arguing to learn: an interface for computer based dialogue games. *Journal of AI in education*, 3(3).
- Traum, D. R. and Allen, J. F. (1994). Discourse obligations in dialogue processing. In *Proceedings of the ACL'94*.

Walker, M. A. (1996). Inferring acceptance and rejection in dialogue by default rules of inference. *Language and speech*, 39(2).

Wittgenstein, L. (1957). *Philosophical Investigations*. Oxford: Blackwell.

MALBOT: An Intelligent Korean Dialogue System using User Modeling*

S.J. Park, K.H. Cha, W.K. Sung,
D.G. Song, J.D. Park,
Language Understanding Lab.
Systems Engineering Research Institute
P.O. Box 1, Yoosung, Taejon 305-600, Korea
{soop, chakh, wksung, dgsong, jdpark}@seri.re.kr

J. Hohle
German National Research Center for Information Technology
D-53754 St. Augustin, Germany
Joerg.Hoehle@gmd.de

Abstract

In this paper we present a plan-based Korean dialogue system as a natural language user interface for a database. MALBOT supports mixed initiative to give users more control over dialogues. Mixed initiative, user modeling, alternative solution suggestion and handling circumlocution are few of MALBOT's characteristics. By employing user modeling, MALBOT reflects the user's preference in dialogues. The user modeling shell system BGP-MS is adapted to suit the requirements of MALBOT's domain. User modeling provides more intelligent system response by reflecting user's preference. The idea of user modeling in dialogue systems has found a good soil in this study.

1 Introduction

Two major dialogue modeling theories are the model using dialogue grammar and the plan-based dialogue model. In the dialogue grammar model, the user has to follow the predefined path as questions are generally followed by answers and proposals by acceptances. The user has very limited room for digression (Jefferson, 1972). From the view point of a finite state machine, speech acts become state transition labels. Utterances are not simply strings of words, but rather are the observable performance of communicative actions, or *speech acts* (Searle, 1969). The basic assumption of the plan-based approach is that humans do not act randomly but rather they plan their actions to achieve goals.

In this paper we adapt a plan-based dialogue model as a natural language interface for a database in the domain of used car buying. Dialogues in this domain are information-seeking and task-oriented. The user's goal is to obtain information about used cars. It is a mutual assumption that the system already knows the user's goal and the user also knows that the system is only able to answer questions within the domain.

A user model is a knowledge source in a system that contains explicit assumptions on all aspects of the user that may be relevant to the behavior of the system (Kass and Finn, 1991). User modeling enables the system to adapt to the users. This adaptive behavior makes the system interact with the users more intelligently. Without user modeling, dialogue patterns between the

*This research was supported by the Ministry of Information and Communication, Republic of Korea

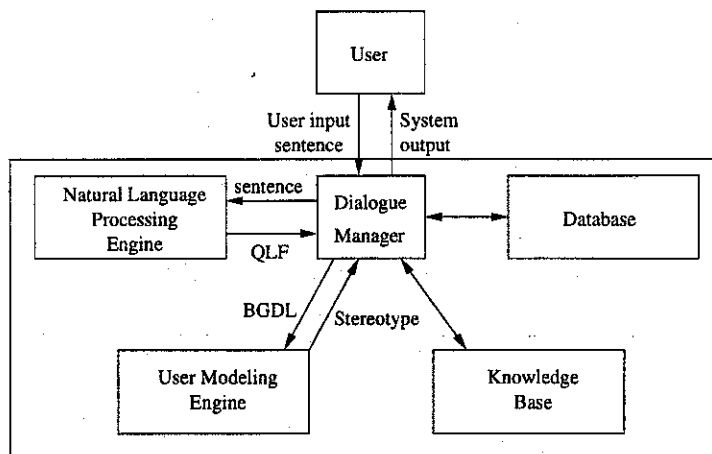


Figure 1: Diagram of MALBOT

users and the system will be the same. If the system utilizes the knowledge of the user and direct the system's behavior to be more intelligent by reflecting the user's preference, the user will be confident with the system's intelligent behavior. Although MALBOT has employed only simple domain specific user models, we think that the idea of user modeling in dialogue satisfies the user's need.

The system keeps six important attributes of the domain as a form. They are *model*, *price*, *color*, *options*, *miles*, and *year*. By using the taxonomy of these six attributes, the system asks questions to obtain enough information to generate a query for the database. For the database system, we collected data of 1000 used cars from The Korea Automobile Press. For system responses, we prepared canned sentences with variable slots which can be replaced by important key words to generate system output effectively. Characteristics of MALBOT are user modeling, mixed initiative, suggesting alternatives and handling circumlocution. The detailed description will be presented in the following sections.

2 System of MALBOT

Overview

The NLP Engine consists of a morphological analyzer, part-of-speech tagger, syntactic analyzer, and semantic analyzer. The dialogue manager and user modeling engine communicate via KQML with the form of BGD¹. The dialogue manager is written in Prolog. It receives QLF² as a meaning representation of the user's input from the NLP engine. Necessary information is transmitted to the user modeling engine while the user and the system exchange messages. We built a domain knowledge base about cars. The knowledge base helps effective and systematic retrieval from the database. Figure 1 shows the system architecture of the MALBOT.

Application Domain of MALBOT

The application of MALBOT is as a system to assist in finding used cars in the database of 1,000 advertisements for used automobiles from a classified advertisement section of a newspaper. More precisely, the MALBOT acts as a dialogue interface to users who want to search for used cars. Using the conceptual taxonomy of the domain, the MALBOT asks questions about necessary constraints until the bottom is reached. It tries to refine what is desired by the user, until at most

¹BGD¹ is the *Belief and Goal Description Language* used in BGP-MS (Kobsa and Pohl, 1995).

²QLF stands for Quasi Logical Form which was introduced in the Core Language Engine of SRI

three advertised objects matching the desired features by the user are found. They can be selected and presented as choices to the user. We collected the data from The Korea Automobile Press.

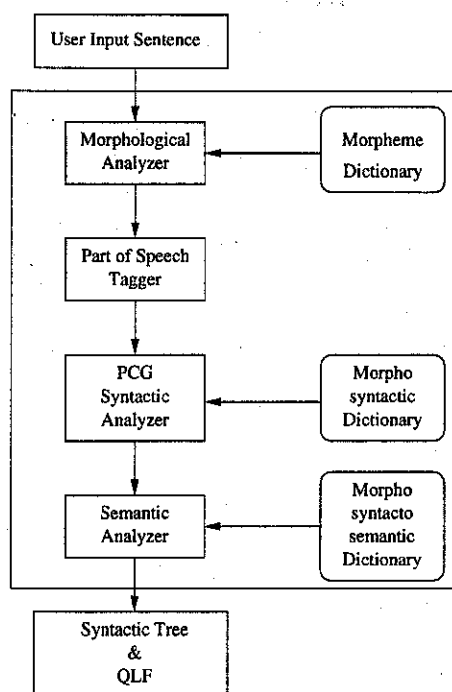


Figure 2: Architecture of natural language processing engine

Natural Language Processing

The natural language processing engine for Korean is based on categorial grammar and quasi logical form for syntactic and semantic analysis. Categorial grammar models Korean syntactic characteristics such as free word order, noun-endings, and various verb-endings. The logical form models each morpheme's semantic structure (Lee et al., 1995). QLF is introduced for the convenient handling of scoping, anaphora and ambiguities.

Based on the compositional semantics derived from Montague theory, QLFs for various morpheme categories are suggested and coded in the dictionary. When two categories are combined by the category cancellation rules, their QLFs are also combined with category and semantic restrictions (Lee et al., 1998). These restrictions resolve semantic/structural ambiguities of the sentence. The derived QLF for the sentence is converted into a logical form through resolution of scope, anaphora and ellipsis.

Figure 2 shows the the architecture of the natural language processing engine. The result of the natural language processing engine is a QLF. Figure 3 shows the semantic representation of the example sentence "흰색 현대 소나타를 사고싶어요." (*I want to buy a white Hyundai Sonata*) Here, "Hyundai" is a maker and "Sonata" is a model name. In Figure 3, N1111 and N1112 represent semantic class hierarchies.

3 Dialogue Manager

The dialogue manager of MALBOT uses a plan-based form filling model. The dialogue manager consists of two parts. The first part extracts mood, modality, predicate, argument of predicates, speech acts, and key words from the user utterances. The second part does database retrieval and system response generation. Since the domain of the system is to retrieve used automobiles

from the database, information-seeking and goal oriented dialogues are the major characteristics of the system. The main goal of the user is to obtain information to buy used cars. So, the system assumes that the user already knows what the goal is.

For database retrieval, we classified the most critical six attributes for this domain. They are *model, price, color, options, miles, and year*. The user's plan can be satisfied by acquiring these six information attributes..

Once the user's stereotype is decided the system prompts questions based on the information which has to be filled in the form. The form has the data-structure of a frame with a set of attribute-value pairs. This form is not only used to generate database queries but also to generate system responses. The system decides what to ask by looking at the unfilled slots of the form. These attributes have different priorities for different stereotypes of users. We predefined the order of important attributes for different stereotypes. So the system asks the question with the highest priority. If the essential information, such as model and price, are satisfied, the system queries the database. If there are too many results from the database the system asks additional questions to narrow down the results to a reasonable size.

The following are the characteristics of MALBOT which makes the system's behavior more intelligent.

Most dialogue system have to deal with the initiative issue. In case of a system driven system, the user is expected to follow the predefined steps which is very limited and provides almost no choice to digress (Goddeau et al., 1996). MALBOT's dialogue management paradigm offers mixed initiative. The initiative can be switched from system to user and vice versa throughout the dialogue. MALBOT initiates a dialogue by asking questions. The user can take the initiative at any time by asking a question during the course of the dialogue. Question answering is one of the key characteristics of information seeking dialogue and most systems follow this pattern. MALBOT also covers various dialogue phenomena like clarification and confirmation.

MALBOT has the ability to adapt to the user. By employing user modeling, MALBOT asks the user questions that are most likely to suit the user. MALBOT has a priority in the questions that have to be asked. User modeling helps the system to decide what to say and how to ask. It also provides helpful advice to the user by evaluating his or her preference. For example, if the user's stereotype has strong emphasis on price or mileage, the system asks about the price information or provides the information about an economical car. If the user wants to buy a Porsche, the system asks about color, extra options and horse power. Mileage may be critical to some users but it maybe irrelevant to other users.

From the user's utterance, the system extracts the meaning and generates a query for the database. Based on the result of database retrieval the system generates its response. When there is no result from the database, the system regenerates the query by changing the slot information which is not too critical to be altered. Suggesting alternative solutions is associated with the user's stereotype. A slot in the form is changed within the boundary of the user's stereotype. Based on a regenerated query, the system provides appropriate and 'close-enough' information to the user.

```
[dcl,
  [modal_honor,
    [modal_serve,
      [show,EV9,\_np[subject];N1111/N1112,
        term(<def, bare, sing, X17,
          [and,
            [named_string, X17, sonata],
            [R4,
              [and,
                [named_string, X15, hyundai],
                [R2,
                  [white, X13], X15]], X17]]), \_np[to];N1111/N13]]]]]
```

Figure 3: Semantic representation in QLF

For example, when the user wants a white Accent³ with the price of \$3,000 which is not in the database, and if there is a red one with the same condition, the system will suggest a red Accent.

When systems ask questions users are usually urged to comply with the question. MALBOT does not require the user to comply with the questions of the system. For example, when the system asks "What color do you want?" then the system expects the answer to be like "I want a red one." This kind of user compliance can be overcome to achieve a more natural and efficient dialogue (Meng et al., 1996). MALBOT not only gives user flexibility in terms of user compliance it can also cover various speech acts such as asking specific questions about the system's inquiry. The following exchange is an example of this.

- U(1): "아토즈를 사고 싶은데요."
(I want to buy an Atoz.)
- S(1): "아토즈는 7대 있습니다. 어떤 가격대를 원하십니까?"
(There are 7 Atozes. What price range do you have in mind?)
- U(2): "흰색 아토즈를 사고 싶어요."
(I'd like to buy a white one.)
- S(2): "흰색 아토즈는 4대 있습니다. 원하시는 옵션은 무엇입니까?"
There are 4 white Atozes. What kind of option do you want?
- U(3): "주행거리가 얼마죠?"
(How many miles on the cars?)
- S(3): "주행거리는 5,500 에서 12,000 킬로미터 입니다."
(The range of miles on the cars is 5,500 km to 12,000km.)

People tend to use indirect expressions in real world dialogues. For example, people use "cheap car" instead of "price of \$100", and "bright color" instead of "white". These phenomena also appear in comparative phrases like "the cheaper car", "the brighter color" and so on. MALBOT covers these kind of dialogue phenomena.

4 User Modeling

The user modeling system BGP-MS⁴ was chosen to serve as the user modeling engine inside MALBOT. BGP-MS gathers and maintains information about a user. Based on the information about the user, BGP-MS tells MALBOT to which stereotype the user belongs.

Stereotypes are the assumptions concerning application domain relevant characteristics of user sub-groups. In order to use the stereotype approach, possible sub-groups are defined within the expected user population whose members are very likely to possess certain homogeneous application-relevant characteristics (Kobsa and Pohl, 1995). For each stereotype, a small number of key characteristics allow the user modeling system to identify the user belonging to the corresponding user sub-group, or as not belonging to this group. Knowledge about these key characteristics is encoded in so-called activation and retraction conditions for each stereotype (Pohl et al., 1995).

Within the domain of cars, it has been shown that the majority of users has a specific interest in a given model. In MALBOT every model is associated with a unique stereotype, which is represented in BGP-MS by having the corresponding stereotype partition contain the concept representing the model by an *is a* relation to the concept of the model. When interest in some particular model is assumed, BGP-MS should therefore activate the corresponding stereotype. In other words, interest in a specific model becomes the primary trigger for a stereotype. However, when the user does not claim interest in a specific model, BGP-MS should find the best matching stereotype using what we herein call the secondary trigger for a stereotype. In the domain of cars, there are six stereotypes, named *Practical*, *Normal*, *Gentle*, *Show-off*, *Fashion* and *Leisure*. Each

³ Accent is a small size car made by Hyundai, Korea.

⁴ BGP-MS was developed at the University of Constance, at the University of Essen and at the German National Research Center for Information Technology(GMD-FIT), Germany.

of these stereotypes contains different defaults or rather priorities for the modeled features of cars. They are *consumption of gas, color, size, price, maintenance cost, and model.*

Once a stereotype is decided, the dialogue manager initiates a system prompt by the priorities of the questions based on each stereotype. The dialogue manager maintains important conditions on each stereotype as well as the order of the system prompts. The system can react to the user's interests more effectively by using the user model than by asking the user to fill out the slots in turn. The dialogue manager also maintains alternative cars for each stereotype. In case no solution is found the dialogue manager will suggest alternatives from this set of cars.

5 Example Scenarios

- S(1): “중고 자동차 검색 대화 시스템 말벗입니다.”
“무엇을 도와 드릴까요?”
(Welcome to MALBOT. I can help you finding cars.)
(What can I do for you?)
- U(1): “BMW를 사고 싶은데요.”
(I'd like to buy a BMW.)
- S(2): “BMW는 12대 있습니다. 어떤 옵션을 원하시죠?”
(There are 12 BMWs. What kind of option do you want?)
- U(2): “94년 이후에 나온 자동차를 원합니다.”
(I want a car which is made after 1994.)
- S(3): “94년 이후에 나온 BMW는 9대 있습니다. 어떤 색상을 원하시죠?”
(There are 9 BMWs after 1994. What color do you want?)
- U(3): “어떤 색상이 있죠?”
(What kind of colors are there?)
- S(4): “검정색 BMW 4대와 흰색 BMW 5대가 있습니다.”
(There are 4-black BMWs and 5 white BMWs.)
- U(4): “검정색은 얼마죠?”
(How much are the black ones?)
- S(5): “검정색 BMW의 가격은 2만 달러에서 3만5천 달러입니다.”
(The price range of black BMWs is from 20,000 dollars
to 35,000 dollars.)
- U(5): “이만칠천 달러 짜리 차를 원합니다.”
(I want the one with the price of 27,000 dollars.)
- S(6): “이만 칠천달러 짜리 검정색 BMW는 없지만 흰색은 있습니다.”
(There is no black BMW but there are two white BMWs
(of about 27,000 dollars.)
- U(6): “주행거리가 얼마죠?”
(How many miles on them?)
- S(7): “주행거리는 6만과 8만 킬로미터입니다.”
(They have 60,000 km and 80,000km.)
- U(7): “너무 많이 뛰었군요.”
(Too many miles on the cars.)
- S(8): “6만 킬로미터 이하인 BMW가 한 대 있습니다.”
(There is one BMW under 60,000 km.)
“색상은 흰색이고 주행거리는 4만 킬로미터입니다.”
(The color is red and the miles are 40,000.)
- U(8): “다른 차는 없나요?”
(Are there any other cars?)
- S(9): “같은 조건의 벤츠와 아우디가 3대 있습니다.”
(There are three Mercedes-Benz and an Audi with the same conditions.)

Above is an example dialogue of MALBOT. At U(1) the user asks about BMW which is a

luxurious car so it triggers the primary condition of the stereotype *Show-off*. Because MALBOT maintains different priorities of asking questions depending on the user's stereotype, MALBOT asks about options which might be the primary concern for those who buy a BMW (see S(2)). In S(2) the system asks about options but the user answers about year instead. To give flexibility to the user, MALBOT does not require the user to comply. In S(3) the system asks about a color and instead of answering the question the user asks more about colors. Here, the user's speech act becomes ASKING. MALBOT is rather tolerant to the user's unexpected dialogue behavior. In U(5) the user specifies the price but there are no cars that match the condition. MALBOT provides alternatives by suggesting BMWs with a different color. Alternatives are provided within the range of the same stereotype. In S(7) and U(7) the system provides miles and the user replies in an indirect way by saying "too many miles". MALBOT covers these kind of circumlocutions by showing other alternatives in S(8).

6 Evaluation

The evaluation is based on the most important aspects supporting intelligent dialogue, according to Edwards and Mason (1988). Initiative concerns messages that are not simply a response to the partner's latest message. In terms of dialogue control, MALBOT supports mixed initiative, where the user or the system can exercise control. As a result, MALBOT provides the user more flexibility: one side can initiate a message at any time, not necessarily in response to the other's immediately preceding message. By using user modeling, MALBOT incorporates user's preferences or characteristics in dialogue. By resolving anaphora and ellipsis the system knows how connections are made from message to message.

We tested MALBOT with ordinary users. Among them, 30% of the users were naive users who have basic knowledge about computers. The rest of them are researchers in our institute. The total number of test users is 178. We logged every dialogue between the users and the system. Then we categorized patterns in which MALBOT could not provide information. Among the results in which MALBOT did not reply appropriately, incomplete parsing of user sentences took two third of the failures. The rest of them results from questions that are out of domain. The reasons for parse failure are lacking dictionary entries, spacing errors and spelling errors. We also found that user's most important concerns are model and price in this domain.

7 Discussion

This paper presents an intelligent dialogue system based on a plan-based algorithm with user modeling. The plan-based method gives users flexibility to behave more naturally in dialogue than the dialogue grammar model where the user follows a predefined path. MALBOT supports mixed initiative to give the user more control over dialogues. User modeling is introduced to reflect users' preferences. User models facilitate the interaction with the user in a more intelligent way. By using user modeling the system can design prompts as well as responses according to users' preferences. User compliance is not mandatory in MALBOT. In natural language dialogue, people tend to use different indirect expressions to achieve their goals. MALBOT supports this kind of circumlocution.

The structure of the dialogue algorithm is simple. The planning mechanism is explicit. We tried to separate components of the system according to their domain dependency. Although it is hard to have a clear division, we prepared for domain portability issues by keeping domain-dependent modules and domain-independent modules separate.

Prospective future work will be in several areas: first, the refinement of the system to be more natural. Second, extending the system to cover various user input sentences to reflect the variety of linguistic behavior of Korean. Third, improving domain portability by achieving effective separation of the system into domain-dependent and domain-independent parts. Finally, applying MALBOT to different domains to test domain portability.

References

- Searle, J.R. (1969). *Speech Acts: An essay in the philosophy of language*, Cambridge University Press.
- Jefferson, G. (1972). Side sequences. In *Studies in Social Interaction*, Free press, New York.
- Lee, W. L. et al. (1995) Chart-Driven Connectionist Categorical Parsing of Spoken Korean. In *Proceedings of the ICCPOL-95*, pages 207-212.
- Lee, G.B. et al. (1998). Research on Semantic Analysis of dialogue sentence in a restricted domain. *Research Report of SERI*.
- Kobsa, A. and Pohl, W. (1995). The User Modeling Shell System BGP-MS *Journal of User Modeling and User-Adapted Interaction* 4(2), pages 59-106.
- Pohl, W. et al. (1995). User Model Acquisition Heuristics Based on Dialogue Acts. In *Proceedings of the International Workshop on the Design of Cooperative System*, pages 471-486.
- Meng, H. et al. (1996). WHEELS: A Conversational System in the Automobile Classifieds Domain. In *Proceedings of ICSLP*, pages 542-545.
- Goddeau, D. et al. (1996). A Form-based Dialogue Manager for Spoken Language Applications. In *Proceedings of ICSLP*, pages 701-704.
- Edwards, J. and Mason, J. (1988). Evaluating the Intelligence in Dialogue Systems. *Int. Journal of Man-Machine Studies*, pages 139-173.
- Kass, E. and Finn, T. (1991). General User Modeling: A Facility to Support Intelligent Interaction. In *Intelligent User Interfaces*, pages 111-128.

Towards an Axiomatization of Dialogue Acts*

Massimo Poesio
University of Edinburgh
HCRC
2 Buccleuch Place
Edinburgh EH8 9LW, Scotland, UK
poesio@cogsci.ed.ac.uk

David Traum
University of Maryland
UMIACS
A. V. Williams Building
College Park, MD 20742 USA
traum@cs.umd.edu

Abstract

Conversations involve all sorts of verbal activities beyond those strictly related to the performance of the task at hand. Among other things, the participants in a conversation have to make sure they both understand what's going on, to manage turn taking, and to keep each other informed about their progress in achieving their task. The participants share information about the status of all of these processes; this suggests that the view of the conversational score they share is rather more complex than assumed in previous accounts. We proposed a preliminary formalization of this more complex view of the conversational score in previous work; in this paper we revise that earlier model, and use our theory of the conversational score to give a partial specification of the effect of the dialogue acts included in the DRI classification.

1 Introduction

Conversations, even task-oriented ones, are a complex business involving all sorts of activities beyond those strictly related to the performance of the task at hand. Among other things, the participants in a conversation have to make sure they both understand what's going on, manage turn taking, and keep each other informed about their progress in achieving their task (Clark, 1996). The information about the status of all of these processes is shared among the participants; this suggests that the view of the CONVERSATIONAL SCORE they share is rather more complex than assumed in previous accounts (Stalnaker, 1979; Lewis, 1979).

We proposed a preliminary formalization of the complex view of the conversational score that emerges from this research in Poesio and Traum (1997). In this paper, we revise that model, and use our theory of the conversational score to partially specify the effect of the dialogue acts included in the DRI classification Discourse Resource Initiative (1997); Allen and Core (1997)—especially those aspects of dialogue act meaning that have not been covered in AI work on speech acts such as Cohen and Levesque (1990b), including how dialogue acts determine which information gets grounded, and how they affect the participants' obligations.

2 Dialogue Acts

Most classic theories of speech acts concentrate on the actions performed by the conversational participants as a way of 'getting the job done'—e.g., instructions to the other conversant, requests for information necessary to accomplish the task, etc. But these actions are only a part of what

*Massimo Poesio is supported by an EPSRC Advanced Fellowship. Thanks to the Rochester TRAINS group, and in particular to Mark Core, Nat Martin, and Len Schubert for comments and suggestions on an earlier version of this paper.

happens in conversations; the participants in a conversation spend a lot of their time making sure they do not talk over each other and ensuring that 'informational' coordination is achieved. Recent theories of speech acts (e.g., Novick 1988; Kowtko et al. 1992; Traum 1994; Bunt 1995) are built on the assumption that a good theory of the actions involved in these aspects of a conversation is as important to a dialogue system as a good theory of task-oriented acts.

The multi-level CONVERSATION ACTS theory, presented in Traum and Hinkelman (1992), maintains the classical illocutionary acts of speech act theory (e.g., **inform**, **request**), now called CORE SPEECH ACTS. These actions are, however, reinterpreted as multi-agent collaborative achievements, taking on their full effect only after they have been *grounded*, i.e., acknowledged. Rather than being actions performed by a speaker to a hearer, the core speech acts are joint actions; the initial speaker and the hearer (called hereafter INITIATOR and RESPONDER, respectively) each contribute actions of a more basic type, the result being the common ground assumed to be the effects of core speech acts.

In addition, Conversation Acts (CA) theory also assumes that three other kinds of speech acts are performed in conversations: acts for TURN-TAKING, GROUNDING, and more complex acts called ARGUMENTATION ACTS; we include in this class both the macro structures of conversation often called GAMES (Carlson, 1983; Levin and Moore, 1978) and the organization of acts according to the rhetorical structure of discourse, as in elaborations. In (Poesio and Traum, 1997) the additional level of LOCUTIONARY ACTS was made explicit in addition to the four levels of the initial proposal. We will not be concerned with turn-taking and argumentation acts here.

The dialogue acts whose interpretation we will discuss are those proposed in the Discourse Resource Initiative (Discourse Resource Initiative, 1997; Allen and Core, 1997), currently the most widely examined proposal for a task-independent set of dialogue acts.¹ The DRI scheme has a somewhat different conceptual organization from CA theory, but it is relatively easy to establish a connection.

We assume throughout that speech acts are just ordinary events, for which we adopt a Davidsonian treatment (Davidson, 1967) as usual in Discourse Representation Theory (Kamp and Reyle, 1993) – more specifically, the version of Davidson's theory proposed by Muskens (1995), in which eventualities are objects of type ϵ and each predicate has an extra argument for the eventuality. We adopt however the standard DRT notation, and write $e : p(\bar{x})$ for $p(\bar{x}, e)$. Each eventuality e is associated in Muskens (1995) with a unique time interval $\vartheta(e)$; we will mostly omit the time intervals below except where necessary to specify the updates.

Locutionary Acts

We use the ternary predicate $e : \text{Utter}(A, P)$ to characterize locutionary acts, where A is an individual, P is a string, and, as mentioned above, e is an eventuality. A locutionary act may consist of an utterance of a single word, a sentence constituent such as an NP, or a complete sentence.²

Core Speech Acts

Core speech acts are dialogue acts which have to do with managing the topic of the conversation, in a general sense. Some of them play a FORWARD-LOOKING FUNCTION: they introduce new social attitudes in the conversation that have to be addressed. The forward-looking acts from the DRI dialogue act coding scheme are shown in (1).

In this scheme acts are hierarchically organized in classes and subclasses; sub-acts maintain all of the properties of the parent act, while also adding additional information about the act.³ The

¹It is being developed by an international team of dialogue researchers from previous coding schemes and speech act taxonomies, and the reliability of this classification scheme has been studied in (Core and Allen, 1997; Di Eugenio et al., 1997).

²We assume that the participants in a conversation also share additional information about a locutionary act such as its syntactic classification if any or its meaning. See (Poesio and Traum, 1997; Poesio, 1997) for details.

³The scheme as developed in Discourse Resource Initiative (1997) included **Info-request** as a sub-class of **Action-directive** — it is a directive in which the directed act is one of making a statement. The authors of Allen

- (1) - **Statement**
 - **Assert**
 - **Reassert**
 - **Other-statement**
- **Influencing-addressee-future-action**
 - **Open-option**
 - **Directive**
 - * **Action-directive**
 - * **Info-request**
- **Committing-speaker-future-action**
 - **Offer**
 - **Commit**
- **Conventional**
 - **Opening**
 - **Closing**
- **Explicit-performative**
- **Exclamation**

current scheme specifies six main act types with subtypes. An initiator is committed to the veracity of her **Statement**. If the statement is used to try to achieve the belief of the addressee (regardless of its success, or the prior belief of the addressee), then it is an **Assert**. If the initiator was already previously so committed, then it is a **Reassert**. An **Other-statement** is a statement that is not an assert or reassert, such as taking a stand on a particular position, without concern to the beliefs of other conversants on this matter.

Influencing-addressee-future-action acts constrain the discourse situation to contain an option for the addressee. **Open-option** does only this; more precisely, it does not count as an attempt to get the addressee to actually do the mentioned act, merely allows it as a possibility for consideration. A **directive**, on the other hand, *does* count as such an attempt. The DRI scheme includes two types of directives to the other agent, depending on what kind of action is directed. **Info-requests** are directives to perform a statement. **Action-directives** are directives to perform another kind of action. Both types of directives also impose an obligation to address the directive itself (though not necessarily to perform the requested action) (Traum and Allen, 1994a). A **Committing-speaker-future-action** act mentions an option of the initiator. A **Commit** act means that the initiator has an obligation to perform the action. An **Offer** is a conditional commitment: if the addressee accepts, then the initiator is committed. **Explicit-performatives** are the traditional speech acts from Austin (1962). The DRI scheme also includes the acts **Opening** and **Closing**, which have to do with the conventional organization of conversations. We will not discuss explicit performatives, conventional acts, and exclamations here. We refer the reader to Allen and Core (1997) for more discussion and examples of these classes, as well as their decision trees.

Other core speech acts are instead classified in the DRI scheme as responses to previous acts: for example, the initiator may accept or reject a previous proposal, or answer a request for information. These acts are called **BACKWARD-LOOKING** in the DRI classification. The backward-looking acts from the DRI scheme playing a function related to the task are listed below; as we will see, other backward-looking acts play functions related to grounding. The specification of such acts

and Core (1997), subsequently decided to make **Info-request** its own main type, since it was often easy to identify using a different syntactic form than other directives (i.e., interrogative vs. imperative mood). While this change is sensible for a coding manual, for reasons of semantic simplicity, we stick with the prior formulation in this paper.

always involves mention of the dialogue act(s) that they are a response to; i.e., all of these acts are implicitly anaphoric on previous speech acts. Again we refer the reader to Discourse Resource Initiative (1997); Allen and Core (1997) for discussion and examples.

- **Agreement**
 - **Accept**
 - **Accept-part**
 - **Maybe**
 - **Reject**
 - **Reject-part**
 - **Hold**
- **Answer**

Hold is the label used for any actions that do not explicitly accept or reject the act they are a response to, but merely postpone the decision.

Grounding Acts

Our model inherits a fundamental assumption of theories such as Clark and Schaefer (1989); Traum (1994): that information has to be **GROUND**ED before it becomes part of the common ground. As in Traum (1994), we assume that grounding is achieved by means of dialogue acts. Acts such as assertions or instructions specify **CONTRIBUTIONS** that have to be **ACKNOWLEDGED** before they become a proper part of the common ground. Acknowledgments can either be performed implicitly or explicitly, by means of linguistics expressions such as *okay* or *gotcha* but also by nodding or by means of expressions such as *uhu*. Here is an example of acknowledgment from the TRAINS-93 corpus:

- (2) utt1: s: take the Avon train to Dansville
 utt2: u: Okay

The participants in a conversation do not always acknowledge contributions right away: they may also signal that they did not understand, e.g., by saying *Sorry, I didn't hear that*.

Some of the backward-looking acts in the DRI classification are concerned with grounding; they are listed below.⁴

- **Understanding-act**
 - **Signal-non-understanding**
 - **Signal-understanding**
 - * **Acknowledge**
 - * **Repeat-rephrase**
 - * **Completion**
 - **Correct-misspeaking**

Turn-taking Acts

The classification in Traum and Hinkelman (1992) also includes a class of acts having to do with the management of the turn—i.e., who is speaking at any given point. Actions in this class include **take-turn**, **keep-turn**, **release-turn**, **assign-turn**. The DRI scheme does not include actions of this type at the moment.

⁴We should note that some of the grounding acts in Traum and Hinkelman (1992) are not included in the DRI scheme - for example, requests for acknowledgments.

Multiple Dialogue Acts

One hypothesis shared both by CA theory and by the DRI proposal is that a locutionary act may generate more than one dialogue act. For example, a locutionary act such as *okay* is typically used to perform actions at both the grounding level and at the core speech act level at the same time; but it is also possible to perform multiple actions at the core speech act level—e.g., an utterance such as *There is an engine at Avon* in the TRAINS domain can be both an **Assert** and an **Open-option**. Following Goldman (1970), we assume that in these cases multiple events are GENERATED by a single locutionary event.

3 The Conversational Score

Grounded and Ungrounded Information

The view of the grounding process proposed by Clark and Schaefer (1989) implies that the conversational score does not simply consist of a record of the information about the conversation that is shared by its participants; instead, it includes both a record of the material that has already been grounded, which we will indicate as *G*, and of the material that hasn't yet been grounded. Following Clark and Schaefer (1989); Traum and Hinkelman (1992), we assume that the ungrounded part consists of a specification of the current 'contributions,' or DISCOURSE UNITS, as they are called in Traum and Hinkelman (1992).

We propose that the view of the conversational score entertained by each CP at any given time (her CONVERSATIONAL INFORMATION STATE) has the structure in (3). We view the conversational information state as a DRS which specifies information about *G* and the discourse units; this DRS gets updated over time as a result of dialogue acts. *G* and the discourse units are also DRSS; we are shortly going to see what kind of information they contain. There are two reasons for these decisions: first of all, the grounding acts refer back to the DUs, as we will see shortly; and second, the modifications to *G* and to the discourse units can be easily modeled as modifications to discourse markers in an extended version of Compositional DRT (Muskens, 1995) with markers denoting DRSS. The CIS also contains information about the currently pending discourse units, which are put together in a list UDUS. The top of UDUS is the Current Discourse Unit CDU—the Discourse Unit to which new material gets added. We write below CDU for $\text{first}(UDUS)$.

	<i>G</i>	UDU	CDU	DU1	DU2	DU3
(3)	$G = \dots$					
	$DU1 = \dots$					
	$DU2 = \dots$					
	$DU3 = \dots$					
	$UDUS = \langle DU3, DU1 \rangle$					
	$(CDU = \text{first}(UDUS) = DU3)$					

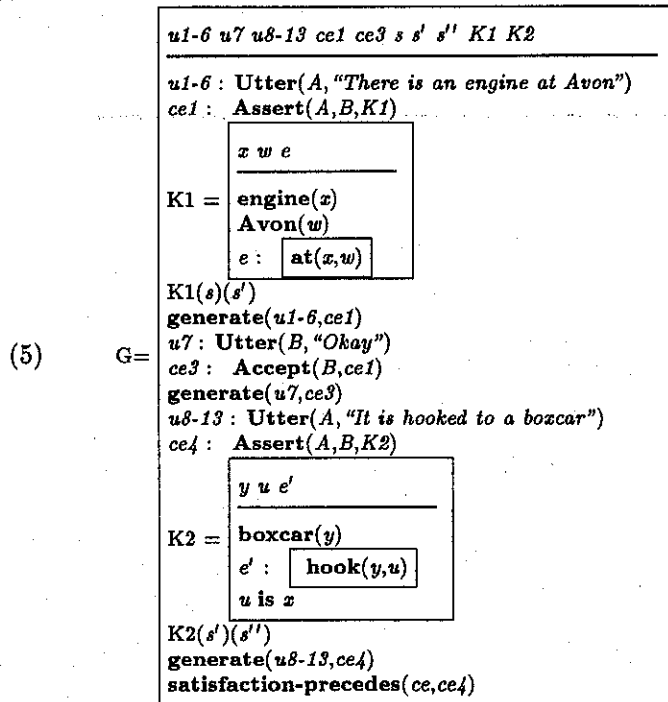
The picture of dialogue we are going to assume below is one in which each act leads to an update of the CIS. All new information gets first added to a DU; this results in obligations of various types and possibly in the responder coming to some conclusions about the intentions of the initiator. Information moves from UDUS into *G* as the result of acknowledgments.

The Conversational Score as a Record of the Discourse Situation

Our second main hypothesis is that the conversational score is, first and foremost, a record of the dialogue acts that take place during a conversation, as well as of the public beliefs, intentions and social commitments of the participants. As such, it can be characterized in terms of the language introduced in DRT to characterize other types of situations. For example, the utterances in (4), if interpreted as an **Assert**, and **Accept**, and an **Assert**, respectively, result in the conversational

participants sharing the information in (5), that includes a record of the occurrence of three locutionary acts and three core speech acts generated by them (we have omitted from (5) all information about smaller locutionary acts such as the uttering of *there*):

- (4) a. A: There is an engine at Avon.
 b. B: Okay.
 c. A: *It* is hooked to a boxcar.



This hypothesis about the conversational score plays two important roles in what follows. First of all, we can assume that agents can reason about the occurrence of dialogue acts and draw some conclusions; most of the updates we will see below are originated by observations of this type. Secondly, we can assume that agents can refer back to dialogue acts just like they do with other events; in this way we can handle the implicit anaphoric reference to events in backward-looking acts.

Compositional DRT

Our formalization is based on DRT – most specifically, on Muskens' formulation of DRT in terms of the theory of types, Compositional DRT (Muskens, 1995). The crucial properties of CDRT to understand what follows are that assignments are treated as first-class objects– of type s – and that discourse entities are viewed as functions from assignments to entities in the domain. DRSS can then be defined as relations between assignments, i.e., objects of type $\langle s, \langle s, t \rangle \rangle$: the DRS $[u_1, \dots, u_n | \varphi_1, \dots, \varphi_m]$ is defined as follows:

$$[u_1, \dots, u_n | \varphi_1, \dots, \varphi_m] = \lambda i. \lambda j. i[u_1, \dots, u_n]j \wedge \varphi_1(j), \dots, \varphi_m(j)$$

where $i[u]j$, the UPDATE OPERATOR, is short for (simplifying somewhat):

$$- \forall v (u \neq v) \rightarrow (v(i) = v(j))$$

For example, the DRS which is the value of the marker K1 in (5) has the following value:

- (6) $[[x \ w \ e | \text{engine}(x), \text{Avon}(w), e : \text{at}(x, w)]] = \{ \langle i, j \rangle \mid i \text{ and } j \text{ are states, } j \text{ differs from } i \text{ at most over } x, w \text{ and } e, \text{ and the values assigned by } j \text{ to } x, w \text{ and } e \text{ satisfy } [\text{engine}(x)], [\text{Avon}(w)], \text{ and } [e : \text{at}(x, w)] \}$

Poesio and Muskens (1997) proposed to extend the standard version of CDRT in order to allow for discourse markers of two new types: ranging over assignments, and ranging over relations between assignments (DRSS). We use that formalism here even though some technical issues concerning how to allow the second modification still have to be addressed.⁵

4 The Update Effects of Dialogue Acts

Cohen and Levesque (1990b) argued that illocutionary acts are not an essential ingredient of a theory of communication; they can be ‘defined away’ by capturing their effect in terms of intentions and beliefs. Our goals here are more modest. The axiomatisation of dialogue acts that we propose below specifies for each dialogue act the update to the conversational score that results when an occurrence of that act is recorded; e.g., what gets grounded as the result of an acknowledgment, or the attitudes that become public (i.e., the corresponding states are recorded in G) as a result of a core speech act. However, we feel it is too early to claim that the update properties we specify completely define the DRI dialogue acts, and that therefore these can be dispensed with.

The update effects are specified using the format:

Name:	Act
Condition on update:	Φ
Update:	Ψ

In the simplest cases, the update simply depends on the occurrence of the dialogue act being recorded in one of the DRS that constitute the CIS; in more complex cases, additional conditions on the CIS are involved. The update condition may also depend on the condition holding in a specific DRS among those that constitute the CIS: e.g., the update resulting from a core speech act being added to a discourse unit are typically different from that that results from that speech act being added to G. We use the shorthand $K::\varphi$ to specify that condition φ must hold in DRS K, meaning:

$$K::\varphi =_{def} \forall ij K(i)(j) \rightarrow \varphi(j)$$

We use the notation $X += K$ to indicate the operation of DRS update in which the value of X is updated by concatenating K to it by means of the CDRT ; operator:

$$X += K =_{def} \lambda i \lambda j X(j) = (X(i); K)$$

Finally, we use two operators for doing list manipulation, **push** and **remove**, defined in turn in terms of a concatenation operator | and a deletion operator on lists:

$$\begin{aligned} \text{push}(X, Y) &=_{def} \lambda i \lambda j X(j) = (Y|X(i)) \\ \text{remove}(X, Y) &=_{def} \lambda i \lambda j X(j) = X(i)Y \end{aligned}$$

Primitives

First of all, a brief introduction to the terminology we use to talk about events and types. We use the term **EVENTUALITY TYPE** to refer to abstracts over conditions describing events or states of type $(\epsilon, (s, t))$, such as $\lambda e. \lambda i. e(i) : \mathbf{Accept}(x(i), e'(i))$ where **Accept** is an event type, x, e and e' are discourse markers⁶ and e' is the event being accepted; or $\lambda e. \lambda i. e(i) : \mathbf{Bel}(x(i), K(i))$, where **Bel** is a state type and K a discourse marker taking values over DRSs. We also refer to event types as **ACTION TYPES**. We use the symbol α to refer to action types, and the symbol σ to refer to state types.

⁵We also assume that discourse markers become accessible in a conversation because each (core) speech act in a discourse segment extends the current **FOCUS SPACE**, as proposed in Grosz and Sidner (1986); we assume here that focus spaces are situations, which conceptually can be thought of as $(\text{world}, \text{assignment})$ pairs, and are implemented as assignments with a distinguished variable w .

⁶We recall that in CDRT discourse markers are functions from assignments to objects in the domain.

Our characterization of the effects of DAs on the CIS makes use, first of all, of the event types **Try**, **Achieve** and **Address**, informally described as follows:

- $e : \text{Try}(A, \alpha)$ means that e is an event of A trying to perform an act of type α .⁷
- $e : \text{Achieve}(A, \sigma)$ means that e is an event of A bringing about the satisfaction of state type σ .
- $e : \text{Address}(A, e')$ means that e is an event of A considering and responding to e' .

Secondly, we assume that the conversational score can include information about agents being in the state having one of the following mental attitudes:

- $s : \text{Bel}(A, K)$: s is a state of agent A believing the proposition expressed by DRS K .
- $s : \text{Int}(A, \tau)$, where τ is either an action type that agent A intends to perform or a state type that A intends to achieve.
- $s : \text{Option}(A, \alpha)$: action type α is one that A is aware that she can perform.

In addition to 'private' attitudes such as **Bel** and **Int**, which are traditional ingredients of formalizations of speech acts (Allen, 1983; Cohen and Levesque, 1990b), our formalization also relies on some *social* attitudes, which relate an agent not only to a course of events or action, but also to a social group. These include:

- $s : \text{SCCOE}(A, B, K)$: this stands for Socially Committed to a Course of Events. It is the public counterpart to individual belief. It means that A is committed to B to K being the case (whether or not she actually privately believes it).⁸
- $s : \text{Obliged}(A, B, \alpha)$ state s is one of A having the obligation to B to perform an act of type α (whether or not she actually intends to) Traum and Allen (1994b).

Typically these states cease to hold after a while, either because e.g., the obligation has been addressed, or because an intention has been dropped. Current states are those whose associated time interval $\varphi(s)$ properly contains the indexical time point **now** ($\text{now} \subseteq \varphi(s)$, in Muskens' notation); some of the acts below update the temporal duration of some of these states making them not current anymore.

Locutionary Acts

As a new utterance is perceived, the current discourse unit is updated with the corresponding locutionary act. This update rule specifies a sort of default co-presence assumption - everything that gets uttered is by default recorded as part of the conversational score. This case differs from the others in that there are no update conditions—the act is not recorded anywhere in the CIS prior to this update. u is a new discourse marker.

Name:	Utter
Condition on update:	
Update:	$\text{CDU} += [u u : \text{Utter}(A, "...")]$

⁷**Try** expresses the notion of present-directed intention (Cohen and Levesque, 1990a, pg. 35) and is related to Cohen and Levesque's **ATTEMPT**.

⁸A default inference can generally be drawn in the case of an honest agent between SCCOE and actual belief, as follows:

$$\forall a, b, K, s, i \quad [s : \text{SCCOE}(a, b, K)](i) \Rightarrow \exists s' \quad [s' : \text{Bel}(a, K)](i)$$

For example, as soon as the first word in (4a), *There*, is perceived, the update to the CDU in (7a) takes place; assuming that the initial constituents of the CIS are empty, and the initial CDU is DU1, the result is the CIS in (7b), where *u1* is a new discourse marker. The same update takes place after each locutionary act.⁹

(7) a. $CDU += [u1|u1 : Utter(A, "There")]$

$G \text{ UDUS } CDU \text{ DU1}$

b.

$G = []$
$DU1 = [u1 u1 : Utter(A, "There")]$
$CDU = DU1$
$UDUS = \langle DU1 \rangle$

Core Speech Acts

A fundamental property of core speech acts is that they impose an obligation on the responder to perform an **Understanding-act** (e.g., acknowledge them) when she recognizes their occurrence. Let *F* be any core speech act, with arguments *A, B, x̄*; then the occurrence of an action of that type in *K* (*G* or a *DU*) results in the following update :

Name:	F
Condition on update:	$K::[e : F(A, B, \bar{x})]$
Update:	$K += [s s : Obligated(B, \lambda s'.s' : Understanding-act(B, e))]$

Some forward-looking actions also impose an obligation on the responder to address them. This is certainly the case for **directives** – arguably, it holds for **Statements** and **Offers**, as well). Let *D* be a forward-looking action of this class, with arguments *A, B, x̄*; then its occurrence in *K* (*DU* or *G*) results in the following update:

Name:	D
Condition on update:	$K::[e : D(A, B, \bar{x})]$
Update:	$K += [s s : Obligated(B, \lambda s'.s' : Address(B, e))]$

The specific update effects for some of the forward-looking acts are shown in Table 1. These formalizations are fairly direct implementations of the specifications in Allen and Core (1997). As mentioned above, we assume that the occurrence of an act such as **Assert** that specializes another act (**Statement**) results in the updates associated both with the more general and with the more specific act.

noindent What distinguishes an assertion from a garden variety statement is the intention to get the responder to believe the claim (one could make a statement in the case where one knows the responder won't believe it, or already believes it). But the achievement of that belief (a successful assertion) is too strong a condition, that defines the perlocutionary act of **convince**: that only results if the responder explicitly **Accepts** the act, which results in a further inference because of the conditional originated from the update. For an example of the consequences of an **Assert**, consider again (4a). The situation after all of the locutionary acts have been processed is as in (8) (we omit here all information about the locutionary acts derived from incremental syntactic and semantic interpretation):

⁹This is actually a simplification, in reality one often can't tell to which DUs various parts of an input utterance will belong. In an extended version of this paper, we will give more details on how to handle this kind of update using *continue* grounding acts to merge new input with existing DUs. For now, the assumption that all new material from a current utterance gets put into the CDU will suffice.

Name:	Statement
Condition on update:	$G::[e : \text{Statement}(A,B,K)]$
Update:	$G += [s s : \text{SCCOE}(A,B,K)]$
Name:	Assert
Condition on update:	$G::[e : \text{Assert}(A,B,K)]$
Update:	$G += [e^1e^1 : \text{Try}(A,\lambda s'.s' : \text{Bel}(B,K)),$ $[e^2e^2 : \text{Accept}(B,e)] \Rightarrow [s s : \text{SCCOE}(B,A,K)]]$
Name:	Influencing-addressee-future-act
Condition on update:	$G::e : \text{IAFutA}(A,B,\lambda e'.e' : \varphi)$
Update:	$G += [s s : \text{Option}(B,\lambda e'.e' : \varphi)]$
Name:	Open-option
Condition on update:	$G::[e : \text{OpOp}(A,B,\lambda e'.e' : \varphi)]$
Update:	$G += [[\neg[e^1e^1] : \text{Try}(A,\lambda s'.s' : \text{Achieve}(A,\lambda e'.e' : \varphi))]]$
Name:	Directive
Condition on update:	$G::[e : \text{Dir}(A,B,\lambda e'.e' : \varphi)]$
Update:	$G += [[e^1e^1 : \text{Accept}(B,e)] \Rightarrow [s s : \text{Obliged}(B,A,\lambda e'.e' : \varphi)]]$
Name:	Committing-speaker-future-action
Condition on update:	$G::[e : \text{CSFA}(A,B,\lambda e'.e' : \varphi)]$
Update:	$G += [s s : \text{Option}(A,\lambda e'.e' : \varphi)]$
Name:	Commit
Condition on update:	$G::[e : \text{Commit}(A,B,\lambda e'.e' : \varphi)]$
Update:	$G += [s s : \text{Obliged}(A,B,\lambda e'.e' : \varphi)]$
Name:	Offer
Condition on update:	$G::[e : \text{Offer}(A,B,\lambda e'.e' : \varphi)]$
Update:	$G += [[e^1e^1 : \text{Accept}(B,e)] \Rightarrow [s s : \text{Obliged}(A,B,\lambda e'.e' : \varphi)]]$

Table 1: Forward-Looking Act Definitions

(8)

G	$UDUS$	CDU	$DU1$
$G = []$			
$u1 \dots u6$			
<hr/>			
$DU1 =$	$u1 : \text{Utter}(A, \text{"There"})$		
	$\dots\dots$		
	$u6 : \text{Utter}(A, \text{"Avon"})$		
$CDU =$	$DU1$		
$UDUS =$	$\langle DU1 \rangle$		

In the meantime, intention recognition takes place. Assuming that the utterance unit consisting of $u1 \dots u6$ is interpreted as an **Assert**, the following update of the CDU takes place (the occurrence of other acts such as a **Release-turn** and perhaps other core speech acts are also possibly inferred):

$ce1$	$K1$	s'
<hr/>		
$ce1 :$	$\text{Assert}(A,B,K1)$	
$x \ w \ e$		
<hr/>		
$K1 =$	$\text{engine}(x)$	
	$\text{Avon}(w)$	
	$e :$	$\text{at}(x,w)$
$K1(s)(s')$	$\text{generate}(u1-6,ce1)$	

(we have glossed over how precisely the semantic interpretation of the utterance unit is computed – see Poesio and Traum (1997) for some details). This update results in an obligation to signal understanding or misunderstanding with respect to $ce1$ and (possibly) in an obligation to address it, which result in the following update of the conversational score:

$$(9) \quad \text{CDU} += \frac{s1 \ s2}{\begin{array}{l} s1 : \text{Obliged}(B, \lambda s'. s' : \text{Understanding-Act}(B, ce1)) \\ s2 : \text{Obliged}(B, \lambda s'. s' : \text{Address}(B, ce1)) \end{array}}$$

As we mentioned earlier and we will see in more detail shortly, acknowledging a DU has the effect of updating G with the information in that DU. The utterance in (4b), *Okay*, has a dual purpose: it serves as an acknowledgment of *ce1*, as well as an acceptance. The acknowledgment leads to the occurrence of *ce1* being grounded, which in turn leads to the updates associated with an **Assert** act according to Table 1, namely:

$$G += \frac{s3 \ e1}{\begin{array}{l} s3 : \text{SCCOE}(A, B, K1) \\ e1 : \text{Try}(A, \lambda s'. s' : \text{Bel}(B, K1)) \end{array}}$$

The crucial property of backward-looking acts is that they remove the obligation to address an act. If **F** is a forward-looking act and **B** an 'appropriate' backward-looking act,¹⁰ then performing an act *e* of type **B** with respect to the occurrence of act *e'* of type **F** removes the obligation:

Name:	B
Condition on update:	$K :: [s : \text{Obliged}(B, \lambda s'. s' : \text{Address}(B, e'))$ $\varphi(s) = t$ $e : \text{B}(B, e')]$
Update:	$K += [t t < \text{now}]$

The specific updates resulting from backward-looking acts are described in Table 2.

Name:	Agreement
Condition on update:	$G :: [e : \text{Agreement}(A, ce)]$
Update:	-
Name:	Accept (Assert)
Condition on update:	$G :: [e : \text{Accept}(A, ce)]$
Update:	[effect specified by conditional in update for Assert]
Name:	Accept (Request)
Condition on update:	$G :: [e : \text{Accept}(A, ce), ce : \text{Directive}(B, A, \alpha)]$
Update:	[work done by conditional]
Name:	Reject
Condition on update:	$G :: [e : \text{Reject}(A, ce)]$
Update:	$G += [s \neg s : \text{SCCOE}(A, B, K)]$

Table 2: Backward-Looking Act Definitions

The DRI scheme includes a single **Accept** act that may be used to address acts of different types; we hypothesize that the act-specific consequences of acceptance are part of the definition of the forward acts themselves, in the conditionals introduced as the result of the performance of the forward acts. Thus, for example, B's acceptance of *ce1* by the *Okay* in (4a) leads to G being updated with the information $[s | s : \text{SCCOE}(B, A, K1)]$.

Grounding Acts

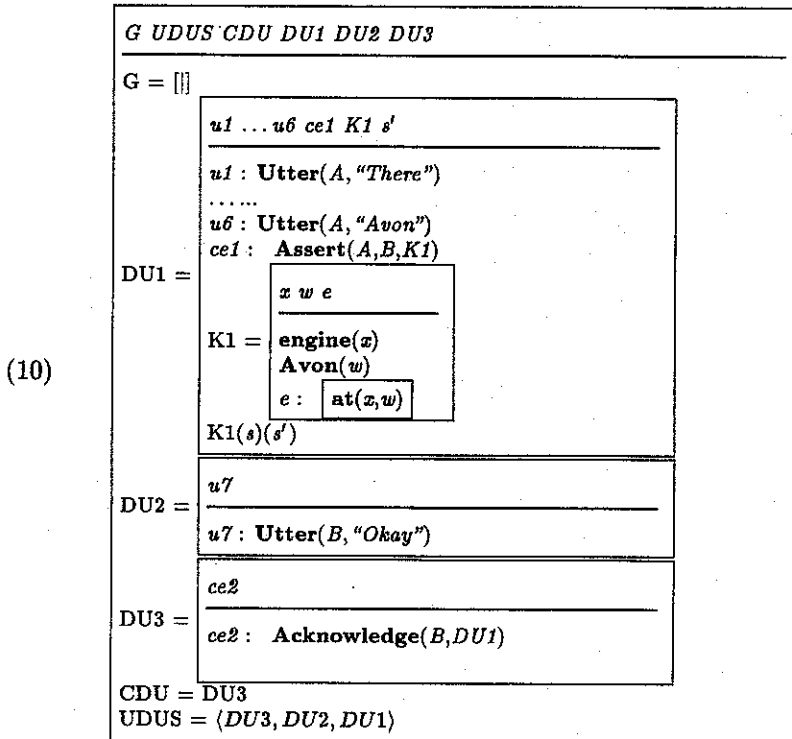
Of the grounding acts, we only consider here **Acknowledge**, that we treat as a predicate *ce* : **Acknowledge**(*A, DU1*) relating a CP *A* to a DU *DU1*. The occurrence of an acknowledgment of *DU1* results in G being updated with that discourse unit, which is then removed from UDUS.

¹⁰We will leave the notion of what counts as 'appropriate' unspecified – see Ginzburg (1995) for the case of questions.

Grounding acts do not seem to ever get added to G; we hypothesize that they are included in their own DUs that also get removed after they update the conversational score.¹¹

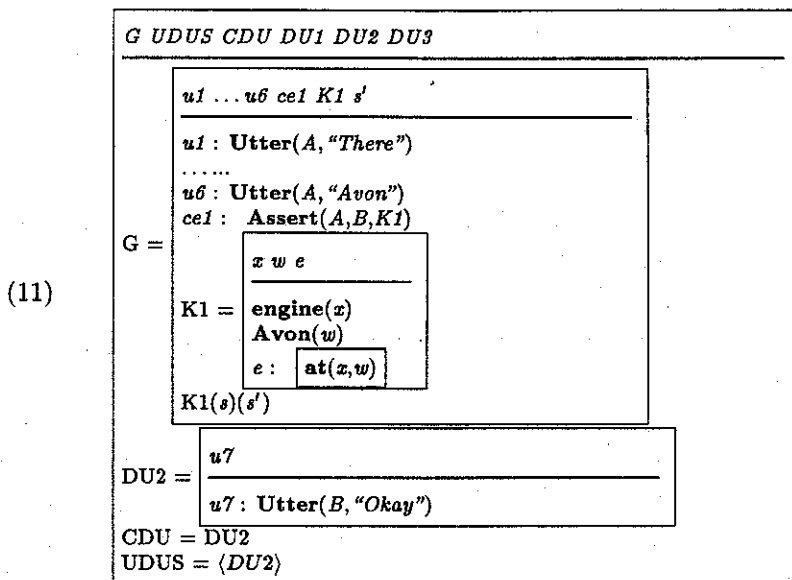
Name:	Acknowledge
Condition on update:	CDU::[<i>u</i> : Acknowledge(<i>A</i> , <i>DU1</i>)]
Update:	G += <i>DU1</i> ;remove(<i>UDUS</i> , <i>DU1</i>);remove(<i>UDUS</i> , <i>CDU</i>)

For example, we hypothesize that the *Okay* in (4b) works as follows. At the end of the first turn in (4) the CIS is as after the update in (9). As the turn is taken by B a new DU is initiated, DU2. (This is an effect of the implicit **Release-turn** performed at the end of (4)a.) The locutionary act *u7* of uttering *Okay* is added to DU2, as is the **Acknowledge** act *ce2* as soon as it is recognized. The result is the situation in (10).



At this point, as a result of the acknowledgment, G is updated with DU1, DU1 and DU3 are removed from UDUS, and DU2 becomes CDU:

¹¹The importance of grounding acts is not that they occur and are objects of discussion, but their effect on restructuring parts of CIS.



5 Conclusions and Future Work

To summarize, we have proposed a preliminary characterization of the DRI acts in terms of a notion of conversational score that takes into account the fact that the construction of the common ground is not an immediate process, and in which the functions of both forward-looking and backward-looking acts can be characterized, if in a preliminary fashion. The inclusion of accounts of these processes, as well as of the process of introduction and removal of social obligations, is the main differences between our proposal and previous formalizations of speech acts such as Cohen and Levesque. More detailed comparisons will only be possible with a more detailed analysis of the basic building blocks.¹²

The current formalization doesn't cover yet some of the DRI acts; of these, 'partial' acts such as **Accept-part** may be particularly tricky to handle since this may involve some notion of structured propositions. And anyway we have concentrated until on clarifying the mechanism by which the conversational score is put together, rather than the attitudes expressed by the acts. We also plan to look at task-dependent actions - i.e., those actions that explicitly manipulate objects related to the task (such as the plan in the TRAINS conversations). Such actions are currently not included in the DRI scheme.

References

- Allen, J. F. (1983). Recognizing intentions from natural language utterances. In Brady, M. and Berwick, R. C., editors, *Computational Models of Discourse*. The MIT Press.
- Allen, J. and Core, M. (1997). DAMSL: Dialogue act markup in several layers. Draft contribution for the Discourse Resource Initiative.
- Austin, J. L. (1962). *How to Do Things with Words*. Harvard University Press, Cambridge, MA.
- Bunt, H. C. (1995). Dialogue control functions and interaction design. In Beun, R., Baker, M., and Reiner, M., editors, *Dialogue in Instruction*, pages 197-214. Springer Verlag.
- Carlson, L. (1983). *Dialogue Games*. D. Reidel, Dordrecht.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge.

¹²CDRT actually incorporates many ideas from dynamic logics—see Muskens (1995) for discussion.

- Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13:259 – 94.
- Cohen, P. R. and Levesque, H. J. (1990a). Persistence, intention and commitment. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communication*, chapter 12. Morgan Kaufmann.
- Cohen, P. R. and Levesque, H. J. (1990b). Rational interaction as the basis for communication. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communication*, chapter 12, pages 221–256. Morgan Kaufmann.
- Core, M. G. and Allen, J. F. (1997). Coding dialogs with the DAMSL scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA. AAAI.
- Davidson, D. (1967). The logical form of action sentences. In Rescher, N., editor, *The Logic of Decision and Action*, pages 81–95. University of Pittsburgh Press, Pittsburgh.
- Di Eugenio, B., Jordan, P. W., Thomason, R. T., and Moore, J. D. (1997). Reconstructed intentions in collaborative problem solving dialogues. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA. AAAI.
- Discourse Resource Initiative (1997). Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.
- Ginzburg, J. (1995). Resolving questions, ii. *Linguistics and Philosophy*, 18(6):567–609.
- Goldman, A. (1970). *A Theory of Human Action*. Princeton University Press, Princeton, NJ.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. D. Reidel, Dordrecht.
- Kowtko, J. C., Isard, S. D., and Doherty, G. M. (1992). Conversational games within dialogue. Research Paper HCRC/RP-31, Human Communication Research Centre.
- Levin, J. A. and Moore, J. A. (1978). Dialogue games: Metacommunication strategies for natural language interaction. *Cognitive Science*, 1(4):395–420.
- Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359.
- Michael E. Bratman, D. J. I. and Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. Technical Report TR425R, SRI International. Appears in *Computational Intelligence*, Vol. 4, No. 4, 1988.
- Muskens, R. (1995). Tense and the logic of change. In Egli, U., Pause, P., Schwarze, C., von Stechow, A., and Wienold, G., editors, *Lexical Knowledge in the Organization of Language*, pages 147–183. John Benjamins, Amsterdam / Philadelphia.
- Novick, D. (1988). *Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model*. PhD thesis, University of Oregon. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.
- Poesio, M. (1997). Underspecified interpretations and a theory of language processing. Submitted.
- Poesio, M. and Muskens, R. (1997). The dynamics of discourse situations. In Dekker, P. and Stolhof, M., editors, *Proceedings of the 11th Amsterdam Colloquium*. University of Amsterdam, ILLC.

- Poesio, M. and Traum, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309-347.
- Stalnaker, R. (1979). Assertion. In Cole, P., editor, *Syntax and Semantics*, volume 9, pages 315-332. Academic Press.
- Traum, D. R. (1994). *A Computational Theory of Grounding in natural language conversation*. PhD thesis, University of Rochester, Department of Computer Science, Rochester, NY.
- Traum, D. R. and Allen, J. F. (1994a). Discourse obligations in dialogue processing. In *Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics*, pages 1-8.
- Traum, D. R. and Allen, J. F. (1994b). Discourse obligations in dialogue processing. In *Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 1-8, New Mexico.
- Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3). Special Issue on Non-literal Language.

Making DRT Suitable for the Description of Information Exchange in Dialogue

Mieke Rats*

Agnes van Leeuwenberchstraat 18
3515 AZ Utrecht, The Netherlands

++31 30 2735100†

mieke.rats@wxs.nl

Abstract

In my paper, I will take the position that although the construction algorithm of DRT produces intuitively appealing representations of the meaning of small sequences of sentences, it needs much more refinement for being able to deliver adequate representations of information exchange in dialogue. I will show some of its inadequacies, present an alternative strategy for dialogues, and point to the consequences for semantic interpretation. The analysis will be based on a thorough study of information exchange in two corpora of naturally occurring information dialogues, one about flight information and one about train time table information.

1 The DRT-algorithm and its Intuitive Attractiveness

Discourse Representation Theory (DRT) (Kamp and Reyle, 1993) describes how semantic representations follow from small sequences of sentences. The construction of these representations, which are called Discourse Representation Structures (DRS's) proceeds incrementally. When the so-called DRT construction algorithm is applied to a sequence of sentences $S_1 \dots S_n$, it first incorporates S_1 into an empty DRS, then it incorporates S_2 into the DRS resulting from S_1 , S_3 into the DRS resulting from S_1 and S_2 , etcetera, until all sentences are interpreted.

The starting point of the construction of a DRS is a syntactic tree. This tree is translated subtree by subtree into a DRS. The translation is done by so-called *construction rules*, which take each time the highest subtree and infer discourse referents and conditions out of it until the whole tree is decomposed. The DRS of the first sentence of a passage is included into an empty DRS, as if there is no information structure into which new information is to be integrated. The DRS of each following sentence is integrated into the DRS of the previous sentence(s).

The advantage of the incremental approach is that it makes it possible to link pronouns with their antecedents. This is only possible when the representation of an utterance containing a pronominal reference is interpreted in relation to the representation of the previous context. To determine the meaning of utterances in dialogue, an incremental approach is also needed, since in dialogue, a lot of context dependent expressions occur.

In particular, elliptic utterances play an important role in the information exchange. Corpus analysis of naturally occurring dialogues (Rats, 1997b; Rats and Bunt, 1997), shows that speakers are very efficient in using their language. Once the context of conversation is set, they will not express more than is really needed to make their point. See, for instance, utterance 8,9, and 11

*I would like to thank Kees Vermeulen for his useful comments on an earlier version of the paper.

†Until the end of december the author has worked in a research project that aimed to build a dialogue manager for a train time table information system at Knowledge-based Systems, Technical Computer Science, Delft University of Technology. Now the project has stopped, she is looking for a new research position in a comparable research field.

in the dialogue (1) of the Schiphol corpus. The answer by utterance 8, the check by utterance 9, and the correction by utterance 11, are elliptic utterances that contain only that information that the speaker wants to give, check, and correct, in the representation that he supposes the other speaker has constructed after utterances 6 and 7.

(1)	6	C:	voor een eh intercontinentale vlucht	6	C:	for an uh intercontinental flight
	7		moet ik daar een uur of twee uur van te voren aanwezig zijn?	7		do I have to be present one or two hours in advance?
	8	I:	Twee uur van te voren	8	I:	Two hours in advance
	9	C:	Een uur van te voren?	9	C:	One hour in advance?
	10	I:	Nee,	10	I:	No,
	11		twee uur	11		two hours
	12	C:	Oke	12	C:	Okay

Another strong point of the DRT algorithm is that it produces intuitively appealing representations of meaning. Looking at spoken discourse, it seems completely natural to describe the meaning of sentences in terms of entities and conditions. In elaborate corpus research of 111 spoken dialogues recorded at the information service of Schiphol Rats (1996b) has shown that each information exchanging utterance in this corpus can be ascribed a topic-comment (entity-condition) structure. The topic-comment structures of the individual utterances can be combined to form topic-comment structures for whole dialogue fragments where the topics form the connecting threads. Dialogue (2) illustrates this.

(2)	1	I:	Inlichting Schiphol	1	I:	Schiphol Information
	2	C:	Ja,	2	C:	Yes,
	3		u spreekt met de Wijn	3		you are speaking with de Wijn
	4		Vlucht KL 550,	4		Flight KL 550,
	5		hoe laat is die gepland?	5		for what time is it scheduled?
	6	I:	Die wordt nu definitief verwacht om vijf voor twaalf	6	I:	It is now definitely expected at five to twelve
	7	C:	Vijf voor twaalf?	7	C:	Five to twelve?
	8	I:	Ja hoor	8	I:	Yes
	9	C:	Oké	9	C:	Okay,
	10		Bedankt	10		Thank you
	11	I:	Tot uw dienst	11	I:	You're welcome
	12	C:	Dag	12	C:	Goodbye
	13	I:	Dag	13	I:	Goodbye

When the information exchange starts, after some dialogue opening acts, in utterance 4, the client introduces a common starting point *Flight KL 550*. This entity serves as a point of attachment for the question that follows: it is the entity about which *for what time is it scheduled* is asked. The answer of the information service takes up this same entity to provide the requested information, as do the reconfirm and the acknowledgement that follow in utterances 7 and 8.

The introduction of a certain entity and the continuation of that entity, is the work of topic management in naturally occurring dialogues (Rats, 1996b). It regulates that information is exchanged in such an orderly and understandable way that it is clear for both speaker and hearer what piece of new information is exchanged about what entity. Topic management regulates that the new information in an utterance will be connected with an entity introduced in a preceding utterance. If there is no preceding discourse or if a new discourse fragment has to be opened, it regulates the introduction of a new entity.

Dialogue fragment (3) shows a topic shift. In the first part, information is exchanged about the next plane to Dublin. Utterance 6 introduces the entity into the discourse domain. Utterance 7 asks information about it, utterance 8 gives the information, utterance 9 reconfirms the

information, and utterance 10 acknowledges the information. Then the information exchange is settled and a new information exchange can be opened, which is done by the introduction of a new discourse entity.

(3)

5	C:	...	5	C:	Could you tell me
6		Zou je mij kunnen zeggen	6		the next plane from Dublin
		het eerstvolgende vliegtuig uit			
		Dublin			
7		wanneer dat aankomt?	7		when will it arrive?
8	I:	Dat is vanavond pas om twintig	8	I:	That is this evening only at
		over zeven			twenty past seven
9	C:	Negentien uur twintig	9	C:	Nineteen hours twenty
10	I:	Ja	10	I:	Yes
11	C:	Ja,	11	C:	Yes
12		want eh.. het voorlaatste was	12		because eh... I suppose the
		zeker die een			penultimate was that one
13		die om kwart voor twaalf	13		which arrived at a quarter to
		aankwam			twelve
14	I:	Juist, ja ja	14	I:	Right, yes yes
		...			

Until now, DRT was only concerned with small sequences of sentences, not really with information exchange in natural occurring dialogue, and not really with topic management, the regulation of discourse entities to be talked about. Nevertheless, it seems a good theory for extension because it seems to start from the right intuitions about meaning. It can deal with context dependency and it is a good starting point for incorporating of the topic-comment distinction in the semantic representation.

2 Problems with the DRT-algorithm

The algorithm has at least two problems. The first is its assumption that initial sentences are integrated into an empty context and that next utterances are solely integrated into a DRS of previous utterances. Precise discourse analysis (Gundel, 1985; Clark and Wilkes-Gibbs, 1986; Gundel, 1988; Rats, 1996b, 1997b; Rats and Bunt, 1997) shows that this is not exactly how it works in discourse and especially not in dialogue. When people communicate, they always assume that their discourse partner has a certain amount of background knowledge with which the information they convey need to be combined. This holds for initial sentences, but also for the incorporation of next sentences.

The second problem of the algorithm is that the end DRS is tacitly assumed to be the common ground for the discourse participants. Each sentence of the discourse simply *adds* information to the common ground. Precise analysis of naturally occurring dialogues shows, however, that utterances in dialogue may aim to bring about other changes like deleting, substituting, confirming, and disconfirming information. In fact, the common ground in dialogue is the result of a stepwise working together of the two participants, where the current speaker tries to make sure he or she is being understood, while the other tries to let the speaker know if he has succeeded (Clark and Schaefer, 1989; Allwood et al., 1992; Traum and Allen, 1992; Rats, 1997b; Rats and Bunt, 1997).

Background Knowledge

There are at least two forms of conversational behaviour in naturally occurring dialogue that show that an interpreter uses his foreknowledge for constructing the meaning and giving the right interpretation for the first information exchanging utterances in the conversation.

Corpus research on the Schiphol corpus (Rats, 1996b, 1997b; Rats and Bunt, 1997) has shown that speakers often apply (in about 50% of the cases) a left-dislocation construction when they

want to introduce the topic of the information exchange. In all cases, the left-dislocated NP introduces a topic into the dialogue, that remains the topic of the utterances that follow until the end of the information exchange. The left-dislocated NP is always a definite NP, most of the time a complex NP with relative clauses, appositions, and prepositional phrases, that give extra information about the referent. Dialogue (4) from the Schiphol corpus illustrates this. Dialogue (5) shows that the same phenomenon occurs in the train corpus.

- | | | | | | | |
|-----|---|----|----------------------------------|---|----|--------------------------------|
| (4) | 1 | C: | Weet u ook | 1 | C: | Do you know |
| | 2 | | of het toestel | 2 | | if the plane |
| | 3 | | dat eh.. van de AL Italia | 3 | | that uh.. of the AL Italia |
| | 4 | | dat is vertrokken uit Rome om | 4 | | that has departed from Rome at |
| | | | tien over tien | | | ten past ten |
| | 5 | | of dat al binnen is? | 5 | | if that has arrived yet? |
| (5) | 1 | C: | ik wilde even weten | 1 | C: | I just wanted to know |
| | 2 | | de laatste trein van Arnhem naar | 2 | | the last train from Arnhem to |
| | | | Den Bosch vanavond | | | Den Bosch this evening |
| | 3 | | hoe laat vertrekt die? | 3 | | at what time will it leave? |
| | 4 | I: | drieëntwintig uur eenenvijftig | 4 | I: | twenty-three hours fifty-one |
| | 5 | C: | dat is de laatste? | 5 | C: | that is the last ? |
| | 6 | I: | ja | 6 | I: | yes |

By introducing the referent by a definite NP the speaker shows that he supposes the information service already has knowledge about the referent he wants to talk about. By introducing it in a left- dislocation construction, in a separate utterance, the speaker asks special attention for the referent. The other speaker is requested to look in his background knowledge and find the entity. To ensure that he would really be able to find the unique referent, the speaker adds as much information as he thinks is important. After this is done, the speaker asks his question about it, as if he assumes that they must first agree on the entity before the other speaker will be able to answer a question about it.

The identification of the topic is an important issue in both corpora. There is another phenomenon that indicates that. It occurs in dialogues, where the entity cannot be easily found. In such a case, the information service starts a clarification sequence to get at agreement about the entity, before she considers the question about it. See for instance dialogue (6) .

- | | | | | | | |
|-----|----|----|------------------------------------|----|----|------------------------------------|
| (6) | 4 | C: | ik wilde vragen | 4 | C: | I wanted to ask |
| | 5 | | hoe laat het vliegtuig uit Malaga, | 5 | | at what time the plane from |
| | | | | | | Malaga, |
| | 6 | | dat vanmiddag om half twee zou | 6 | | that would leave this afternoon |
| | | | vertrekken, | | | at half past two, |
| | 7 | | aankomt | 7 | | will arrive |
| | 8 | | Het vluchtnummer is niet precies | 8 | | The flight number I don't know |
| | | | bekend.. | | | exactly.. |
| | 9 | | of althans.. | 9 | | although at least.. |
| | 10 | I: | Welke maatschappij? | 10 | I: | Which company? |
| | 11 | C: | Wat? | 11 | C: | What? |
| | 12 | I: | Welke maatschappij? | 12 | I: | Which company? |
| | 13 | C: | Ja | 13 | C: | Yes |
| | 14 | | dat weet ik ook niet | 14 | | that I don't know either |
| | 15 | | Ik weet alleen | 15 | | I only know |
| | 16 | | dat ie uit Malaga zou vertrekken | 16 | | that it should leave from Malaga |
| | | | om half twee | | | at half past two |
| | 17 | I: | Ja, | 17 | I: | Yes, |
| | 18 | | maar eh.. is het een charter- | 18 | | but uh.. is it a charter flight, a |
| | | | vlucht, een lijnvlucht? | | | scheduled flight? |

In this dialogue, the client doesn't give the right information to enable the information service to find the right referent. In such a case, the information service will first start some clarification sequences to get the referent right, before she goes on answering the question. Dialogue (7) from the train corpus shows the same principle. Here, the information service must first be certain about the information about the train connection, before she is willing to answer the question.

- | | | | | | | |
|-----|----|----|---|----|----|---|
| (7) | 7 | C: | welke treinen van Amsterdam Schiphol naar Delft Centraal rijden er op dertig mei tussen half negen 's ochtends en tien uur 's ochtends? | 7 | C: | which trains from Amsterdam Schiphol to Delft run at the 30th of May between half past eight in the morning and ten o'clock in the morning? |
| | 8 | I: | tussen half negen tot tien uur 's morgens en vanaf Schiphol naar Delft? | 8 | I: | between half past eight and ten o'clock in the morning and from Schiphol to Delft? |
| | 9 | C: | ja | 9 | C: | yes |
| | 10 | I: | Delft Centraal? | 10 | I: | Delft Centraal? |
| | 11 | C: | heel graag | 11 | C: | please |
| | 12 | I: | en op een nog [uh] op een bepaalde dag op dertig mei zei u he? | 12 | I: | and also on a [uh] on a certain day on the 30th of May you said he? |
| | 13 | C: | [uh] ja | 13 | C: | [uh] yes |
| | 14 | I: | nou er zijn verschillende mogelijkheden | 14 | I: | Well, there are several possibilities |

The examples make clear that after the first turn in the conversation, interpretation directly follows the construction of the meaning. It is the aim of the introduction of the topic, that the other speaker will immediately try to find it in his background knowledge. This specific domain knowledge needs to be available to enable the speakers to continue to the next step in the information exchange.

Dialogue (8) shows that knowledge of the specific topic becomes part of the common ground because one of the speakers is uttering it. The client introduces his knowledge with the introduction of the topic. The information service shows her knowledge about it while checking a characteristic that was taken from her knowledge of the specific topic: the arrival time.

- | | | | | | | |
|-----|---|----|--|---|----|---|
| (8) | 1 | C: | het vliegtuig Los Angeles-Amsterdam | 1 | C: | the plane Los Angeles-Amsterdam |
| | 2 | | is dat al geland? | 2 | | has it already landed? |
| | 3 | I: | de vlucht van tien over drie, bedoelt u? | 3 | I: | the flight of ten past three, you mean? |
| | 4 | C: | ja | 4 | C: | yes |

The previous examples also make clear that the speakers in these dialogues must have some general knowledge about the kind of topic they will be talking about and what kind of characteristics it will have. A flight has a departure place, a departure time, a flight number, and is performed by a certain company. A train connection has to take place within a certain time span, it has a departure place, an arrival place, and a date. Together they form a knowledge structure that speakers get available in the common ground with the introduction of a topic flight or a topic train connection. How would the speakers of dialogue (8) otherwise be able to find the referents of the NPs *flight number* and *company*? How would speakers otherwise be able to come to an agreement about their specific domain topic?

So three kinds of background knowledge come into play: the knowledge of the client of his own specific flight or train connection, the knowledge of the information service concerning the same flight or connection, and a kind of lexical knowledge about the kind of characteristics flights and train connections typically have (Bartsch, 1987).

The first two become part of the common ground during communication. The third consists of lexical knowledge structures for flights and train connections, which enable speakers to build clear representations of the meaning of their partner's utterances on the one hand. On the other hand, these knowledge structures will enable the information service to build the unique referent representations, she will need to be able to give the right answers.

These language phenomena show then, that speakers in both dialogue corpora do not start the dialogue from scratch. For understanding the exact meaning of the utterances in the dialogues at least some lexical foreknowledge need to be invoked. This holds for the first utterances, but also for utterances that follow.

Update

Research on spoken dialogue makes clear that the information exchange cannot simply be reduced to a straightforward question answer mechanism, where the first speaker just poses his question about the task domain and the second speaker gives the answer. Often, problems in the communication will occur, that need to be solved before the question is posed and the answer is given.

Nevertheless, it is possible to distinguish two phases in the information exchange, a *query phase* and an *information phase*. During the query phase, the speakers try to come to an agreement about the topic and the question of the client, while during the information phase, the speakers come to an agreement about the answer (Rats, 1996a, 1997a; Rats et al., 1997).

In each phase, the speakers each have their own role. In the query phase, it is the client, who has to introduce his topic, and who has to pose his question about it. The information service has to show that he has heard, understood, and interpreted the question, and that he has enough information to identify the topic and find the answer. In the information phase, it is the information service, who has to present the answer, while the client has to show that he has heard it, understood it, and that it is in fact the information that he needed.

A speaker may show that he has heard, understood, and interpreted the other speaker's utterance in three ways: either he utters a positive acknowledgement, or he shows positive processing by moving on to the next appropriate dialogue act, or both. Dialogue fragment (9) shows an information phase of a train dialogue in which no problems arise. An important characteristic of the information phases in these dialogues is that the information presentation has an interactive form (Rats et al., 1997; Rats, 1997c). The information service presents the travel plan in a step-wise way, giving one piece of new information with each turn. After each step in the information presentation, the caller shows that he has processed the step by an acknowledgement. Dialogue fragment (9) shows that positive acknowledgements have different linguistic forms. The most simple form is an *yes*, but positive acknowledgements also repeat part of or an entire partner's previous utterance.

(9)	1	I:	dat wordt[u:h] acht uur negentien vertrekken uit Hilversum	1	I:	that will be[u:h] depart from Hilversum at eight hours nineteen
	2	C:	acht uur ja	2	C:	eight hours yes
	3	I:	en dan overstappen Amsterdam	3	I:	and then change Amsterdam
	4	C:	[u:h] ja	4	C:	[u:h] yes
	5	I:	vertrek Amsterdam acht vijftig	5	I:	departure Amsterdam eight fifty-five
	6	C:	acht vijftig	6	C:	eight fifty-five
	7	I:	komt u negen nul drie aan Amsterdam Lelylaan	7	I:	will arrive at Amsterdam Lelylaan at nine o three
	8	C:	oké	8	C:	okay
	9		dus acht uur negentien vertrekken	9		so depart at eight hours nineteen
	10		oké	10		okay
	11		hartstikke bedankt	11		thank you very much

In Schiphol dialogue (10), the answer follows after the question. Also in this dialogue, no communication problems occur.

- | | | | | | | |
|------|----|----|------------------------------|----|----|---------------------------------|
| (10) | 1 | I: | Informatie Schiphol | 1 | I: | Schiphol Information |
| | 2 | C: | de machine uit Londen | 2 | C: | the plane from Londen |
| | 3 | | KL 128 | 3 | | KL 128 |
| | 4 | | hoe laat wordt die verwacht? | 4 | | at what time is it expected? |
| | 5 | I: | dan moet ik even kijken | 5 | I: | then I just have to look |
| | 6 | | een ogenblikje | 6 | | a moment |
| | 7 | I: | die komt om drie uur binnen. | 7 | I: | it will arrive at three o'clock |
| | 8 | C: | dank u wel | 8 | I: | thank you very much |
| | 9 | I: | tot uw dienst | 9 | I: | you are welcome |
| | 10 | | dag | 10 | | good bye |

Communication problems do occur, however. In such cases, speakers have shown to perform the following dialogue acts:

1. **Reconfirm.** A speaker expresses a reconfirm, if he is not sure that he has heard the other speaker's utterances well and he wants him to repeat, or to confirm previously given information. See, for instance, utterance 6 in the information phase of Schiphol dialogue (11).

- | | | | | | | |
|------|---|----|---------------------------------|---|----|--------------------------------------|
| (11) | 1 | C: | eh.. het KLM vliegtuig | 1 | C: | uh.. the KLM plane |
| | 2 | | dat om zes uur in New York | 2 | | that will depart at six o'clock |
| | | | vertrekt | | | from New York |
| | 3 | | hoe laat komt dat aan | | | at what time will it arrive tomorrow |
| | | | morgenochtend op Schiphol? | | | morning at Schiphol? |
| | 4 | I: | om zes uur | 4 | I: | at six o'clock |
| | 5 | | dat komt aan om vijf over zeven | 5 | | it will arrive at five past seven |
| | 6 | C: | Vijf over zeven? | 6 | C: | five past seven? |
| | 7 | I: | Ja | 7 | I: | yes |

2. **Question.** A speaker will apply a question, if he needs extra information. See, for instance, utterance 5 in the query phase of train dialogue (12).

- | | | | | | | |
|------|---|----|----------------------------------|---|----|-----------------------------------|
| (12) | 1 | C: | ik wilde graag even weten hoe | 1 | C: | I would like to know |
| | | | laat de trein vanuit Ermelo naar | | | |
| | | | Amsterdam gaat via Amersfoort | | | |
| | 2 | | hoe laat de trein vanuit Ermelo | 2 | | at what time will the train leave |
| | | | naar Amsterdam gaat | | | from Ermelo to Amsterdam |
| | 3 | | via Amersfoort | 3 | | via Amersfoort |
| | 4 | I: | Ermelo | 4 | I: | Ermelo |
| | 5 | | en hoe laat wilde u | 5 | | and at what time do you |
| | | | vertrekken of aankomen | | | want to depart or to arrive? |
| | 6 | C: | rond een uur of vijf vertrekken | 6 | C: | depart around five |

3. **Check.** A speaker will apply a check, if he wants to disambiguate or complete information, or if he wants to check information that he has inferred from the other speaker's utterances. See, for instance, utterance 3 in the query phase of Schiphol dialogue (13).

- | | | | | | | |
|------|---|----|-------------------------------|---|----|-------------------------------|
| (13) | 1 | C: | het vliegtuig Los Angeles- | 1 | C: | the plane Los Angeles- |
| | | | Amsterdam | | | Amsterdam |
| | 2 | | is dat al geland? | 2 | | has it already landed? |
| | 3 | I: | de vlucht van tien over drie, | 3 | I: | the flight of ten past three, |
| | | | bedoelt u? | | | you mean? |
| | 4 | C: | ja | 4 | C: | yes |

4. **Correction.** The speaker will start a correction, if he notices that the other speaker did not process his utterance as intended. A correction consists of a statement and eventually a negative acknowledgement.

- | | | | | | | |
|------|---|----|---|---|----|--|
| (14) | 1 | I: | de treinen vanuit Zaltbommel
naar Utrecht vertrekken rondom
het hele en het halve uur | 1 | I: | the trains from Zaltbommel to
Utrecht depart around the hour
and a half past an hour |
| | 2 | C: | maar hoe laat precies? | 2 | C: | but at what time exactly? |
| | 3 | I: | vanavond? | 3 | I: | tonight? |
| | 4 | C: | nee | 4 | C: | no |
| | 5 | | [uh] morgen overdag | 5 | | [uh] tomorrow daytime |
| | 6 | I: | tweeëntwintig en tweeënvijftig | 6 | I: | twenty-two and fifty-two |

The presence of these so-called *dialogue control acts* (Bunt, 1994, 1995), reconfirms, questions, checks, and corrections, in case of negative processing and positive acknowledgements in case of positive processing makes clear that each step in the communication needs in principle to be agreed on before a next step is taken. This means that the common ground of a dialogue comes into being interactively. The assumption of the DRT-algorithm, that each sentence of the discourse directly adds information to the common ground is not true for dialogue. In fact, the information content of an utterance will only be added to the common ground, when the other speaker agrees with that action. He can only agree if he has properly heard, understood, and interpreted the speaker's utterance. And when he doesn't agree, he will propose to delete the proposed addition and substitute it for something else.

3 DRT for Dialogue

Until now DRT was mainly concerned with the construction of DRSs as such. The aim is to extend the coverage in a traditional way by dealing with more linguistic phenomena like more syntactic constructions, different kinds of anaphors (Asher, 1993), and presuppositions. The sentences for which the DRSs are constructed, are most of the time thought up or taken from written text rather than taken from naturally occurring spoken dialogue. So if we want to make use of DRSs to build a common ground for dialogue, we will be confronted with two tasks. The first task will be to give DRSs a place in an interactional context, where the common ground comes into being by cooperation between the two dialogue participants and the application of background knowledge is inevitable. The second task will be to extend the DRT fragment with typical spoken language constructions. In this paper, I will confine myself to a rough proposal for the first.

Towards an Interactive Common Ground

In dialogue, the common ground comes into being structured and interactively. At the end of the information exchange, it will reflect the agreed information structure of the two dialogue participants. Precise corpus research has shown, that speakers aim to agreement in a stepwise way, where in principle each new information unit is fixed before the next is considered. When problems in communication occur they will be solved, in most of the cases, information unit by information unit.

If applied to dialogue, the integration of DRSs can not always simply be a matter of adding the DRS of the next turn to the DRS of the previous turn. When a speaker applies a reconfirm or a check, the addition will have to be agreed upon first. A speaker may also aim to cause a change in the DRS that was constructed of the previous turn. In case of a negative acknowledgement, the speaker requests his dialogue partner to delete information, while in case of a correction, he asks to perform a substitution. For all these actions, the place of change need to be clear from the presentation.

I will show how this should work by constructing a common ground of dialogue (1) given below as (15). In this dialogue a communication problem occurs about the answer. The construction will be given in figure 1 and 2.

- | | | | | | | |
|------|----|----|--|----|----|--|
| (15) | 6 | C: | voor een eh intercontinentale vlucht | 6 | C: | for an uh intercontinental flight |
| | 7 | | moet ik daar een uur of twee uur van te voren aanwezig zijn? | 7 | | do I have to be present one or two hours in advance? |
| | 8 | I: | Twee uur van te voren | 8 | I: | Two hours in advance |
| | 9 | C: | Een uur van te voren? | 9 | C: | One hour in advance? |
| | 10 | I: | Nee, | 10 | I: | No, |
| | 11 | | twee uur | 11 | | two hours |
| | 12 | C: | Oke | 12 | C: | Okay |

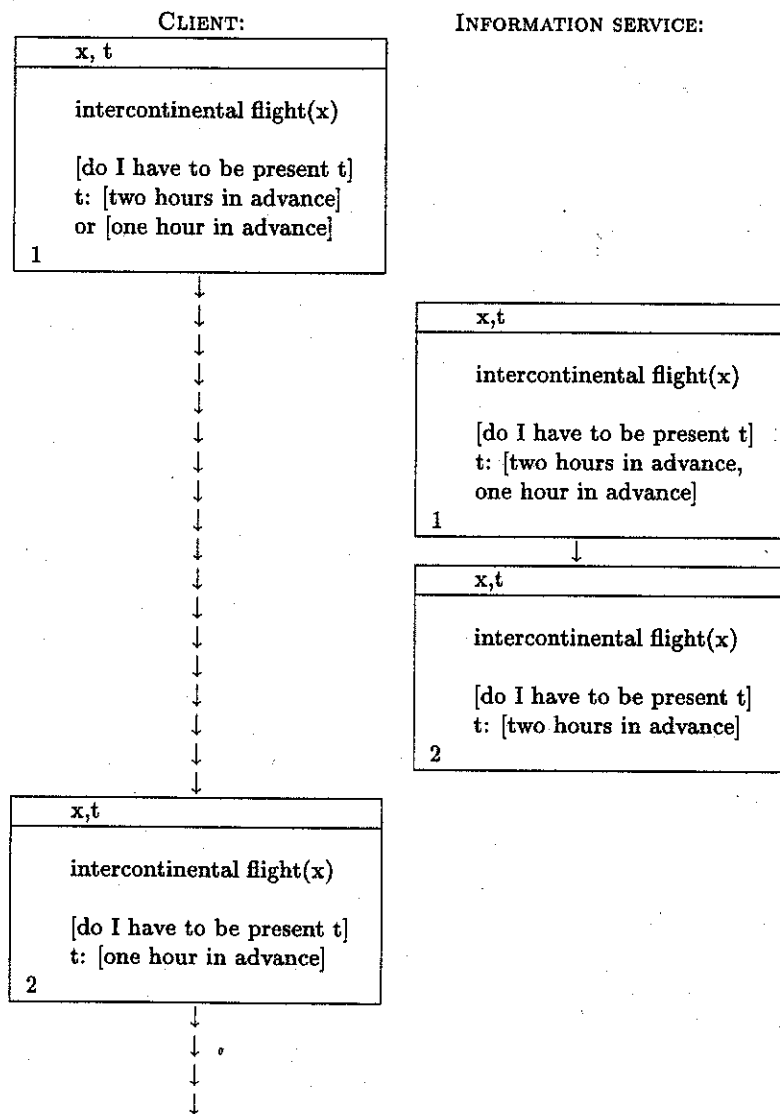


Figure 1: Information update in dialogue (15).

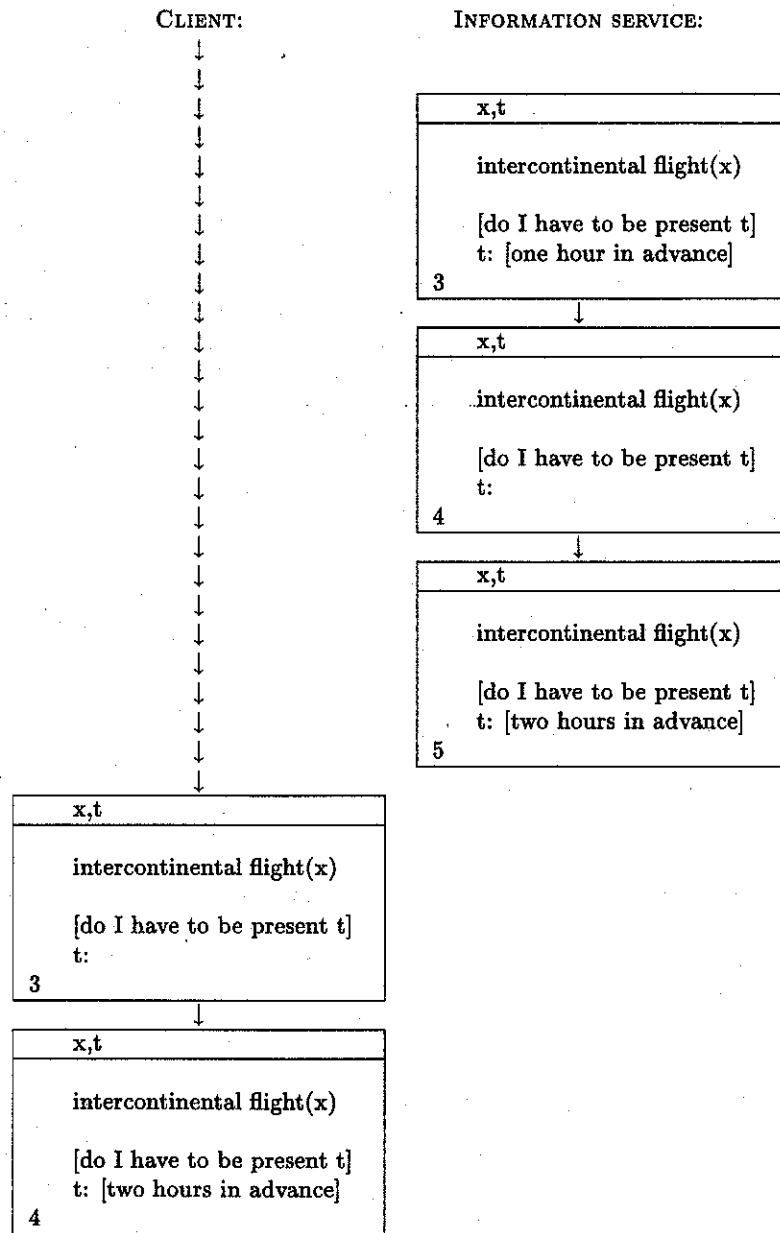


Figure 2: Continuation of the information update in dialogue (15).

The figures show the information updates from the perspective of one speaker, the information service (I). Although speakers aim at a mutually agreed common ground, it is clear that during the process of communication, there cannot immediately be one. So during communication, each speaker will have to keep a record of his partner's representation of the common ground, and the changes he proposed for it.

Utterance 6 of the dialogue in (15) introduces the topic of the information exchange, "*an intercontinental flight*". The information service is asked to evoke his background knowledge about intercontinental flights. Utterance 7 asks a question about it: if the client, has to be present one or two hours before. The information service is requested to check the items *one hour before* and *two hours before* for intercontinental flights. In Rats (1997b); Rats and Bunt (1997), this

items are called the *foci* of the question. The focus of a dialogue act is defined as the information that is asked, checked, confirmed, or corrected, dependent of the dialogue act, with respect to the topic. See also Vallduví (1990); Hoepelman et al. (1991), for the same kind of focus definition. The first DRS in the first column of figure 1 shows the update the speaker has aimed after his first turn.

The construction of this DRS roughly followed the DRT-algorithm. It is not a complete DRS, since the fragment of DRT does not provide an analysis for left-dislocation constructions and alternative questions, and to give a full analysis is beyond the scope of this paper. Instead, only the most important ingredients are given, the topical referent *an intercontinental flight* and the information elements that are asked to be checked with respect to the topic, *one hour before* and *two hours before*. The way these foci are represented differs from the way the algorithm would prescribe. They are represented as possible values of the time referent *t*. It is at this point in the representation, where the content of the dialogue control acts that follow need to be applied to.

Utterance 8 gives the answer to the question. It only expresses the focus, the information element that need to be changed in the context that was constructed during processing the preceding turn. The first and the second DRS in the second column of figure 1 shows the update aimed by the second turn. The value of *t* is changed in one of the alternatives. Since the question is immediately answered, the information service seemed to have no problems in interpreting the topical reference and the question about it.

With utterance 9, the client reconfirms the update he thinks the information service has proposed. He only expresses the focus, since the framework of interpretation is still given. The DRSs 1 and 2 of the first column in figure 1 shows the update suggested by utterance 9.

The value that is proposed by the reconfirm is wrong, so the information service performs a negative acknowledgement with utterance 10. She wants the client to take the wrong value from the representation. With utterance 11, the information service gives the correct value. Again, only the focus is expressed. The DRSs 4 and 5 in the second column show the update meant by this utterance.

The client accepts this proposal by uttering *Oke (Okay)*, which means that he will update his information state accordingly. This is shown by DRSs 2,3, and 4 of the first column in figure 2. We see now that at the end of the exchange the end DRSs of both speakers is the same, which was exactly the aim of the communication.

Integration of Background Knowledge

Our discussion in the beginning of section 2 has shown that the common ground should in principle not only consist of semantic representations that are derived from the syntactic form of the utterances. It will also have to contain representations originating from the lexicon. I will show how this could work for dialogue fragment (16) .

(16)	4	C:	ik wilde vragen	4	C:	I wanted to ask
	5		hoe laat het vliegtuig uit	5		at what time the plane from
			Malaga,			Malaga,
	6		dat vanmiddag om half twee zou	6		that would leave this afternoon
			vertrekken,			at half past two,
	7		aankomt	7		will arrive
	8		Het vluchtnummer is niet pre-	8		The flight number I don't
			cies bekend..			know exactly..
	9		of althans..	9		although at least..
	10	I:	Welke maatschappij?	10	I:	Which company?

In utterances 4,5,6, and, 7, the client introduces his topic and poses his question about it. Roughly, the DRT algorithm would lead to the following DRS:

x, y, t_1, t_2, n, z
<p> the plane(x) Malaga(y) x from y x would leave at t_1 this afternoon half past two(t_1) x arrive at t_2 $t_2 = ?$ flightnumber(n) </p>

I propose however, to incorporate background knowledge in the following way. I imagine that looking up the word *plane* in the lexicon will not only lead to the introduction of a new referent x and the condition $plane(x)$, but also referents that obey the conditions $departure\ place(x_1, x)$, $departure\ time(x_2, x)$ $arrival\ time(x_3, x)$ $flightnumber(x_4, x)$ $company(x_5, x)$.etcetera. The conditions that follow from the surface text should then be interpreted in the light of the lexical knowledge. For example, x from y will be recognized as a specification of the departure place of x , x would leave at t_1 as departure time of x etcetera. In this way, the information service will build a representation of the topic using the lexical knowledge, resulting in the DRS:

$x, x_1, x_2, x_3, x_4, x_5$
<p> the plane(x) Malaga(x_1) departure place(x_1, x) departure time(x_2, x) arrival time(x_3, x) $x_3 = ?$ flightnumber(x_4, x) </p>

Subsequently, she will try to find the flight in her database, that will obey to this representation. When she will need extra information to find the unique referent, as in our example, she will start a clarification sequence. Since it can be assumed that the client will have the same lexical knowledge, we can see that he could solve the anaphor *company* in the same way as the information service solved *flightnumber*.

Of course this is a first rough proposal for the extension of the algorithm. Lot's of details still have to be made precise in further research. But I trust that this example gives a clear indication of how I intend to procede.

4 Summary

I have shown that there are at least two problems with the DRT algorithm. The first is that it doesn't provide the opportunity to interact about the DRS-construction. The second is that it doesn't give a place to background knowledge. I propose to solve the problems by a redefinition of the algorithm. To meet the first problem, the topic and the focus of the information exchange need to given a marked place in the DRS on the one hand and other actions than simply *adding* should be allowed. To solve the second problem, lexical knowledge should become part of the DRS with the introduction of the topical referent. I gave a rough proposal of how these problems could be solved technically. Of course, much more detailed refinement is needed.

References

- Allwood, J., Nivre, J., and Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9(1):1-26.

- Asher, N. (1993). *Reference to Abstract Objects*, volume 50 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dordrecht.
- Bartsch, R. (1987). Frame representations and discourse representations. Technical report, Institute for Language, Logic and Information, Department of Philosophy, University of Amsterdam.
- Bunt, H. (1994). Context and dialogue control. *Think*, 3:19–31.
- Bunt, H. (1995). Dynamic interpretation and dialogue theory. In Taylor, M., Néel, F., and Bouwhuis, D., editors, *The Structure of Multimodal Dialogue*, volume 2. John Benjamins Publishing Company.
- Clark, H. and Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, 13:259–294.
- Clark, H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22:1–39.
- Gundel, J. (1985). 'shared knowledge' and topicality'. *Journal of Pragmatics*, 9:83–107.
- Gundel, J. (1988). Universals of topic-comment structure. In Hammond, M., Moravcsik, E., and Wirth, J., editors, *Studies in syntactic Typology*, pages 209–239. John Benjamins Publishing Company.
- Hoepelman, J., Machate, J., and Schnitzer, R. (1991). Intonational focusing and dialogue games. *Journal of Semantics*, 8:253–275.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic, Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dordrecht.
- Rats, M. (1996a). Repair in VIOS. Technical Report 96-15, Alparon, Delft University of Technology.
- Rats, M. (1996b). *Topic Management in Information Dialogues*. Ph.D. thesis, Tilburg University.
- Rats, M. (1997a). Dialogue management with limited information. Technical Report 97-02, Alparon, Delft University of Technology.
- Rats, M. (1997b). Information update in dutch information dialogues. In Landsbergen, J., Odijk, J., van Deemter, K., and van Zanten, G. V., editors, *Computational Linguistics in the Netherlands 1996, Papers from the Seventh CLIN Meeting*, pages 145–159. IPO, Center for Research on User-System Interaction.
- Rats, M. (1997c). Towards a better information presentation for vios. Technical Report 97-06, Alparon, Delft University of Technology.
- Rats, M. and Bunt, H. (1997). Information packaging in dutch information dialogues. In Dybkjaer, L., editor, *Proceedings of the Third Spoken Dialogue and Discourse Workshop*, pages 52–73. Odense: Maersk-McKinney Moller Institute for Production Technology.
- Rats, M., van Vark, R., and de Vreught, H. (1997). Information presentation for a spoken public transport information system. In *ACL/EACL-97 Workshop on Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications*, pages 106–113.
- Traum, D. and Allen, J. (1992). A "speech acts" approach to grounding in conversation. In *Proceedings International Conference on Spoken Language processing (ICSLP-92)*, pages 137–140.
- Vallduví, E. (1990). *The Informational Component*. PhD thesis, University of Pennsylvania.

Modal Subordination in Questions*

Robert A.M. van Rooy[†]
ILLC/University of Amsterdam
vanrooy@philo.uva.nl

Abstract

In this paper it is discussed how questions should change information states in dynamic semantics. The main claim is that the meaning of following utterances can be dependent on questions in a similar way as these meanings can be dependent on quantificational sentences, or more broadly, that this dependence is one of 'modal subordination'. The most important contribution of this paper, however, is to show how this dependence between questions and later utterances can be accounted for in a general and systematic way.

1 Introduction

The central idea behind dynamic semantics is that each sentence is interpreted with respect to a context of interpretation, and that this single context contains two kinds of information: (i) what is commonly assumed about the *subject matter* of conversation, and (ii) what is assumed about the *conversational situation* itself. The first kind of information is used to determine whether what is expressed is true and appropriate or not, while the latter kind of information is used to determine what was expressed by context-dependent utterances in the first place. We need to account for questions in a dynamic semantics because also for the analysis of questions we need a context that contains both kinds of information. First, a context should contain enough information about the subject matter of conversation to determine whether a question can be appropriately asked, and whether a certain answer is a partial or a complete answer to a question. Second, a context should contain enough information about the discourse especially to determine what is expressed by an answer. So the question is how we should represent a context in dynamic semantics such that it contains both of these kinds of information, and how questions should *change* the context. This is the question I will address in the next two sections. In section 4 I will generalise my proposed way of analysing questions to other constructions.

2 Questions in Dynamic Semantics

Groenendijk & Stokhof (1984, 1997) have forcefully argued that the meaning of a question is the set of its mutually inconsistent possible complete answers. As a result, the intension of a question is a set of mutually exclusive propositions thought of as the set of all alternative complete exhaustive answers to the question. This treatment allows all kinds of different interrogatives as having denotations in the same category. On the basis of this uniform treatment they can give a general characterisation of the notion of answerhood and it allows them to give a general definition of *entailment* between all kinds of interrogatives simply by inclusion of denotation (intension).

There seems to be a natural way to implement their approach towards questions in dynamic semantics: assume that information states should be modelled in terms of partitioned structures where each cell of the partition represents the complete answer to a question relevant in the

*This article owes an obvious debt to the work of Van den Berg, and of Groenendijk & Stokhof.

[†]This research is done as part of the 'Sources and Streams of Information: Presupposition and Topic-Content'-project, sponsored by the Dutch Organisation for Scientific Research (NWO).

discourse. According to this modelling, answers eliminate cells of an existing partition, while questions partition the context in a more fine grained way, and thus introduce partitions.

Unfortunately, this won't do, because partitions normally do not contain enough information to determine what is expressed by elliptical answers; the constituent answer *John and Mary* means something different as a response to *Who walks?* than as answer to the question *Who doesn't walk?*, although the partitions induced by these questions are the same.¹ To account for this difference we have to make a distinction between the above two questions, a distinction that the partitions themselves do not offer us. For this reason, Groenendijk & Stokhof (1984) make essential use of so-called *abstracts*, *n*-ary relations in intension. The chosen abstract needed to determine what is expressed by the answer is the abstract that underlies the partition induced by the question. What is essential is that these abstracts underlying the partitions contain more information than the partitions themselves.

What is expressed by an answer depends obviously on the foregoing question, but also the meaning of a *question* can be dependent on foregoing discourse, and so for two reasons. First, because what is expressed by a wh-question is normally context-dependent in the sense that in different contexts the domain over which the values of a wh-phrase vary might be different. Just as what is expressed by the assertion *Everybody went to the concert* depends on the contextually given domain of quantification, so does what is expressed by the question *Who went to the concert?* depend on a contextually given domain. The second way in which a question can be anaphorically dependent on foregoing discourse is that it can contain an explicit pronoun, as when *A* says *I bought a car* and *Q* responds by asking *How much did it cost?*

More interesting, however, is that a question can contain a pronoun that anaphorically refers back to a wh-phrase used in an earlier question by the same agent: *Who went to the party?*, and *What did he/they bring as a present?* What makes this sequence interesting from a dynamic semantics point of view is that (i) it shows a kind of *donkey effect*; the meaning of the sequence of the two wh-questions seems equivalent with the meaning of the single multiple wh-question *Who, of those that came to the party brought what present?*, and (ii) the example shows that we can anaphorically refer back with explicit pronouns to wh-phrases of a question.

But not only in a *question* can we anaphorically refer back with a pronoun to an earlier mentioned wh-phrase, but also in a normal *assertion*. Suppose *Q* asks *Who went to Mary's party?*, then *A* can answer by saying *At least John was one of them*. In this case we used a plural pronoun, but sometimes we can even refer back to a wh-phrase with a *singular* pronoun, as when *A* answers the above question by saying *I don't know, but Mary liked him a lot*. In the latter case *A* seems to imply that a notion of *uniqueness* is involved; he can only appropriately answer the above question in this way when he believes or presupposes that only one person went to the party. It seems that we cannot refer back to wh-phrases with explicit personal pronouns that occur in subject position, even if uniqueness is assumed. The reason is, I believe, that pronouns referring back to wh-phrases only have an *attributive* reading, and should, in our example, be read as something like *the one(s) who went to Mary's party*. At other times that we can refer back to a wh-phrase with a singular pronoun the pronoun gets a *distributive* reading. Suppose *Q* asks *Who can fully understand Albert's papers on dynamics?*, then *A's* answer *I don't know, but he will be a trained mathematician* really means that *anybody* who can fully understand Albert's papers on dynamics must be a trained mathematician.

But we don't need an explicit pronoun to 'refer' back to a wh-phrase of a question, it is well known that constituents with falling, *focal*, accent can 'refer' back to wh-phrases in questions too. Intuitively, a wh-phrase sets up a set of alternatives, while a constituent in focus picks out one of these alternatives. Whether a constituent has focal stress or not seems normally truth-conditionally irrelevant, but once we quantify in one way or another over the alternatives, things change; the sentence *John only introduced Bill to Mary* claims that for any pair in the set of alternatives, only the pair $\langle \textit{Bill}, \textit{Mary} \rangle$ is such that John introduced the first to the second. The relevant set of alternatives is restricted in a context-dependent way, and can, for instance, be determined by a foregoing question. For illustration, in the context of the question *Which*

¹if we assume that all worlds have the same domain.

gentlemen did John already introduce to which of their diner partners?, the above assertion doesn't seem false if John already introduced the gentlemen of the diner to each other, but has only started introducing the gentlemen to their lady diner partners (Von Stechow, 1990). A similar example is that in the context of the question *Who drank what kind of liqueur at the party last night?* the quantificational answer *Every guy had a litre of beer, while the girls went for the whiskey* only talks about the drinking behaviour of the guys and dolls at the party last night.

The above examples all suggest that the meaning of sentences can be in the same way dependent on questions as they can be on foregoing *quantified statements*. First, just like we can refer back to quantificational phrases with plural pronouns, as in *Most friends of Sue want to marry a Swede. They believe it will make them happy*, we saw above that we can also refer back to *wh-phrases* with plural pronouns. Second, some of the earlier discussed examples suggest that in those cases that we refer back to a *wh-phrase* by a *singular* pronoun, this singular pronoun must receive an *exhaustive* and *distributive* interpretation. The same pattern we see with singular pronouns referring back to quantificational phrases, as with Churchill's *Every soldier deserves a medal. He risked his own life for his country's sake*. Third, just like to analyse the so-called 'telescoping' cases with quantification, as Roberts' (1989) *Every player chooses a pawn. He puts it on square one*, also to analyse sequences of questions like *Who brought what cake? and How much did he eat of it?* we need to account for the *functional dependencies* between the values of the variables involved.

Most of the above discussed cases where the appropriateness or content of what is expressed by an utterance depends on an earlier question motivated in Van Rooy (1997b) the claim that the abstracts underlying the partitions induced by questions should be introduced to the discourse, and can be anaphorically taken up by elliptical answers, explicit pronouns, and anaphoric quantifiers. The reason is that these abstracts can be thought of as properties or relations, and as such determine sets and relations between sets needed to account for the above phenomena. I still believe this proposal goes a long way, but I will implement the main idea behind the specific proposal of that paper in a rather different way here. First, I won't assume anymore that we should account for the anaphoric dependence on questions by means of abstracts underlying these questions. The reason is that in some cases the anaphoric dependence cannot be accounted for by means of the abstract underlying the question by itself.² In the following example, *Does John have a sister? and is she nice?*, the interpretation of the pronoun of the second question depends on an indefinite in the first question, but because the first question is a yes/no question, the abstract associated with this question is just a 0-ary relation by means of which we cannot account for the anaphoric dependence. Instead I will use the very similar, but somewhat richer, formalism of Van den Berg (1996) to account for the anaphoric dependencies on and between questions and quantificational sentences. Although abstracts are not used to account for anaphoric dependencies, I will assume that questions still introduce abstracts, but only because in this way we can represent contexts in a relatively simple way. In Van Rooy (1997b) it was argued that questions introduce both partitions and abstracts, one for the analysis of entailment, embedded questions, and appropriateness, while the other for determining what is expressed by an answer, and for anaphoric dependencies. But once we have abstracts we don't need the partitions anymore, because we can always determine the partition induced by (a set of) abstract(s). Suppose that f is the abstract denoted by a variable introduced by a question. Then we can easily determine the partition induced by this relation on a set of worlds C : $\{\{w' \in C \mid f(w') = f(w)\} \mid w \in C\}$, or equivalently an equivalence relation: $\{\langle w, w' \rangle \in C \times C \mid f(w) = f(w')\}$.

3 Formalisation

The formal system that I will use is a variant of Van den Berg's (1996) *Dynamic Plural Predicate Logic, DPPL*. Van den Berg has shown that his formal system can account for *dependencies*

²To be sure, I could account for this dependency too, by making use of Van Rooy's (1997a) analysis of descriptive and functional pronouns. I won't, mainly because Van den Berg's analysis can be more easily extended to cover other cases of modal subordination where uniqueness need not play a role.

between sets/plurals, and I will extend his account to questions and other cases that show the phenomenon of *modal subordination*. The main differences from his system are that (i) I don't use negative transitions, but limit myself to the positive part of his proposal, (ii) I use a *two-sorted* language, where with respect to states variables range either over sets of *individuals*, or over sets of *worlds*, and (iii) I will treat (specifically used) indefinites in a somewhat different way as Van den Berg does. This two-sorted language and this different way of analysing indefinites are mainly needed to show that my analysis is systematically generalisable such that it can account for anaphoric dependencies in modal contexts, such as belief attributions.

The *syntax* for the language will not be specified explicitly, but is given implicitly in the semantic clauses. The *Models* are models for plural intensional logic, and will be triples $\langle \wp(D), \wp(W), I \rangle$, where D is the set of individuals, W the set of worlds, and I the interpretationfunction. An *assignment* for DPPL is a partial function from individual-variables to subsets of D , and world-variables to subsets of W . I use one distinguished variable, i , that represents in each state what the actual world could be. A *state* in DPPL is a set of assignments. Intuitively, each state in a context represents to the *hearer* what the *speaker* could be talking about. The ordering relation between states is defined in terms of their elements as follows: $G \leq H$ iff $\forall g \in G : \exists h \in H : g \subseteq h$. In the functional formulation of DPPL that I will use, a *context* is represented by a set of states, i.e. a set of sets of assignments. The ordering relation between states carries over to an ordering relation between contexts: $S \leq S'$ iff $\forall G \in S : \exists H \in S' : G \leq H$. Entailments between formulae are defined in terms of the update function $[\cdot]$ to be defined below, and this latter ordering relation between contexts in the usual way: $A \models B$ iff $\forall S : [A](S) \leq [A \wedge B](S)$.

Dependence

If G is a set of assignments, and x a variable, then $G(x)$, the value of x in G , denotes the set $\{g(x) : g \in G\}$. As usual, the value of variable y with respect to an assignment might be *dependent* on the value of variable x . What is special about DPPL is that this dependency carries over to states, the elements of a context. Thus, the elements of the value of variable y in state G might be dependent on the elements of the value of the variable x in G . Suppose that there are only two men in our universe, *John* and *Bill*, and that John loves *Mary*, while Bill loves *Sue* and *Dazy*. If we now would represent the sentence *Every man loves a woman* by $\forall x[Man(x) \rightarrow \exists y[Woman(y) \wedge Love(x, y)]]$, any assignment that assigns John to x , will assign Mary to y , while only some assignments that assign Bill to x will assign Sue to y , others will assign Dazy to y . Suppose we collect the three relevant assignments in G , then $G(x)$ will denote the set $\{John, Bill\}$, while $G(y)$ will denote the set $\{Mary, Sue, Dazy\}$. By looking at these sets alone the dependencies that are encoded in the states are still hid. To find those dependencies we have to look at subsets of G that assign to x a particular value. Let us say that $G|_{x=d}$ denotes the maximal substate G' of G such that $G'(x) = \{d\}$, $G' = \{g \in G \mid g(x) = d\}$. Now we can say that the value of y in our above example depends in G on the value of x , because $G|_{x=John}(y) = \{Mary\}$ differs from $G|_{x=Bill}(y) = \{Sue, Dazy\}$. Going from statics to dynamics, DPPL is defined in such a way that after the interpretation of the formula $\forall x[Man(x) \rightarrow \exists y[Woman(y) \wedge Love(x, y)]]$ with respect to our toy model, we will result in such a state G that encodes the above dependencies. Because these dependencies are encoded, a following sentence like *He_x gives her_y flowers to prove this*, or a distributively read *They_x prove this by giving them_y flowers* will be interpreted such that the sentence is true if John gives flowers to Mary, and Bill to each one of Sue and Dazy. How exactly this distributive reading of these sentences is accounted for is explained below.

Before we can state our interpretation rules, we first have to define the auxiliary notions of *subsetassignment*, *maximization* and *distribution*, introduced and more extensively discussed in Van den Berg (1996).

Subset Assignment

Subsetassignment introduces a subset y of an earlier variable x , for which A holds, and preserves all dependencies that x has:

$$(1) \quad S[y \subseteq x] := \{H | \exists G \in S \ \& \ \exists D \subseteq G(x) \ \& \\ H = \{[\bigcup_{d \in D} G|_{x=d} [y := d] \cup G|_{x \notin D}]\}$$

This introduces y on a subset D of the values that x takes. On D , y takes the same values as x , ($G|_{x=d}[y := d](y) = d$), outside D , y stays undefined ($(G|_{x \notin D}(y) = \emptyset)$, where $G|_{x \notin D} = \{g \in G | g(x) \notin D\}$). This notion can be extended in the obvious way to a definition of $S[y_1 \subseteq x_1, \dots, y_n \subseteq x_n]$, but I will leave this to the reader.

Maximization

To define the operator M_y that maximizes the output of its argument relative to the variable y , we first have to define some abbreviations. First, $G[x := D]$ will be the assignment introduction of x to $D : \{g[x := d] | g \in G \ \& \ d \in D\}$. Second, the introduction of x to a subset of the value of y : $H \leftarrow_{x \subseteq y} G$ iff $\exists D : H = G[x := D] \ \& \ H(x) \subseteq G(y)$. Third, an ordering relation between states relative to a variable $H <_x G$ iff $H(x) \subset G(x) \ \& \ H \subseteq G$. The definition has the effect that the set $H(x)$ is bigger than $G(x)$, and that the dependencies are preserved. Now we can define the maximization operator:

$$(2) \quad M_{y \subseteq x}(A)(S) = \{H \in [A](\{G\}) | \exists G \in S : \neg \exists G', H' (G' \leftarrow_{y \subseteq x} G \ \& \\ H <_x H' \ \& \ H' \in [A](\{G'\}) \}$$

This operator maximizes the output H for a given variable y , for each G in S . It is important to realise that according to this definition H is only demanded to be the largest for y , nothing is said about other variables. In Van den Berg (1996) it's assumed that all variables are introduced with their own maximization operator, but I will assume this only for quantifiers, not for *specifically used* indefinites in A represented by $\exists x A$, the indefinites that will give rise to more fine-grained partitions. To assume that most variables associated with indefinites are not introduced with their own maximization operator is essential, I believe, to account for modal subordination in case of sequences of *attitude attributions* and other cases with epistemic modalities involved.

Distribution

To account for sequences like *Every man loves a woman. He gives her flowers to prove this*, and the non-collective reading of *Every man loves a woman. They gives them flowers to prove this* we have to define a distribution operator. Every sentence is interpreted with respect to a state, which is a set of assignments. Such a state G will assign to variables y and z introduced by quantificational sentence *Every_y man loves a_z wife*, the set of all men (in x) that love a wife, and to z the set of all woman (in v) loved by a man. However, both to account for the dependencies between the values of x and the values of y , and to account for the universal force of a following sentences like *He gives her flowers to prove this*, and of the distributive reading of *Most imaginative men give them flowers to prove this*, we have to define a distribution operator. The distribution operator, $\delta_x A$, is variable dependent, and is essentially a parallel conjunction indexed by the values of x . State H is an element of $[\delta_x A](\{G\})$ iff for each d in $G(x)$ it holds that A is true in H . Things are somewhat more complicated because it might be that there are several elements of G that assign the same undefined value to x . These assignments should not be involved in the checking whether A is true or false, but should rather be ignored. The following interpretation rule results:

$$(3) \quad [\delta_x A](S) = \{H | \exists G \in S : \forall d \in G(x), \\ H|_{x=d} \in [A](\{G|_{x=d}\}) \ \& \ H|_{x=*} = G|_{x=*} \}$$

In the scope of the distribution operator δ_x a formula is not only statically interpreted for each value of x separately, but the distribution operator will also have important *dynamic* effects; it *creates dependencies* between values of variables. If the embedded sentence contains a quantifier or indefinite that introduces a discourse referent y , the value of this discourse referent might depend on the value of x . If so, a new dependence between values of x and values of y in the resulting state H is created. Actually, in the formal system all dependencies will be created this way, and therefore the distribution operator will be used in a lot of interpretation rules. Above we have

taken over Van den Berg's definition of distribution. For the analysis of questions, however, I will use a generalisation of the above definition; distribute for each state G over all non-referential variables defined in G . Fortunately, it is quite obvious how to do this operation, denoted by δA , and the exact definition can therefore be left to the readers imagination.

Translating the above mentioned sequence *Every man loves a wife. He/They gives her/them flowers to prove this* as $Every_y^x(Man(y, i), Some_z^v[Woman(z, i) \wedge Love(y, z, i)])$. $\delta_y gfp(y, z, i)$, we get the so-called 'telescoping' or 'distributive' reading, on the assumption that every man loves only one wife (due to the translation of the indefinite by a quantifier). The *collective* reading for the plural variant of the second sentence we get if we don't use the distribution operator in the second sentence, and translate it as: $gfp(y, z, i)$.

Interpretation rules

Now we can give a recursive definition of the other clauses. To account for the atomic case, we have to note that for each G in S , $G(i)$ need not be a singleton set. However, we want to interpret atomic sentences with respect to a single world. So we first look at subsets of G that assign to i the same world, and then check the atomic clause for this subset. After this checking is done, the resulting subsets will be 'glued' together again.

$$(4) \quad [P(x_1, \dots, x_n, i)](S) = \{ \{g \mid \exists w \in G(i) : g \in G|_{i=w} \ \& \ < G|_{i=w}(x_1), \dots, G|_{i=w}(x_n) \> \in I_w(P) \mid G \in S \}, \\ \text{if for each } G \text{ in } S, \text{ and for each } x_j : \text{ if } x_j \text{ is a 'referential' variable,} \\ \text{then } \forall g, g' \in G : g(x_j) = g'(x_j) \neq *, \\ \text{undefined otherwise.}$$

The definedness condition says that all 'referential' variables, variables that are introduced by 'specifically' used indefinites, as we will see soon, should for each G in S have the same interpretation w.r.t. each element of G . Although also wh-phrases will introduce variables, these will not be 'referential' variables, and therefore this constraint does not apply to pronouns referring back to wh-phrases. Still, it will be demanded that for each world, variables that represent singular pronouns should always denote a singleton set w.r.t. a state.

Once we understand the interpretation rule for the atomic clause, the rules for conjunction and negation are straightforward:

$$(5) \quad [A \wedge B](S) = [B]([A](S)) \\ (6) \quad [\neg A](S) = \{ \{g \in G \mid \neg \exists h \subseteq g : h \in \bigcup \bigcup [A]({G}) \} \mid G \in S \}$$

The interpretation rule for the 'existential' sentences is special in that it introduces partitions to the context. To which cell an element of a new state belongs depends (partly) on the value it assigns to the introduced variable. I will also assume that we reserve a special set of individual variables, the 'referential variables', that are introduced by these 'specifically' used indefinites. I will denote these variables by r 's. As a result, for every H in $[\exists r A](S)$ it holds that $H(r)$ is a singleton set:

$$(7) \quad [\exists r A](S) = \bigcup \{ [A]({G[r := \{d\}]}) \mid G \in S \ \& \ d \in D \}$$

The constraint formulated by the interpretation rule for atomic clauses relate to the variables introduced by these 'existential' sentences.

Quantification

Now we give the interpretation rule for quantification, one of the two rules that will be discussed in this section for which we defined all the machinery given above. For simplicity I will only look at distributive quantifiers. As in all cases of quantification, also for the analysis of $Q_y^x(A, B)$ two sets are compared; the maximal subset of the value of x , the context set, that satisfies A , and the maximal subset of this first set that satisfies B , where both A and B are interpreted distributively.

This is the static part of the story, and is exactly what you would expect. The dynamic part of the story is that variable y is introduced and gets as value the maximal subset of x that satisfies $A \wedge B$, and that the variables created by A and B are also introduced, but the values of those variables might depend on the value of y . The following interpretation rule takes care of all of this (where $A[y/y']$ is A with all free occurrences of y replaced by fresh y'):³

$$(8) \quad \llbracket Q_y^x(A, B) \rrbracket(S) = \{ \{ h \mid \exists K \in \llbracket M_{y' \subseteq x}(\delta_{y'}(A[y/y'])) \rrbracket(\{G\}[y' \subseteq x]) \} \& \\ \exists H \in \llbracket M_{y \subseteq y'}(\delta_y(B)) \rrbracket(\{K\}[y \subseteq y']) \& \exists w \in G(i) : \\ [Q](K|_{i=w}(y'), H|_{i=w}(y)) \& h \in H|_{i=w} \} \mid G \in S \}$$

Questions

At this point we are almost ready to state the interpretation rule for questions. To account for all the things suggested in section 2, our interpretation rule has to be such that (i) the domain over which the values of a wh-phrase vary might be context-dependent, (ii) anaphora occurring in questions should be treated distributively within a state, (iii) all questions should introduce abstracts to the discourse from which we could determine the Groenendijk & Stokhof kind of partition induced by this question, and (iv) wh-phrases, and indefinites and quantifiers occurring within questions all introduce variables to the discourse whose values are sets to which we can 'refer' back by explicit pronouns and quantificational phrases, such that the values of these sets might be dependent on each other. The following interpretation rule takes care of all of this:

$$(9) \quad \llbracket A?_q^{x_1 y_1, \dots, x_n y_n} \rrbracket(S) = \{ \{ h \mid [h^q / \hat{x}_{1 y_1}, \dots, \hat{x}_{n y_n} \mid \delta A]_{H|_{i=w}} \} \mid w \in G(i) \& \\ H \in \llbracket M_{x_1 \subseteq y_1, \dots, x_n \subseteq y_n}(\delta A) \rrbracket(\{G|_{i=w}\}[x_1 \subseteq y_1, \dots, x_n \subseteq y_n]) \& \\ h \in H \} \mid G \in S \}$$

Where the abstract $\hat{x}_{1 y_1}, \dots, \hat{x}_{n y_n} \mid A|_G$ is that function f such that $\forall w \in W: f(w) = \{ \langle d_1 \dots d_n \rangle \in G(y_1) \times \dots \times G(y_n) \mid [A](\{G^{x_1/d_1, \dots, x_n/d_n, i/w}\}) \neq \{\emptyset\} \}$ and where $G^{x/d} = \{g \mid x := \{d\} : g \in G\}$.

A multiple wh-question like *Who was teasing whom?* will be represented by something like $Tease(x_1, x_2, i)?^{x_1 y_1, x_2 y_2}$, where the values of y_1 and y_2 are the domains over which the wh-phrases vary. The question will introduce the variables x_1 and x_2 , and their values will be the maximal subsets of y_1 and y_2 of, respectively, the teasers and the teased. The question will also introduce the state-dependent abstract $\hat{x}_{1 y_1}, \hat{x}_{2 y_2} \mid \delta x_1, x_2(Tease(x_1, x_2, i))|_G$ to the discourse, if G is the relevant state. This abstract will denote a function from worlds to a relation between the teasers in y_1 , and the teased ones in y_2 .

The variables introduced by a question can now be used to determine what is expressed by an answer. The dialogue Q: *Who went to party?*, A: *John and Mary*, can now be represented by something like $WP(x, i)?_q^{x y}$. $x = (John \heartsuit Mary)$, where \heartsuit denotes some lovely kind of conjunction that unites two NP-denotations into a set. Because in state G resulting after the interpretation of the question, x denotes the maximal set of individuals in y that went to the party, the answer means that John and Mary were the only (relevant) ones who went to the party. If you don't want to build this exhaustivity into the semantics, you can also represent the answer by $(John \heartsuit Mary) \subseteq x'$, where $y \subseteq x$ is interpreted just like you would expect.

Suppose we are in a situation where we know that some of the boys were teasing the girls non-collectively, then the question *Who was teasing whom?* can be appropriately answered by a discourse like *John teased Mary. Bill teased Sue and Dazy. And Harry teased all three of them* from which we implicate that John only teased Mary, Bill only Sue and Dazy, and Harry all three of them. As observed by Büring (1996), the noun phrases *John*, *Bill* and *Harry* used in these answers typically have rising accents, while the noun phrases *Mary*, *Sue and Dazy*, and *all three of them* will get a falling accent. The falling, focal, accent indicates exhaustivity, and Büring hypothesises that rising, topical, accent indicates the existence of a further open question. This cannot be

³Note that according to this interpretation rule not only variable y is introduced whose values satisfy $A \wedge B$, but also variable y' , whose values only need to satisfy A . We seem to need this variable to account for sequences like *No professor came to the party. They were too busy with their work.*

quite right, however, because although *Harry* will get topical accent, after the third sentence the relevant question is exhaustively answered. The hypothesis of Hendriks & Dekker (1996) seems more appropriate. They propose that topical accent indicates the use of a non-monotone anaphor. But then, how should we account for the difference between the use of topical and focal accent?; normally it is assumed that also constituents with focal accent pick out particular elements of a relevant set of alternatives. But the difference can be accounted for within our framework. Very informally, constituents that are either in topical or focal stress ‘anaphorically’ refer back to variables introduced in the foregoing discourse. In both cases the interpretation of the constituent should be a (real) subset of the value of the variable the constituent ‘anaphorically’ refers back to. The difference between the use of topical and of focal accent referring back to variable x is, I would like to propose, that the interpretation of the constituent in *topical* accent should be a (real) subset of the value of x in state G for every world w in $G(i)$, while it need only be a subset of the value of x in the whole state G in case of *focal* accent. This gives rise to the prediction that the topical accent on the noun phrase *John* in the answer *John teased Mary* to the question *Who teased whom?* gives rise to the expectation that more individuals than John were teasing, while the focal accent on *Mary* in this answer only indicates that as far as we knew, John could have teased other individuals too. I believe these predictions are in accordance with our intuitions.

4 Dependencies on worlds

Even though the above analysis to account for anaphoric dependencies on and among questions is just a generalisation of Van den Berg’s analysis of dependencies on quantifiers, some readers still might find it too difficult. In this section I will argue that this complexity pays off, because the above formalism can be extended such that we can also account for other cases that show modal subordination behaviour, cases where anaphoric dependencies across *epistemic contexts* are involved. Moreover, I will argue that this way of handling these anaphoric dependencies is to be preferred to other treatments, although I will only compare my treatment with the account of modal subordination that has been proposed recently by Bart Geurts.

Just like individual variables in states refer to sets of individuals, so do word-variables refer in states to sets of worlds. Until now we haven’t made much use of world-variables. That is, we only used the distinguished variable i that refers to the set of worlds of which each element, according to each state, could be the actual world. But the reason for using a two-sorted language in the formal system was that this enables us to state in our object-language that we sometimes quantify over worlds, and moreover that these world-variables can be introduced to the discourse. The analogy with the quantificational case should be clear; just like *Every man loves a wife* introduces two variables, where the values of the variable introduced by *a wife* are functional dependent on the values of the variable introduced by *Every man*, I now propose that a belief attribution like *John believes that a man walks in the park* also introduces two variables, where the values of the variable introduced by *a man* are dependent on the values of the world-variable introduced by *believe*. This can be accounted for by the following interpretation rule (where $K_a(w)$ denotes the set of worlds compatible with what a believes in w):

$$(10) \quad \llbracket Bel_k^i(a, A, i) \rrbracket(S) = \bigcup \{ \llbracket M_{k \subseteq j}(\delta_k(A^i/k)) \rrbracket(\{G'\} \{k \subseteq j\}) \mid \exists G \in S : \\ G' = \{g \mid \exists w \in G(i) \ \& \ g \in G|_{i=w} \ \& \\ \exists H \in \llbracket \delta_k(A) \rrbracket(\{G|_{i=w}\} \{k \subseteq j\}) : K_a(w) \subseteq H(k) \} \}$$

This interpretation rule has the effect that (i) we eliminate assignments that assign to i worlds where a doesn’t believe A , and (ii), that not only ‘specifically’ used indefinites in the ‘main’ context partition the context in a more fine-grained way, but that this is also the case for indefinites used inside embedded sentences of belief attributions. You might think of the interpretation rule for belief sentences as having the effect that a formula of the form $Bel_k^i(a, \exists r A, i)$ is really interpreted as if it were of the form $\exists r \llbracket Bel_k^i(a, A, i) \rrbracket$. But due to the *distributive* interpretation of *embedded* sentences there is an important difference between the two formulae; in the latter

case the object introduced is an ordinary existing individual, while in the former case the object introduced is an *individual concept*, and thus, by our definedness condition for atomic clauses, we can (normally) not refer back to this introduced 'object' in the main context. For instance, if we translate the sequence *Mary believes that a man is walking in the park. She believes that he is whistling* by $Bel_k^i(m, \exists r[Man(r, i) \wedge WiP(r, i)], i)$. $Bel_k^i(m, Whisl(r, i), i)$, then the sentence is predicted to be true in every world in which Mary believes that a man who is walking in the park is whistling. Although each H in $\llbracket M_{k \subseteq j}(\delta_k(\exists r[MWP(r, i)]^i/k)) \rrbracket(\{G'\}[k \subseteq j])$ assigns to k the exhaustive set of worlds in $G'(j)$ where a man is walking in the park, the pronoun *he* in the second sentence need not refer to the exhaustive set of men walking in the park for each world in $H(k)$, due to the specific interpretation of indefinites. Suppose for concreteness that $G(k) = \{w, w'\}$, and that in w , d is the only man walking in the park, while in w' both d' and d'' are doing so. In that case $\llbracket \delta_k(\exists r[MWP(r, i)]^i/k) \rrbracket(\{G\})$ will have two elements, H and K , with $H|_{k=w}(r) = \{d\}$ & $H|_{k=w'}(r) = \{d'\}$, and $K|_{k=w}(r) = \{d\}$ & $K|_{k=w'}(r) = \{d''\}$. Note that because in a state H for different worlds in $H(k)$ different individuals might be involved, Mary doesn't have to believe of a specific man that he is walking in the park, and we cannot refer back in the 'main' context to this indefinite.

To some extent this looks much like what has been proposed recently by Geurts (1995, to appear). According to Geurts' analysis, modal expressions introduce discourse referents that refer to sets of world-assignment pairs in the model. Other modal expressions in the following discourse can then be interpreted with respect to the set of world-assignment pairs that satisfies the embedded sentence of the first modal statement, because these modal expressions anaphorically refer back to the discourse referents that denote these contexts. Although my account of modal subordination gives rise to very similar predictions as Geurts', there are, I believe, a few decisive advantages for my account.

First, because Geurts introduces discourse referents that refer to sets of world-assignment pairs in the model, to account for embedded modal statements he needs a whole hierarchy of assignmentfunctions in the model to avoid circularity. In the theory stated above, however, circularity is no problem, because I don't use assignmentfunctions in the model. So my analysis is, in good Amsterdam tradition, truly *non representational*, while his account is not.

Second, consider the following variant of Geach's (1967) famous example of *intentional identity* attributions with more agents involved, where the indefinite and pronoun can have no *de re* interpretation: *Hob thinks a witch walks in the park, and Nob believes that she whistles, too*, with stress on 'whistles'. Geurts (to appear) claims that he can account for such examples, but I will argue that his analysis leaves one of the main features of intentional identity attributions unexplained. Intuitively, for the above case, we have to account for two 'difficult' facts: First, that Nob believes what is attributed to Hob in the first conjunct, and that the pronoun *she* in the second conjunct can be interpreted somehow. Second, that in worlds consistent with what is believed by both Hob and Nob, the same witch is involved; that is, we don't seem to attribute to Hob and Nob two purely existential beliefs that might be completely unrelated to each other. It is easily seen that my analysis accounts for both intuitions. Geurts' analysis however, only takes care of the first intuition; the second requirement is not met. According to Geurts' analysis it is possible that there is a world w consistent with what both Hob and Nob believe, where there are two witches walking in the park, d and d' , but for Hob this world is among his belief alternatives because he believes that d is walking in the park, while for Nob this is so because his beliefs are based on a belief 'about' d' . According to my analysis such a situation is impossible, and Geurts admits that his analysis leaves this intuition unexplained. He proposes that this 'extra' should be left to pragmatics. But this is unwanted, because one of the main goals of the 'dynamic enterprise' is to formalise the proposals of traditional pragmatic theories.

Third, and perhaps empirically most important, according to my analysis the modally subordinated contexts are, in distinction with Geurts' analysis, not separated from the global context. As one consequence of this, we don't have to do anything special for the analysis of *de re* attributions, while he would. Thus, I could represent a *de re* belief attribution simply by $\exists r[Bel_k^i(a, Pr, i)]$, while Geurts must handle this case in a rather ad hoc way, because the variables introduced in the main context are, according to his analysis, not accessible to the subordinated contexts anymore.

Above we have discussed attitude attributions, but in an even simpler way we can also account for the famous sequence with which the recent discussions on modal subordination all started: *A wolf might come in. It would eat you first.* (Roberts, 1989).⁴ That is, we can analyse such sequences when we interpret epistemic possibility and necessity as follows:

$$(11) \quad \begin{aligned} [\Diamond_k^j A](S) &= \{H \mid \exists G \in S : H \in [M_{k \subseteq j}(\delta_k(A^{i/k}))](\{G\}[k \subseteq j]) \& \\ &\quad H(i) \cap H(k) \neq \emptyset\} \\ [\Box_k^j A](S) &= \{H \mid \exists G \in S : H \in [M_{k \subseteq j}(\delta_k(A^{i/k}))](\{G\}[k \subseteq j]) \& \\ &\quad (H(i) \cap H(j)) \subseteq H(k)\} \end{aligned}$$

This analysis is simpler than the analysis of belief attributions because I assume that epistemic *might* and *would* statements are *global* checkings with respect to states, and thus have the same value in each world in a state. Note that for the analysis of epistemic modality and belief attributions the set of worlds that forms the restriction might be the set of worlds consistent with what is presupposed. This will be the case when there is no salient modal subordinated context around, and technically when the variable j is i . Note that this is again something that would be difficult to handle by Geurts, because he treats modally subordinated contexts as being unrelated to the global one. Similarly, the oddity of the sequence A: *It might be that a man is walking in the park* B: *No man is walking in the park* A: *He might whistle* is immediately explained in the above formalism, while this is not so for Geurts'. The reason is, again, that in his framework main and subordinated contexts are treated as being unrelated with each other, while this is not the case for ours.⁵

5 Conclusion and Outlook

In this paper I have discussed some ways in which the meaning of utterances can depend on foregoing questions. In particular I argued that the dependence on questions shows the same kind of 'modal subordination' behaviour as utterances whose meaning is dependent on foregoing quantificational sentences. In the main part of the paper I proposed a particular formalism to account for these dependencies, and show that this formalism is general enough to account also for other constructions that show 'modal subordination' behaviour. In future work I will give a more detailed analysis and discussion of the cases of modal subordination mentioned in the fourth section of this paper, and extend it such that we can also account for modal subordination behaviour where conditionals, negations and disjunctions are involved. Once we have done this, we can account for the sometimes unusual behaviour of presuppositions triggered in these environments within the *satisfaction* account of presuppositions.

References

- Berg, M. van den, (1996), *The Internal Structure of Discourse*, Ph.D. thesis, University of Amsterdam.
- Büring, D. (1996), *The 59th Street Bridge Accent*, Ph.D. thesis, University of Tübingen.
- Frank, A. (1997), *Context Dependence in Modal Constructions*, Ph.D. thesis, University of Stuttgart.
- Geach, P. (1967), "Intentional identity", *Journal of Philosophy*, 64, pp. 627-632.
- Groenendijk, J. and M. Stokhof, (1984), *Studies on the Semantics of Questions and the Pragmatics of Answers*, Jurriaans BV, Amsterdam.

⁴It should be noted that also Van den Berg (1996) briefly suggests that his formalism can be used to account for these kinds of sequences. What he suggests, however, is rather different from my concrete proposal.

⁵Also in Annette Frank's (1997) recent analysis of modal subordination main and subordinated contexts are related to each other, and as a result she can also account for the latter kind of discourse.

- Groenendijk J. and M. Stokhof, (1997), "Questions", In: Benthem J. van and A. ter Meulen (eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam, pp. 1055-1124.
- Geurts, B. (1995), *Presupposing*, Ph.D. thesis, University of Stuttgart.
- Geurts, B. (to appear), "Presuppositions and anaphors in attitude contexts", accepted for *Linguistics and Philosophy*.
- Hendriks, H. and P. Dekker, (1996), "Links without location", In: P. Dekker and M. Stokhof (eds.), *The proceedings of the 10th Amsterdam Colloquium*.
- Roberts, C. (1989), "Modal subordination and pronominal anaphora in discourse", *Linguistics and Philosophy*, 12, pp. 683-721.
- Rooy, R. van, (1997a), *Attitudes and Changing Contexts*, Ph.D. thesis, University of Stuttgart.
- Rooy, R. van, (1997b), "The context-dependence of questions and answers", In: P. Dekker, M. Stokhof, Y. Venema, *The Proceedings of the 11th Amsterdam Colloquium*, Amsterdam.
- Stechow, A. von, (1990), "Focusing and background operators", In: *Discourse Particles, Pragmatics and Beyond*, John Benjamins, Amsterdam.

A Discourse Theory of Manner and Factive Adverbial Modification*

Dr. Adam Zachary Wyner
Department of English,
Bar Ilan University,
Ramat Gan, 52900, Israel
mswyner@pluto.mscc.huji.ac.il

Abstract

In the paper, similarities are demonstrated between sentences with manner and factive adverbs and discourses with correlated adjectives which predicate of a pronoun which refers to an antecedent event or fact. It is argued that the intrasentential adverbial cases should be analyzed in terms of the intersentential adjectival cases. An account, the Discourse Theory of adverbial modification, is provided for the intersentential cases using a version of DRT (Asher, 1993), then extended to the intrasentential cases. It is argued that this theory is superior to the current and widespread "Association Theory" of adverbial modification, in which particular sorts of adverbs are adjoined to particular adjunction sites. Furthermore, it is argued that adverbs are not be combinators with respect to events or facts.

Introduction

In this paper, I show that intrasentential adverbs of manner and "fact" behave, in several respects, like sentences which adjectives which predicate of a pronoun which refers to an antecedent event or fact. I argue that the intrasentential cases should be analyzed in terms of the intersentential cases. I provide an account in DRT of the intersentential cases, then extend this to the adverbial cases. The organization of the paper is as follows. In the first section, I discuss some of basic properties of manner and factive adverbs, as well as some of the current assumptions about the syntax and semantics of adverbial modification. I point out some of the problems of these assumptions. In the second section, I compare the intrasentential adverbial cases to intersentential adjectival cases and suggest analyzing the adverbs in terms of discourse. In the third section, I provide an explicit account of both the intra and intersentential cases in terms of Asher's (1993) version of DRT.

1 Manner Adverbs, Factive Adverbs, and the Association Theory

The analysis of adverbs involves a great range of phenomena and problems; in this paper, I restrict our attention. As is well known, certain adverbs such as "rudely" are ambiguous between a manner and what will be called here a "factive" interpretation (see Parsons 1990, Moore 1993, and Wyner 1997a) on the properties of factive adverbs). The manner interpretation of (1a) is paraphrased as in (2a), and the factive interpretation is paraphrased as in (2b). Other adverbs which have the "factive" interpretation are "stupidly", "wisely", and others; other adverbs such as "passionately" are unambiguously adverbs of manner.

*I would like to thank Jonathan Ginzburg for helpful discussion of some of the topics of this paper. Of course, I bear responsibility for the remaining errors and unclarity.

- (1) a. Bill left rudely.
 b. Rudely, Bill left.
- (2) a. The manner in which Bill left was rude.
 b. That Bill left was rude.

In current analyses (for example, Jackendoff (1972), McConnell-Ginet (1982), Travis (1988), Parsons (1990), and Bowers (1993) among many others), the difference between these interpretations depends on distinct hierarchical attachment sites for the adverbs. For example, we might say that the adverb "rudely" in its manner interpretation is a VP-adjunct while in its factive interpretation is an IP-adjunct, schematized as in (3).

- (3) a. [IP Bill [VP₁ [VP₂ left] rudely]]
 b. [IP Rudely [IP Bill [VP₁ left]]]

Let us refer to such a theory as the "Association Theory" (see Wyner (1994) for a fuller discussion of some of these issues). In the Association Theory, distinct syntactic and/or semantic properties are associated with distinct positions in the sentence; that is, with respect to adverbial semantics, the sentence is syntactically and semantically heterogenous, for we find different properties at the VP and sentence-levels. The distinct adverbs are, in each case, sensitive to the distinctions of these positions. In a "classical" Montogovian semantics for adverbs (Thomason and Stalnaker 1973), the distinct sorts of adverbs might apply to distinct semantic types, manner adverbs as functions from VP types to VP types, and factive adverbs from sentence types to sentence types. In a Neo-Davidsonian Event Theoretic approach, one might claim that only at the VP level is an "event" argument available for manner adverbial modification, while at the IP level, we have some proposition-like entity suitable for the application of factive adverbs. Yet another approach suggests that manner adverbs modify the main predicate, while factive adverbs modify an abstract higher predicate such as "ACT" (see McConnell-Ginet (1982) and subsequent similar proposals).

A wide variety of semantic and syntactic evidence is provided in the literature to support the Association Theory. Let us consider manner adverbs first. An early and long lasting assumption in formal grammatical thinking has been that since manner adverbs appear to modify the verb or the event of the sentence, they should appear in close proximity to that which they modify (Keyser 1968); thus, manner adverbs modify the verb, so should appear in close proximity to the verb. More specific tests are whether the adverb appears under the scope of an operator which is on or within the VP or whether the adverb appears under the scope of another adverb which is said to be a sentence operator. For instance, consider negation (4a), a universal quantifier in object position (4b), and the relative interpretation of the adverbs in (4c).

- (4) a. Bill didn't kiss Jill passionately.
 b. Bill kissed every girl passionately.
 c. Stupidly, Bill kissed Jill passionately.

In each case, we find interpretive restrictions. In (4a), negation has scope over the manner adverb, and not vice versa; that is, we can negate the manner, but there is no interpretation in which passionateness is attributed to Bill's not kissing Jill. In (4b), the universal quantifier in object position has scope over the manner adverb; each kissing was passionate, not that the "collection" of kissings was passionate. Finally, in (4c), the factive adverb has scope over the manner adverb; that is, what was stupid was that Bill kissed Jill in a manner which was passionate, and not vice versa; nor is there an interpretation in which the adverbs are "autonomous" – where the stupidity and passionateness are attributed to the kissing irrespective of one another. We still need to determine, for this case, what the "scope" of the factive adverb is, but assume for the moment that it is the sentence. By similar considerations, we might conclude that an adverbial modifier such as "with wet lips", among others, is also a VP modifier. Furthermore, we can "scramble" the adverbs in (5); the sentences are synonymous.

- (5) a. Bill kissed Jill passionately with wet lips.
- b. Bill kissed Jill with wet lips passionately.

Contrast these observations with factive adverbs.

- (6) a. Stupidly, Bill didn't kiss Jill.
- b. Stupidly, every boy kissed Jill.
- c. Stupidly, Bill kissed Jill passionately.

In (6a), the factive adverb appears outside the scope of negation; that is, what was stupid was that Bill didn't kiss Jill, rather than it is false that it is stupid that Bill kissed Jill. In (6b), the factive adverb can appear outside the scope of the universal quantifier in subject position; what was stupid was that the group of boys kissed Jill, though it may not have been stupid with respect to some particular individuals. In (6c), as noted above, the factive adverb applies to expressions which contain manner adverbs, and not vice versa. Such evidence has been taken to indicate that, semantically, the factive adverb is not in the VP, but may have scope over the entire sentence.

Among the syntactic evidence, we may consider the syntactic positions in which the adverbs are said to appear and VP-ellipsis (these matters are discussed more extensively in Wyner (1997b)). For example, if it is the case that manner adverbs associate with the VP, then they should not appear in a position where they must adjoin to the IP; where factive adverbs associate with the IP, then they should not appear in a position where they must appear adjoined to VP. For example, (7a) with a manner adverb and (7b) with a factive adverb are often claimed to be "ungrammatical". Furthermore, if a manner adverb is in a VP while a factive adverb is not, then the manner adverb may be "copied" into an ellided VP, while a factive should not. In (8), we see that indeed, this is the case; that is, it generally seems to be the case that we understand the ellided VP to mean that Will's kissing was passionate, but not that Will's kissing was stupid.

- (7) a. Bill passionately may have kissed Jill.
- b. Bill may have kissed, stupidly, Jill.
- (8) a. Bill kissed Jill passionately, and Will did too.
- b. Stupidly, Bill kissed Jill, and Will did too.

However, the syntactic evidence is not unequivocal. First, it is not clear that the examples in (7) really are "ungrammatical", particularly in comparison with "benchmark" cases of ungrammaticality. For example, manner adverbs can appear sentence initially as in (9a) while factive adverbs can appear within the VP between a NPI licenser object and NPI licensee indirect object as in (9b), interpreted as in (9c).

- (9) a. Passionately, Bill kissed Jill.
- b. Bill gave no gifts, stupidly, to any grandchild.
- c. It was stupid of Bill to give no gifts to any grandchild.

A great deal seems to depend on intonation, and it is not yet clear that intonation determines or indicates syntactic adjunction site (see Steedman 1997). There may be other factors contributing to the surface distribution of adverbs which determine that a manner adverb is best at the end of a sentence, whereas the factive adverb is best at the beginning. In addition, VP-ellipsis data suggest that the manner adverb in (9a) is "abstractly" within the VP, while, as McCawley (1982) and Wyner (1997b) show, in (9b), the factive adverb is "abstractly" not in the VP. For example, (10a) can be interpreted as (10b), where each kiss was passionate; in contrast, (11a) can be interpreted as in (11b), where the "joint" giving was stupid, but it can't be interpreted as in (11c), where each giving was stupid.

- (10) a. Passionately, Bill kissed Jill, and Will did too.
- b. Bill kissed Jill passionately, and Will kissed Jill passionately.

- (11) a. Bill gave cheap gifts, stupidly, to Jill, and Will did too.
 b. That both Bill and Will gave cheap gifts to Jill was stupid, though what Bill did or what Will did separately in and of itself needn't have been stupid.
 c. That Bill gave cheap gifts to Jill was stupid, and that Will gave cheap gifts to Jill was stupid.

Furthermore, whether VP-ellipsis relies on constituent structure depends on the analysis of VP-ellipsis (see Dalrymple, Shieber, and Pereira (1991), Lappin (1996) and Wyner 1997b).

The data in (9)-(11), the open questions, and the alternative analyses raise problems for the Association Theory, but they are not knockdown arguments. For example, one could amend the Association Theory with some sort of adverbial movement. But clearly, this is an ad hoc strategy to save the analysis; it has not been clearly proposed or defended in the literature, nor is it obviously consistent with current syntactic frameworks which employ movement such as Minimalism (Chomsky 1995). Still, this is a relatively weak criticism of the Association Theory.

2 Discourse Comparisons

However, discourse cases are a critical problem for the Association Theory. First, we consider some of the basic cases of pronominal reference and extensions to anaphora to abstract objects. Then we consider how these observations extend to adjectival predication and parallels to adverbial modification. Discourse Representation Theory (DRT, see Kamp and Reyle 1993, Asher 1993) is designed to account for cases of intersentential anaphora as in (12).

- (12) a. A man kissed Jill. He was tall. And he was fat.
 b. Bill kissed Jill. She was thin. And she was short.
 c. Jill didn't kiss a man_i. *He_i was tall.
 d. Jill kissed every man_i. *He_i was tall.

We know that proper names and indefinites can serve as discourse antecedents of pronouns in (12a) and (12b); the pronouns depend on some antecedent discourse referent. Furthermore, as (12a) and (12b) show, we may add intersective adjectives such as "tall/short" and "fat/thin" in either order; they simply predicate additional properties of the antecedent. These cases also show that whether the antecedent is in a subject position or an object position is irrelevant; in general, we relate the pronouns and antecedent in these cases without resorting to a syntactic relationship between them, for the antecedent is made available semantically. Obviously, the structural relationship between the clauses is irrelevant as well; these are distinct sentences after all. (12c) and (12d) show well-known restrictions on anaphora under negation and universal quantification.

Peterson (1997) and Asher (1993) show that we can have pronominal anaphora to a variety of abstract objects (propositions, facts, and events), and we can type verbs by the sort of abstract object they predicate of. The expression in (13) can serve as the antecedent of the pronouns in (14).

- (13) Bill kissed Jill.
 (14) a. It indicated that Bill liked Jill. *Fact Anaphora*
 b. Will believed it. *Proposition Anaphora*
 c. It occurred at 10pm. *Event Anaphora*

How (13) is the antecedent of these diverse sorts of predicates is not entirely clear: it may be that (13) simultaneously provides a variety of sorts of abstract objects; alternatively, the different sorts of objects could be created "on the fly" by the predicate itself, perhaps assuming some initial, default sort. We shall consider this again later.

The adjectival predicates "is passionate" and "is stupid" appear to behave in the same way. In (15b) and (15c), "it" refers to Bill's kissing Jill. Under the relevant interpretations, "is passionate"

predicates of the event of Bill's kissing Jill and "is stupid" predicates of the fact of Bill's kissing Jill.

- (15) a. Bill kissed Jill.
- b. It was stupid.
- c. It was passionate.

(15a) and (15b) taken together seem to be synonymous with (16a), the adverbial form with the factive interpretation. (15a) and (15c) taken together seem, for our purposes, synonymous with (16b).

- (16) a. Stupidly, Bill kissed Jill.
- b. Bill kissed Jill passionately.

As an anonymous reviewer pointed out, there is a difference between the adverbial case and the adjectival case, which is interesting in and of itself, but not relevant to our discussion; in particular, (15a) and (15c) together seem to attribute passionateness equally to both participants, while (16b) seems to attribute passionateness just to Bill. This observation points to issues concerning the relationship between adverbs and arguments (on this, see Wyner 1998a) which need not concern us here. This aside, the adjectival and adverbial cases contribute the same information.

We might imagine two analyses for the intersentential cases which preserve the view that one sort of adjective applies to the VP while another applies to some syntactic or semantic expression "larger" than the VP. In one approach, the pronoun might refer to a VP meaning, say one which contains the subject (the so-called VP-Internal Subject Hypothesis in Pollock (1989) and others). Another slightly different approach might be that there is pronominal reference to "the kissing", some abstract entity. However, as with the VP approach, the kissing must have the properties, spatio-temporal location, and participants specified in the antecedent; that is, it must be the kissing specified by the information in the whole antecedent sentence. And, if there is negation or quantification, these too, as we see below, have an effect. In any case, these issues indicate that it is difficult to distinguish a VP element from a sentence element in such a way to distinguish between the cases of anaphora. We will, then, have to consider what entity is the antecedent in each case, and where that entity is defined.

One other comment is in order before considering addition data. The intrasentential cases might be more felicitous with expressions such as "also", "yet", "and", "but", and so on. "That" as a pronoun or "one of those" might work better in some cases than "it". I have not used "did it" or "was done". As these refinements are not particularly relevant to the core issues, we consider the simplest cases.

The observations above, that the adjectival and adverbial cases provide the same information, are not, in and of themselves, particularly novel; the intersentential cases might be viewed as paraphrases of the adverbial cases. However, what is novel is how we use this observation. Obviously, as with the cases in (12)-(14), the adjectives contribute information, but in such a way as has nothing to do with attachment to or within a previous sentence. The core question of this paper is, if attachment is irrelevant in the adjectival cases, why should it be relevant in the adverbial cases? My claim is just that attachment in the adverbial cases is just as irrelevant in the adjectival cases, moreover, that I can provide an explicit theory in which the adverbial and adjectival cases are more closely related. Turning the issue the other way, if one were to argue that the analyses for the adjectives and adverbs differ, one would have to say why. And it is not clear how this could be explained. The claim that the adverbial cases should be assimilated to the adjectival cases is directly contrary to the basic analysis of the Association Theory, where the adjunction site of the adverb is used to explain the interpretation of the adverb.

Let us build up the observations systematically. It is quite clear that whether we add information by adjectives or by adverbs, no matter what the order, we get the same result; all the cases in (17) are synonymous. Adverbs can be contained within expressions which serve as pronominal antecedents; the discourses can be extended, gradually adding more and more information.

- (17) a. Bill kissed Jill with wet lips. It was passionate.
- b. Bill kissed Jill passionately. It was with wet lips.
- c. Bill kissed Jill. It was passionate. It was with wet lips.
- d. Bill kissed Jill. It was with wet lips. It was passionate.
- e. Bill kissed Jill passionately with wet lips.
- f. Bill kissed Jill with wet lips passionately.

The examples in (17c) and (17d) are a critical problem for the Association Theory, for while it might be possible to build a theory in which the pronoun in (17a) and (17b) refer to a VP-like entity which also contains the subject, this possibility is unavailable for (17c) and (17d). The problem is that the second occurrence of the pronoun in these examples can refer to the information comprised of the first two sentences; obviously, these don't comprise a VP, yet the adjective can still predicate of them; for example, "Bill kissed Jill" and "It was passionate" can be, together, the antecedent for the pronoun "It was with wet lips", allowing us to conclude that Bill's kissing Jill was both passionate and with wet lips.

We want next to consider more complex discourse cases in order to see whether the adverbial orderings we saw above hold in discourse as well. First, we should observe that it is not entirely clear whether expressions with factive adverbs or adjectives can serve as pronominal antecedents. For example, while the pronoun with the factive predicate "indicates" in (18a) might be interpreted along the lines of (18b), it is not certain that we have a "fact" antecedent; (18a) and (18b) seem better expressed as (18c), where the judgement is the subject. Furthermore, propositional anaphora (18d) does not clearly have as antecedent "Stupidly, Bill kissed Jill"; certainly event anaphora in (18e) does not have "Stupidly, Bill kissed Jill" as antecedent. We could, in contrast, have a manner adverb in the sentence which, as a whole, serves as antecedent for either fact, proposition, or event anaphora.

- (18) a. Stupidly, Bill kissed Jill. It indicates that we no longer tolerate affection in public.
- b. That Bill kissed Jill was stupid indicates that we no longer tolerate affection in public.
- c. The judgement that Bill kissed Jill was stupid indicates that we no longer tolerate affection in public.
- d. Will believed it.
- e. It occurred at 10pm.

We see that, in general, factive adverbs and adjectives can't appear in expressions which are antecedents of discourse anaphora. While it is not clear why this might be the case, it does not effect the basic point I wish to make in this paper. Speculatively, the restriction seems to be some like the following: if factive adverbs coerce the expression they operate on into a compatible type, a fact, and if the other predicates, "indicate", "believe", and "occur" similarly coerce their antecedents, we might account for the observations in (18) if we assume that coercion is not iterative, one doesn't make a fact expression into a fact expression. But, more consideration must be given to these matters. However, the observation that factive adverbs do not seem to appear well in the antecedents to pronouns will serve us later.

At this point, we see that adverbial and adjectival cases have, largely, the same information. In the following, the adjectival cases pattern like the adverbial cases in (4) and (8) with respect to the relative scope of adverbs, negation, and universal quantification. Since we find similar patterns, it would make sense to account for them uniformly. Moreover, given that the Association Theory says nothing about the intersentential adjectival cases, we should consider whether the independently required account for them can be used for the intrasentential adverbial cases.

As we saw earlier, the interpretive order of the adverbials in (19a) follows the paraphrase in (19b), not (19c); that is, we may apply a factive adverb to an expression containing a manner adverb, but not vice versa. To this, we can add the observation that pronominal anaphora in (19d) and (19e) follow just the same pattern; while the pronoun in (19d) can have as antecedent "Bill's kissing Jill passionately", the pronoun in (19e) can not have as antecedent "that Bill kissed

Jill was stupid" as antecedent. To the extent that the discourse in (19e) is well-formed, it must be the manner of the kissing itself which was passionate; although, there is something odd about such continuations, for it is not clear what is stupid, the kissing itself or the passionate kissing. A discourse such as (19e) is difficult because it is unclear what is being predicated of and what its properties are. To sum up, there is a meaning which is unavailable in the sentence (19a), the paraphrase (19c), and in the discourse (19e).

- (19) a. Stupidly, Bill kissed Jill passionately.
 b. That Bill kissed Jill in a manner which was passionate was stupid.
 c. *The manner in which Bill's kissing Jill was stupid was passionate.
 d. Bill kissed Jill passionately. It was stupid.
 e. Stupidly, Bill kissed Jill. *It was passionate.

(19e) is particularly a problem for the Association Theory; yet the VP antecedent for manner predicates might be a saving strategy; we eliminate this below.

The facts of (19c) and (19e) are not simply a matter of intrasentential or VP predication, for we get the same patterns in extended discourse.

- (20) a. Bill kissed Jill. It was passionate. It was stupid.
 b. Bill kissed Jill. It was stupid. It was passionate.

(20a) and (20b), to the extent that it is well-formed, have just the same interpretation as the good cases in (19a), (19b), and (19d). The other interpretation, as we've seen before, is simply unavailable no matter what the order of the predicates. (20b) is a fundamental problem for the Association Theory; as we've seen earlier, we can use pronominal anaphors in such cases; and the issue of attachment is irrelevant here. Clearly, we can not explain the restriction on interpretation which we find here in structural terms; rather, it appears that we must refer to some semantic restriction. And, if we need such a semantic restriction on the relative order of application in the intersentential cases, then for simplicity, we can use it for the intrasentential cases as well.

Let us consider quantifier and negation cases, which are said to be evidence for the Association Theory. In (21a), the intended antecedent for the pronoun is "Jill's kissing every boy", which fails; in (21b), the intended antecedent is "Jill's not kissing Bill", which again fails. In contrast, the cases in (22) work just fine: in (22a), what was stupid was the group of girls kissing Bill; in (22b), Jill's not kissing Bill.

- (21) a. Jill kissed every boy. *It was passionate.
 b. Jill didn't kiss Bill. *It was passionate.
 (22) a. Every girl kissed Bill. It was stupid.
 b. Jill didn't kiss Bill. It was stupid.

As before, the issue of attachment is irrelevant here. We must account for why the cases in (21) are ill-formed, but those in (22) are fine.

Consider the cases in (21) in a bit more detail, comparing them to similar adverbial cases. A theory which says that the antecedent of the pronoun in (21) is a VP or the verb meaning would fail to account for the data, for the VP or verb meaning would seem to be available as an antecedent to the very same extent as they are in non-negated, non-quantified sentences; we should, in principle, have interpretations much like the adverbial forms in (4), where the adverb appears "under" the scope of the quantifier and negation. Interestingly, the cases in (21) are parallel to the NP anaphora cases, which suggests a similar account. It would seem then, following this parallel, that at the level at which discourse anaphora is resolved, there are no appropriate discourse referents available as antecedents for the pronouns in (21). Clearly, the universal quantifier and negation determine, in these cases, what discourse referents are available at this level; that is, first we apply the universal quantifier and negation, which then determines the domain of discourse referents. If adverbs, like the intersentential cases, depend on discourse referents, and if the domain of those

referents is determined, in part, by application of the universal quantifier and negation, then we can get the effect of the relative scope of adverbs with respect to quantifiers and negation without having to stipulate it in terms of attachment site. In particular, the adverb will apply to any position where the appropriate discourse antecedent is available. As the intrasentential cases in (21) show, the universal quantifier and negation make the relevant discourse referent unavailable; because we are dealing with discourse, the anaphora is determined only after the application of the quantifier and negation. This implies in the adverbial case that a manner adverb can't apply after negation or have scope over the universal quantifier, for the discourse referent is unavailable. But this suggests something rather novel; within the VP, before the application of the quantifier or negation, an appropriate discourse referent is available for the adverb, and after the application it is not. It just so happens that the quantifier and negation apply over VPs as well as over sentences, giving the impression that manner adverbs can't apply outside of the VP; in these cases, there is no antecedent for the adverb, just as in the discourse cases in (21). It should, then, be no surprise that just in those cases where a manner adverb can't appear, that is, where it applies after negation or the universal quantifier, are just those cases in which the discourses of the form in (21) are ill-formed.

We have yet to explain a variety of other problems: why manner adverbs can apply outside of the VP; why factives don't block manner adverbs entirely; why the discourse referent is, after negation and quantification, unavailable; and why manner adverbs, unlike manner adjectives, can't refer to some event outside the sentence in which they appear. But we will return to these issues in the next section.

In contrast to manner predicates, universal quantifiers and negation do not prevent anaphoric reference in (22), with the factive predicate. Again, attachment is irrelevant. Clearly, the antecedent expressions for both manner and factive predicates are the same, yet the adjectives, like the adverbs, behave distinctly. This suggests strongly that the sort of antecedent available to the factive adjective is distinct from the manner adjective – one an event and the other a fact. Then, it would appear that sentences with universal quantification and with negation can be facts, but not events, as claimed in the literature. The factive adjective predicates of this. The question is, at what level is the fact determined, and what is the element which determines the fact.

3 A Discourse Analysis

I will provide an account of the observations above in a DRT framework. The novelty of the account is the uniform, compositional treatment of intrasentential adverbs as intersentential adjectives related anaphorically to events or facts. The analysis achieves several goals. First, it undermines the Association Theory; the key to adverbs is not the arbitrary, fixed adjunction site of the adverb, but what sorts of semantic entities are available to the adverb and at what point in the sentence. Restrictions on the distribution and interpretation of adverbs arise out of the interaction of independent elements. Second, the analysis provides a novel view of manner adverbs, in contrast to the current "standard" Neo-Davidsonian view in which manner adverbs are eventive combinators. Third, the analysis provides part of an account of factive adverbs. And finally, despite the distinctions between factive and manner adverbs, they are treated very much on a par; thus, we provide an underlying uniformity to the theory of adverbial modification.

I make use of Asher's (1993) version of DRT, including subDRSs, predicative DRSs, accessibility, a "characteristic function", abstract entities such as facts, and anaphora to these abstract entities. However, I differ from Asher, particularly on the treatment of adverbs. Asher (1993: 85-86) essentially adopts the standard Neo-Davidsonian analysis. Moreover, Asher (1993: 88) introduces the event discourse referent at a particular point in the syntax-semantics. I will argue that, given the observations above, this analysis of adverbs is not satisfactory.

I review DRT as needed, then apply it to the observations above. A Discourse Representation Structure K is an ordered pair of sets, $\langle UK, ConK \rangle$, where UK is a set of discourse referents of K and $ConK$ is a set of conditions of K , some of which are predications on the individuals in UK , others of which are complex. A sentence can be represented as a single DRS and serve as

the context against which we interpret subsequent sentences. Given two DRSs in a discourse, we “merge” by set-theoretic union, the set of discourse referents of one DRS $K1$ with that of the other $K2$ as well as the conditions of one with the conditions of the other. This is DRS-update, which I will write as $UPDATE(K1, K2)$. Asher introduces lambda-abstraction over conditions in DRSs in order to give us compositional rules; these yield predicative DRSs. In order not to be concerned with unnecessary technical details, we assume lambda-conversion is such that when a predicative DRS $K1$ applies to another predicative DRS $K2$, both DRSs being of the relevant types, the referent sets and the condition sets of each DRS merge, respectively, by set-theoretic union (see Asher, 1993: 100).

Asher incorporates a Neo-Davidsonian analysis into DRT, allowing events as discourse referents, abstraction over them in the basic predicate, and adverbs as combinators over events. As in Neo-Davidsonian Event theory, a predicate like “kiss” is a three place relation between the participants and the event argument. However, unlike contemporary applications of event theory (Parsons, 1990; Landman, 1997), we will not make use of thematic roles. Nor is tense or aspect relevant here. Asher (1993: 86) claims that the *Infl* node introduces an event discourse referent into the universe of discourse. Furthermore, proper names are used as predicates of discourse variables for simplicity.

A sentence with a pronoun is represented as a DRS with an incomplete meaning in that the value of the pronoun must be determined with respect to some other discourse referent; an incomplete condition is given as $z = ?$, where z is a discourse referent in the DRS of the sentence in which the pronoun appears, and $?$ is eventually replaced by the value of the other discourse referent. We must determine what values $?$ can be and what conditions might hold in identifying this value. An example appears in (23). We wish to resolve the discourse in (23a); the relevant parts are in (23b)-(23f).

- (23) a. Bill kissed Jill_i. She_i cried.
 b. Bill kissed Jill.
 c. $\langle \{x, y, e1\}, \{Bill(x), Jill(y), kiss(x, y, e1)\} \rangle$
 d. She cried.
 e. $\langle \{z, e2\}, \{cry(z, e2), z = ?\} \rangle$
 f. $\langle \{x, y, z, e1, e2\}, \{Bill(x), Jill(y), kiss(x, y, e1), cry(z, e2), z = ?\} \rangle$

Given the conditions for determining the value of $?$, discussed below, the result appears as in (24a), where “Jill” is the one who cries. Note, however, that $?$ could also be resolved as z ; that is, the pronoun could be used deictically, not anaphorically, and the resulting DRS is (24b).

- (24) a. $\langle \{x, y, z, e1, e2\}, \{Bill(x), Jill(y), kiss(x, y, e1), cry(z, e2), z = y\} \rangle$
 b. $\langle \{x, y, z, e1, e2\}, \{Bill(x), Jill(y), kiss(x, y, e1), cry(z, e2), z = z\} \rangle$

How do we determine the value of $?$ This depends on subDRSs, subordination, complex DRSs, and accessibility, which are sketched here. DRSs can be complex, in particular, a DRS $K1$ can contain a DRS $K2$ as a condition; we say that $K2$ is a subDRS and that $K2$ is subordinate to $K1$. Furthermore, we have complex conditions, ones made up of subDRSs and DRS logical operators or relations. As indicated below, a quantifier relation or negation introduce complex conditions; DRSs which are under the scope of a universal or negation are subDRSs of a DRS. Some expressions may introduce a discourse referent into the universe of discourse, others may not. To account for anaphora in discourse, DRT uses the notion of accessibility, which is defined with respect to subordinate DRSs, of one discourse referent to another as in Asher (1993: 76-77). Basically, a discourse referent y is accessible to $?$ only where the universe of discourse containing y is in the immediate set of discourse referents (i.e. if $z = ?$ is a condition of $ConK$, then y has to be in UK), or of a superordinate DRS referent set, or if $z = ?$ is part of a complex conditional expression, then y is in the referent set of the DRS immediately to the left. Pronouns, represented by $?$, can only be resolved by replacing $?$ with an accessible discourse referent. We have examples below.

Consider a universal quantifier. Asher (1993: 89-91) claims that in (25), the event argument has wide scope over the universal quantifier.

(25) Jill kissed every boy.

In such an interpretation, there is one event, the “sum of kissing of every boy”. The DRS appears as in (26). The universal quantifier itself is expressed as a conditional relation between subDRSs, from individual boys to events in which Jill kisses him.

(26) $\langle \{x, e\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle \rightarrow \langle \{\}, \{kiss(x, y, e)\} \rangle \rangle$

Notice that there is an event discourse referent, and since it is in the “main” DRS, not a subDRS, it is accessible to anaphora in a discourse continuation. In contrast, the discourse referents *y* for individual boys are in a subDRS. Discourse anaphora fails in (27), for the pronoun does not have an accessible discourse referent.

(27) Jill kissed every boy_{*i*}. *He_{*i*} cried.

There are three problems with this analysis specifically with respect to the event argument. First, as Asher (1993: 92) notes, the universal quantifier has a unique representation among the determiners, which otherwise are expressed as relations between the nominal restriction and the remainder of the sentence; the reason for this distinction is unexplained; optimally, quantificational expressions will have uniform forms. Second, as Landman (1997) points out, the interpretation of (25) with a “wide” scope event argument is wrong, for there must have been individual hitting events. Landman claims that the universal quantifier must have wide scope over the event argument. Third, Asher’s analysis makes the wrong predication about event anaphora in (28) in comparison to the adverbial case in (29). We saw that event anaphora is blocked in the intrasentential case with a universal quantifier, just the opposite of what Asher would allow.

(28) Jill kissed every boy. *It was passionate.

(29) Jill kissed every boy passionately.

(28) suggests that the event discourse referent, like the individual referents, is contained in the subDRS introduced by the universal quantifier; but in the adverbial case, somehow “discourse reference” is available. I will return to the different between inter and intra sentential modification later. In the meantime, I would suggest that the universal quantifier is just like the other quantificational expressions which appear in (Asher, 1993: 92). While there may be problems with this, something along these lines seems called for. I will indicate the quantificational relation by the expression enclosed between / ... / ; the relation holds between two DRSs. The other important element is that quantificational relations subordinate the event argument. So, we represent (25) as in (30).

(30) $\langle \{x\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle$

Such a representation is consistent with the interpretation as well as the restriction on anaphora. Neither individual events of Jill hitting a boy nor individual boys are available for discourse anaphora.

Next, let us consider the representation for VP negation as in (31).

(31) Jill didn’t kiss Bill.

There are a variety of issues to consider here, but for my purposes I will keep things as simple as possible. Asher (1993: 81) claims that while negation can have, with appropriate abstraction, scope over V, VP, I, or S, the preferred reading, Asher (1993: 217) suggests, is one in which negation appears over the event argument. There are a host of issues I leave for future research and which are not relevant to the discussion at hand (see Asher, (1993: 214-221), (Kamp and Reyle, (1993: 102), and note that I do not discuss focus here). The important point for our purposes is that the event argument appears in a subDRS. Therefore, I assume the expression of (31) as in (32).

(32) $\{\{\}, \{-\{\}, \{\{\{x, y, e1\}, \{Jill(x), Bill(y), kiss(x, y, e1)\}\}\}\}\}$

Note that, as with the universal quantifier, negation introduces a subDRS which contains all the discourse referents to Jill, Bill, and the event; therefore, we don't have discourse anaphora to these referents, which is consistent with the facts in (33a) and (33b). Yet, we still have to account for the contrast in the adverbial case, (33c).

- (33) a. Jill didn't kiss Bill_i. *He_i cried.
 b. Jill didn't kiss Bill. *It was passionate.
 c. Jill didn't kiss Bill passionately.

Let us now consider the analysis of adverbs. In Asher, as is standard in the Neo-Davidsonian framework, a manner adverb is a VP function from relations between individuals and events to relations between individuals and events. Consider (34). For simplicity, let us assume a VP meaning with the object NP already in place. We assume that (35) and (36) combine to produce the VP in (37), the DRS for "kissed Jill passionately". We introduce, at *Infl*, the discourse referent event *e1*, and then the subject; the result is (38). Notice that only once the event discourse referent is introduced it is available for anaphora.

(34) Bill kissed Jill passionately.

(35) $\lambda x \lambda y \langle \{y\}, \{kiss(x, y, e), Jill(y)\}\rangle$

(36) $\lambda P \lambda x \lambda e \langle \{\}, \{P(x)(e), passionate(e)\}\rangle$

(37) $\lambda x \lambda e \langle \{y\}, \{kiss(x, y, e), Jill(y), passionate(e)\}\rangle$

(38) $\langle \{x, y, e1\}, \{kiss(x, y, e1), Jill(y), Bill(x), passionate(e1)\}\rangle$

Asher (1993: 86) claims that event anaphora should follow the same patterns as individual anaphora. It is clear that in some sense, in this example, this claim is true, for the event individual appears as a discourse referent. But with respect to adverbial modification, the claim is not true, for the relation of the adverb to the event argument is mediated by combination, not anaphora. Other event predicative adverbs have similar forms, and therefore can be added to the sentence irrespective of order.

Note, finally, that while I've given a type to the adverb which applies to VPs, in principle, we could have a variety of types, so that the adverb could apply to the V. As usual, we simply have abstraction and "dummy" arguments over a variety of expressions. The only condition on the types is whether the adverb can be a combinator with respect to the event argument; in turn, this depends on assumptions about the "localization" of the event argument in the syntax and semantics.

However, Asher predicts that the adverb can not apply after I or after the subject, for in these positions, we have no abstraction over the event argument. This suggests that a sentence in which the adverb applies over the subject should be ill-formed, perhaps as ill-formed as (33b); in both cases, an "appropriate" expression for the adverb or adjective is unavailable.

(39) Passionately, Bill kissed Jill.

Yet, clearly (39) is not ungrammatical, as I've argued elsewhere (Wyner 1997b). One could suggest movement and reconstruction, but this too seems unmotivated. For instance, it wouldn't explain why (40) is so much worse than (39), though in a formal sense, they are largely the same.

(40) *Passionately, Bill didn't kiss Jill.

Instead, it would appear better to suppose that adverbs and operators apply in-situ; in contrast, (40) is just as bad as (33b), and I would claim, for precisely the same reason. If this is the case, then we must have an alternative expression of the adverb.

The alternative form becomes clearer if we consider the contrast between (34) and the corresponding adjectival case in (41).

(41) Bill kissed Jill. It was passionate.

Given that (34) and (41) are synonymous for our purposes, (38) ought to be the DRS for (41), yet as we see, it isn't quite. What is interesting is that we use a completely different means to arrive at virtually the same translation. In particular, (41) has two DRSs, (42a) and (42b), which we put together with DRS UPDATE to form (42c); then, we solve for ? by seeking an appropriate, accessible antecedent. Indeed, such an event referent is available, so ? is solved as in (42d).

- (42) a. $\langle\{x, y, e1\}, \{kiss(x, y, e1), Bill(x), Jill(y)\}\rangle$
 b. $\langle\{e2\}, \{passionate(e2), e2 = ?\}\rangle$
 c. $\langle\{x, y, e1, e2\}, \{kiss(x, y, e1), Bill(x), Jill(y), passionate(e2), e2 = ?\}\rangle$
 d. $\langle\{x, y, e1, e2\}, \{kiss(x, y, e1), Bill(x), Jill(y), passionate(e2), e2 = e1\}\rangle$

The formal differences are that in the adjectival case, "passionateness" is predicated of a kissing event derivatively, conditioned by referent accessibility, irrespective of other syntactic or semantic structural conditions; in the adverbial case, predication of the event argument depends on the relative semantic position of the adverb and the event argument. Moreover, given these different means of derivation, one might expect the cases to behave in different ways, to have different interpretations, to be effected by distinct conditions; however, as we've seen, such cases are essentially parallel. As mentioned earlier, since the discourse case seems required independently, we might try to reduce the adverbial case to the discourse case. We would generalize this to other adverbs which predicate of events.

Let us see how a different set of assumptions would work. Conceptually, the combination of the manner adverb and the expression it applies to is a form of discourse; UPDATE applies between them. In other words, there are discourse functions within the sentence. In contrast to Asher, I suppose that the basic predicate introduces an event referent into the predicative DRS. In (43), we have the translation of the VP.

(43) $\lambda x \langle\{y, e1\}, \{kiss(x, y, e1), Jill(y)\}\rangle$

The adverb has a variety of "dummy" types to apply to predicates. Furthermore, UPDATE applies between DRSs along the way. The predicative DRS version of "passionately" which applies to VPs is in (44); note that we incorporate the UPDATE function on the pair of DRSs, and that this applies after arguments have gone into the first DRS.

(44) $\lambda P \lambda x2 [UPDATE(P(x2), \langle\{e2\}, \{passionate(e2), e2 = ?\}\rangle)]$

Applying (44) to (43), we have the result in (45a), which reduces to (45b). (45c) is the result after we supply a subject.

- (45) a. $\lambda x2 [UPDATE(\lambda x \langle\{y, e1\}, \{kiss(x, y, e1), Jill(y)\}\rangle(x2), \langle\{e2\}, \{passionate(e2), e2 = ?\}\rangle)]$
 b. $\lambda x2 \langle\{y, e1, e2\}, \{kiss(x2, y, e1), Jill(y), passionate(e2), e2 = ?\}\rangle$
 c. $\langle\{x2, y, e1, e2\}, \{kiss(x2, y, e1), Bill(x2), Jill(y), passionate(e2), e2 = ?\}\rangle$

Given that e1 is accessible to ?, we can resolve e2 = ? as in (46).

(46) $\langle\{x2, y, e1, e2\}, \{kiss(x2, y, e1), Bill(x2), Jill(y), passionate(e2), e2 = e1\}\rangle$

Clearly, we could do very much the same thing for a variety of syntactic and semantic types. (39) presents no special problems. Thus, we can use the discourse analysis in adverbial cases where we apply it to predicative expressions and introduce an event referent in the basic predicate. The intersentential and intrasentential expressions have the same translations in DRT, and we essentially arrive at them in the same way.

A problem appears to arise. In this analysis, we need a condition not overtly expressed in the analysis that adverbs are eventive combinators. Consider the expression in (45c), where we have not yet resolved the ?. Suppose the discourse in (47).

(47) Lee spoke to Kim. Bill kissed Jill passionately.

Obviously, the passionateness in the second sentence can't be about the speaking event in the first sentence, but only about the kissing event of which it is part. But, if (45c) were to represent the second sentence, and the sentences are put together by UPDATE before the resolution of ?, then the event of Lee's speaking to Kim is an accessible event discourse referent to $e_2 = ?$; in principle, passionateness could predicate of Lee's speaking to Kim, which is the wrong interpretation. This must be prohibited. Such intersentential anaphora must be available in the discourse case, however, as we see below, in a limited fashion. In the analysis of adverbs as eventive combinators, this is never a problem, for the adverb always relates to the event of the expression to which the adverb applies. But, this makes implicit something which should be made explicit, for it does nothing to make clear the similarity and difference of the intra and intersentential cases; that is, in a compositional semantics, we should attribute the difference between these cases to some expression and its effect. In particular, the combinators analysis encodes a locality constraint: the event(s) being modified must be the very event(s) of the expression to which the adverb applies. Yet, this is built into the means of combination – lambda-abstraction and conversion – rather than some particular element of the adverb.

Interestingly, this locality constraint also applies in the discourse case, something the combinator view of adverbs can't account for. Moreover, since we independently need a locality condition in the intersentential case, we can apply it in the intrasentential case as well. In (48), we are highly predisposed to anaphorically relate the pronoun to the kissing event alone, but not the speaking event alone, and certainly not the event which is the combination of the speaking and kissing. It would appear that the most local or recent event is prominently available as the antecedent to the pronoun. In this, it is similar to the adverbial case.

(48) Bill spoke to Jill. And Bill kissed Jill. It was passionate.

Yet, there is a difference between the intra and intersentential constructions. While it may be the case that the pronoun in (48) is, in principle, "free", is not anaphorically dependent on a discourse referent in the preceding sentence, this can not be so in the adverbial case; that is, the adverbial case is only well-formed where the event which is passionate is the very event expressed in the sentence in which the adverb appears. In the sense that there must be a highly "local" antecedent, intrasentential event anaphora appears like a reflexive pronoun.

Let us suppose the following informal intersentential discourse locality constraint, which is part the system of anaphora. Suppose, as in (48), three sentences with DRSs in linear order, respectively, K_1 , K_2 , and K_3 ; in K_3 we must solve for an event discourse referent ?. Let us assume that we can distinguish K_1 and K_2 by the time we process K_3 as well as to indicate that K_2 immediately precedes K_3 in the discourse. Given this, ? is anaphorically linked most felicitously to an accessible discourse referent in the domain of the immediately preceding DRS, namely UK_2 , not UK_1 , or DRSs even earlier in the discourse; this says, in effect, to find the discourse referents for ? in the most recent DRS. This does not preclude that in the intersentential case, the ? may be "free" to refer to another discourse referent not in the preceding sentence(s).

For adverbs, we assume the discourse translation (44) as the basic expression and with the same locality condition, but we associate with the morpheme "-ly" some particular properties to "tighten" the condition. We suppose that "-ly" in the adverb signals first that we have a predicative DRS, so the adverb applies within the sentence. This implies that the anaphora is most felicitously resolved within the sentence. Second, the adverbial morpheme indicates that the expression $z = ?$ has the properties of a reflexive pronoun; that is, it is referentially dependent on some preceding discourse referent; there is no "free" interpretation of ? in this instance. Let us suppose that as a reflexive, the antecedent must be in the immediately preceding DRS. We could say more directly: the discourse referent for an event reflexive must be in the discourse set of the immediately preceding DRS to which the adverb applies. This approach makes the anaphoric relationship between the adverb and the expression to which it applies explicit, maintains the similarity between the inter and intra sentential cases, and allows us to pinpoint the compositional contribution of the adverbial morpheme.

Let us consider the factive adverbs. As mentioned earlier, we can argue that factive adverbs predicate of facts. In the following, I omit details which are not relevant to our discussion (on factive adverbs, see Parsons (1990), Moore (1993), in some detail Wyner (1997a); on fact and fact predication, see Peterson (1997) and Asher 1993). In contrast to manner adverbs, factive adverbs are not understood as combinators, say functions from predicates of facts to predicates of facts, for a “fact” entity is not claimed to be part of the underlying argument structure of a basic predicate, as is the event argument. Rather, a fact is an abstract entity, which may be created “on the fly” from antecedent expressions. In the analysis of factive adverbs below, the factive adverb creates a fact entity and predicates of it; in this way, it is not anaphorically dependent on some prior discourse referent. Given this is the case, then it is less surprising that factive adverbs predicate of expressions which in other cases appear to be inaccessible to anaphora. It also suggests a discourse explanation for surface word order differences between manner and factive adverbs: manner adverbs, which depend on a discourse antecedent, appear felicitously after expressions which have introduced the antecedent; factive adverbs, which are not dependent on discourse antecedents, but rather introduce them, do not have this surface ordering constraint.

We suppose a domain of entities, f_1, \dots, f_n , which are entities in the domain of the predicate FACT. We can predicate of these entities just as of other entities. Following Asher (1993: 145), a DRS or predicative DRS must characterize a fact entity; if K_1 is a DRS, we will write $K_1.CHAR.f$, read as “the DRS K_1 characterizes the fact f ”. This expression, $K_1.CHAR.f$, can only appear as a condition within a DRS; thus K_1 is a subDRS. The CHAR relation to a DRS tells us which fact we have. In particular, any two intensionally indistinguishable DRSs are in the CHAR relation to one and the same fact; and, moreover, no two intensionally distinguishable DRSs are in the CHAR relation to one and the same fact; in other words, there is a one-to-one relation between intensionally distinct DRSs and distinct facts. While facts are similar, in many respects, to propositions and DRSs, they are distinct; for the sake of discussion, we will assume that the FACT predicate introduces factive presuppositions, which are intended to account for the particular properties of factive entities in contrast to these other expressions.

With this, we can give a novel, discourse-based analysis of the factive adverb: it introduces a condition to a DRS K_1 ; the condition is itself a DRS K_2 ; K_2 has a fact individual in its domain of discourse referents; K_2 has a condition, which is the CHAR relation between that fact and the DRS K_3 to which the adverb applies; K_2 also has the condition in which the fact is predicated of. Where the factive adverb applies to DRSs K , we can express the factive adverb abstractly as in (49). $+Con(K)$ is a function from a DRS K to the set of conditions of K .

(49) “stupidly” translates as:

$$\lambda K[\text{UPDATE}(K, \{\{\}, \{\{\{f_1\}, \{\{\{\}, +Con(K)\}.CHAR.f_1, stupid(f_1)\}\}\}\})]$$

Consider the factive applied to a particular K , (50), which results in (51). Let us refer to the DRS for “Bill kissed Jill” in (50) as K_1 and the whole DRS in (51) as K_2 .

$$(50) [stupidly(\{\{x, y, e\}, \{Bill(x), Jill(y), kiss(x, y, e)\}\})]$$

$$(51) \langle \{x, y, e\}, \{Bill(x), Jill(y), kiss(x, y, e), \{\{f_1\}, \{\{\{\}, \{Bill(x), Jill(y), kiss(x, y, e)\}.CHAR.f_1, stupid(f_1)\}\}\}\} \rangle$$

There are a variety of observations to make. First, notice that since the fact discourse referent is introduced in a subordinate DRS, it is inaccessible as a discourse referent to discourse anaphora in a discourse continuation. We argued for this earlier and apply it here to block anaphora to expressions with factive adverbs. Second, the set of discourse referents in K_1 is the set of discourse referents in K_2 ; that is, if a discourse referent is accessible to anaphora in K_1 , it remains accessible to anaphora in K_2 ; in particular, anaphora to Bill, Jill, and the kissing event are maintained; in this respect, factive adverbs are referentially transparent. The subordinate DRS which expresses the fact is dependent on these discourse referents as well. Third, we want factive adverbs to apply in a variety of syntactic and semantic locations just like manner adverbs; we allow the factive adverb to apply to a variety of predicative DRSs, supplying “dummy” arguments which

are reabstracted over; but in the end, it would supply the same resultant DRS. I provide this below. Finally, there are a great variety of issues to address about factive adverbs; this analysis touches on some of them; others are beyond the scope of this paper.

Consider the discourse case.

(52) Bill kissed Jill. It was stupid.

As I've claimed that the adverbial case and the discourse case are virtually synonymous, we want our DRS for "it was stupid" to be similar to the adverb; in (53), we add the condition $f1 = ?$ for pronominal anaphora, eliminate the lambda-abstract and the UPDATE operation. Here, K is a free variable over DRSs; we suppose that it is the most salient DRS in the context, usually the preceding discourse, but perhaps the most salient topic. If we have the discourse in (52), then the DRS for "Bill kissed Jill" is provided, resulting in (54).

(53) $\langle \{ \}, \{ \{ f1 \}, \{ \{ \}, +Con(K) \}.CHAR_{f1}, stupid(f1), f1 = ? \} \} \rangle$

(54) $\langle \{ x, y, e \}, \{ Bill(x), Jill(y), kiss(x, y, e), \{ f1 \}, \{ \{ \}, \{ Bill(x), Jill(y), kiss(x, y, e) \} \}.CHAR_{f1}, stupid(f1), f1 = ? \} \} \rangle$

Resolving $?$, we see that the only accessible discourse referent is $f1$ itself, and $f1 = f1$. Were some other fact, $f2$, introduced into an accessible domain of discourse referents, we could have $f1 = f2$ only where the DRSs which characterize these facts are intensionally indistinguishable, otherwise not. In this sense, the condition $f1 = ?$ is redundant; the pronoun has a pleonastic function. This suits the observation that adjectival constructions appear like raising verbs with pleonastic subjects or that they can have sentential subjects.

- (55) a. It was stupid of Bill to kiss Jill.
 b. It was stupid that Bill kissed Jill.
 c. That Bill kissed Jill was stupid.

In other words, the intrasentential and the intersentential function of the factive expression work just the same.

Just as manner adverbs, factive adverbs appear to have a locality constraint; interpretively, they must apply only to the expression in which the adverb appears; the abstraction over DRSs in the adverbial form does just this. Since the pronoun is pleonastic, nothing need be said about it as a reflexive. For factive adverbs, the adverbial morpheme just indicates that the adverb applies within a sentence. We might observe that for both manner and factive adverbs, lambda-abstraction serves as a convenience; we could eliminate it and suppose that adverbs must satisfy the DRS variable in some constrained fashion. This might emphasize the discourse nature even more, but it is not clearly an advantage in anything more than the technical sense. Beside this issue, factive adverbs, like manner adverbs, can be given a variety of syntactic and semantic types. For example, we have the VP form of "stupidly" in (56); given a VP and a subject, the result is (57).

(56) $\lambda P \lambda x_2 [UPDATE(P(x_2), \langle \{ \}, \{ \{ f1 \}, \{ \{ \}, +Con(P(x_2)) \}.CHAR_{f1}, stupid(f1), f1 = ? \} \} \rangle)]$

(57) $\langle \{ x, y, e1 \}, \{ kiss(x_2, y, e1), Bill(x), Jill(y), \{ f1 \}, \{ \{ \}, \{ kiss(x_2, y, e1), Bill(x), Jill(y) \} \}.CHAR_{f1}, stupid(f1), f1 = f1 \} \} \rangle$

We could do this for other types and get the same result.

With our basic cases of adverbs, we can turn to consider more complex cases: relative adverbial ordering, and the interactions with quantification and negation. Given the analysis of the adverbs in terms of discourse and locality constraints, parallel adverbial and adjectival cases result in the same DRSs; moreover, they are derived in virtually the same way. As was our goal, the interactions of manner and factive adverbs with other adverbs, quantification, and negation are, fundamentally, the same as the discourse cases. Thus, for brevity, I provide just the relevant interesting cases, pointing out which other cases follow.

It is clear that the manner expression may appear in the fact of which the factive predicate predicates, but never vice versa. In the analysis, this follows, for the fact is determined relative to the DRS which the factive expression applies to, and this may contain an event discourse referent as well as predicates of events. However, the event discourse referent or its properties are not determined relative to facts, for information pertaining to the fact is too low in the DRS to have any impact. In (58) we see a final expression where first the manner expression applies, then the factive expression applies to it; in (59), first the factive applies, then the manner. (58) has the interpretation where what was stupid is the fact which contains the information that Bill kissed Jill passionately; in addition, there is an event of Bill's kissing Jill passionately. In (59), there is an event of Bill's kissing Jill passionately, yet, what is asserted to be stupid was just Bill's kissing Jill. As in the intersentential cases, I assume that in the intrasentential cases, only the relative order of application of the adverbs is relevant, and we assume the adverbs apply in-situ. Other than this, the position of the adverbs in the syntactic structure is not relevant.

- (58) $\langle\langle x, y, e1, e2 \rangle, \{kiss(x2, y, e1), Bill(x), Jill(y), passionate(e2), e2 = e1, \{f1\}, \{\{\{, \{kiss(x2, y, e1), Bill(x), Jill(y), passionate(e2), e2 = e1\}\} \} \} \} \} \} \rangle \text{CHAR_f1, stupid}(f1), f1 = f1 \rangle\rangle$
- (59) $\langle\langle x, y, e1, e2 \rangle, \{kiss(x2, y, e1), Bill(x), Jill(y), passionate(e2), e2 = e1, \{f1\}, \{\{\{, \{kiss(x2, y, e1), Bill(x), Jill(y)\} \} \} \} \} \} \rangle \text{CHAR_f1, stupid}(f1), f1 = f1 \rangle\rangle$

Thus, with one basic underlying expression for adverbs and adjectives, we can account for the relative interpretive order in the intra and intersentential cases.

We can turn to the cases of negation and universal quantification to see some of the advantages of this analysis. Consider first the case where a manner adverb appears under the scope of negation. Negation takes the predicate DRS or DRS it applies to, putting it into a negated subDRS.

- (60) a. Bill didn't kiss Jill passionately.
 b. $\langle\langle \{, \{-\{\}, \{\{\{x, y, e1, e2\}, \{Jill(x), Bill(y), kiss(x, y, e1), passionate(e2), e1 = e2\}\}\}\} \rangle\rangle$

If, in contrast, we were to apply the manner adverb after negation, then the result would be as in (61), where the information from the manner adverb is found outside the DRS to which negation applies.

- (61) $\langle\langle e2, \{-\{\}, \{\{\{x, y, e1\}, \{Jill(x), Bill(y), kiss(x, y, e1)\}\}\}, passionate(e2), e2 = ? \rangle\rangle$

Here, the manner adverb applied, ? is a reflexive anaphor which looks for a discourse referent within the DRS in (61); this referent can't be e2 given the conditions on the reflexive anaphor. Since there is no other accessible referent, ? is not resolved and the expression is ill-formed. The discourse case in (62) has the translation as (61).

- (62) Bill didn't kiss Jill. It was passionate.

Again, the event of kissing is inaccessible as a discourse discourse referent to ?, which means that the pronoun can't refer to the kissing. Nor can it refer to "not kissing", for the discourse antecedent of ? must be an event. Since ? is not an anaphor, we may have the "free" interpretation as in (63).

- (63) $\langle\langle e2, \{-\{\}, \{\{\{x, y, e1\}, \{Jill(x), Bill(y), kiss(x, y, e1)\}\}\}, passionate(e2), e2 = e2 \rangle\rangle$

Precisely the same principles work for the case with the universal quantifier. Here I provide just the examples and their translations. Where the manner adverb of (64) applies before the universal quantifier, we have the DRS in (66), which yields just the right interpretation. If the manner adverb applies after the universal quantifier, or where we have the discourse case in (65), we have the DRS in (67). In the intrasentential case, conditions on reflexive anaphora are not satisfied, and the expression is ill-formed; in the intersentential case, ? may be e2 itself, having the free interpretation, which may be one of the kissings by Jill of a boy.

(64) Jill kissed every boy passionately.

(65) Jill kissed every boy. It was passionate.

(66) $\langle \{x\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e1, e2\}, \{kiss(x, y, e1), passionate(e2), e2 = e1\} \rangle \rangle$

(67) $\langle \{x, e2\}, \{Jill(x), passionate(e2), e2 = ?, \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle$

In contrast, consider where the factive cases. We look at the factive adverb, where it applies after negation and the universal quantifier. In (68b), there are two conditions in the main DRS, each a DRS. The first condition represents "Bill didn't kiss Jill", and the second condition represents the fact of Bill's not kissing Jill, which stupid predicates of. Negated statements can be facts. The expression is well-formed since factive adverbs are not anaphorically dependent on an accessible discourse referents other than the fact the adverb introduces. The intersentential case has exactly the same translation.

(68) a. Stupidly, Bill didn't kiss Jill.

b. $\langle \{ \langle \{ \langle \{ \langle \{ \langle \{x, y, e1\}, \{Jill(x), Bill(y), kiss(x, y, e1)\} \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle$
 $\langle \{f1\}, \langle \{ \langle \{ \langle \{ \langle \{x, y, e1\}, \{Jill(x), Bill(y), kiss(x, y, e1)\} \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle$ CHAR-f1,
 $\text{stupid}(f1), f1 = f1 \rangle \rangle$

The case with the universal quantifier works in much the same way. There are two DRS conditions, the first which represents "Jill kissed every boy", and the second, predication of "stupid" of the fact that Jill kissed every boy.

(69) a. Stupidly, Jill kissed every boy.

b. $\langle \{x\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle$
 $\langle \{f1\}, \langle \{ \langle \{ \langle \{ \langle \{x, y, e1\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle \rangle$ CHAR-f1,
 $\text{stupid}(f1), f1 = f1 \rangle \rangle$

As we see, the approach I have outlined here applies straightforwardly to account for factive adverbs and adjectives applying to expressions in such a way as to have wide scope over negation and universal quantification. Where the universal has wide scope over the factive adverb, we assume that the quantifier has been stored, then quantifies in.

We have a problem, however, where negation has scope over the factive adverb in the intrasentential case; the intersentential case is no problem, for the predicate is always outside the scope of negation. Since we have allowed the factive adverb to appear in a variety of types that it may apply in a variety of syntactic and semantic expressions, there is no reason in principle why it can't apply in a position where it would go under the scope of negation, much as the manner adverb. At this point in the theory, there is no explanation for this. However, this is not a drawback of this particular approach to factive adverbs. First, it appears that in all the current approaches to factive adverbs, one must stipulate that the factive adverb has widest scope over negation. Second, the observation about the relative scope of factive adverbs with respect to negation is part of extensive of observations about the relative interpretation of factive adverbs or restrictions on cooccurrence as discussed in (Wyner 1997a). At this point in our understanding of factive adverbs and facts, it is not clear how to account for these observations, and it will be left to future research. Optimally, we would like this to be a consequence of "facticity".

We can make one final note before closing. We can address the following issue: what is the difference between an optional adverb and an obligatory adverb? Where manner and factive adverbs are discourse expressions, they are optional in just the same way as discourse continuations are optional. Obligatory adverbs, then, are not given a discourse analysis, but are essentially arguments; the examples in (70) support this, for obligatory adverbs do not have analogous discourse continuations.

(70) a. Jill behaved poorly.

b. Jill resided in Los Angeles.

(64) Jill kissed every boy passionately.

(65) Jill kissed every boy. It was passionate.

(66) $\langle \{x\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e1, e2\}, \{kiss(x, y, e1), passionate(e2), e2 = e1\} \rangle \rangle \rangle$

(67) $\langle \{x, e2\}, \{Jill(x), passionate(e2), e2 = ?, \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle \rangle$

In contrast, consider where the factive cases. We look at the factive adverb, where it applies after negation and the universal quantifier. In (68b), there are two conditions in the main DRS, each a DRS. The first condition represents "Bill didn't kiss Jill", and the second condition represents the fact of Bill's not kissing Jill, which stupid predicates of. Negated statements can be facts. The expression is well-formed since factive adverbs are not anaphorically dependent on an accessible discourse referents other than the fact the adverb introduces. The intersentential case has exactly the same translation.

(68) a. Stupidly, Bill didn't kiss Jill.

b. $\langle \{ \{ \{ \{ \{ \{ \{ x, y, e1 \}, \{ Jill(x), Bill(y), kiss(x, y, e1) \} \} \} \} \} \} \}, \{ \{ f1 \}, \{ \{ \{ \{ \{ \{ x, y, e1 \}, \{ Jill(x), Bill(y), kiss(x, y, e1) \} \} \} \} \} \} \} \} \rangle \text{CHAR-}f1, \text{stupid}(f1), f1 = f1 \rangle \rangle$

The case with the universal quantifier works in much the same way. There are two DRS conditions, the first which represents "Jill kissed every boy", and the second, predication of "stupid" of the fact that Jill kissed every boy.

(69) a. Stupidly, Jill kissed every boy.

b. $\langle \{x\}, \{Jill(x), \langle \{y\}, \{boy(y)\} \rangle / \text{every } y / \langle \{e\}, \{kiss(x, y, e)\} \rangle \rangle, \langle \{f1\}, \{ \{ \{ \{ \{ \{ \{ x, y, e1 \}, \{ Jill(x), Bill(y), kiss(x, y, e1) \} \} \} \} \} \} \} \} \} \rangle \text{CHAR-}f1, \text{stupid}(f1), f1 = f1 \rangle \rangle$

As we see, the approach I have outlined here applies straightforwardly to account for factive adverbs and adjectives applying to expressions in such a way as to have wide scope over negation and universal quantification. Where the universal has wide scope over the factive adverb, we assume that the quantifier has been stored, then quantifies in.

We have a problem, however, where negation has scope over the factive adverb in the intrasentential case; the intersentential case is no problem, for the predicate is always outside the scope of negation. Since we have allowed the factive adverb to appear in a variety of types that it may apply in a variety of syntactic and semantic expressions, there is no reason in principle why it can't apply in a position where it would go under the scope of negation, much as the manner adverb. At this point in the theory, there is no explanation for this. However, this is not a drawback of this particular approach to factive adverbs. First, it appears that in all the current approaches to factive adverbs, one must stipulate that the factive adverb has widest scope over negation. Second, the observation about the relative scope of factive adverbs with respect to negation is part of extensive of observations about the relative interpretation of factive adverbs or restrictions on cooccurrence as discussed in (Wyners 1997a). At this point in our understanding of factive adverbs and facts, it is not clear how to account for these observations, and it will be left to future research. Optimally, we would like this to be a consequence of "facticity".

We can make one final note before closing. We can address the following issue: what is the difference between an optional adverb and an obligatory adverb? Where manner and factive adverbs are discourse expressions, they are optional in just the same way as discourse continuations are optional. Obligatory adverbs, then, are not given a discourse analysis, but are essentially arguments; the examples in (70) support this, for obligatory adverbs do not have analogous discourse continuations.

(70) a. Jill behaved poorly.

b. Jill resided in Los Angeles.

- Thomason, R. and R. Stalnaker. (1973) A Semantic Theory of Adverbs, *Linguistic Inquiry* 4. 195-220.
- Travis, L. (1988) The Syntax of Adverbs, in McGill Working Papers in Linguistics.
- Wyner, A. Z. (1994) *Boolean Event Lattices and Thematic Roles in the Syntax and Semantics of Adverbial Modification*, Ph.D. dissertation, Cornell University. Distributed by the Cornell Linguistics Circle, Ithaca, NY.
- Wyner, A. Z. (1997a) On Factive Adverbs, ms. presented at Sinn and Bedeutung '97, Humboldt University, Berlin. Bar Ilan University, Ramat Gan, Israel.
- Wyner, A. Z. (1997b). On Adverbial Modification and Ellipsis, ms. Bar Ilan University, Ramat Gan, Israel.
- Wyner, A. Z. (1998a) Subject-Oriented Adverbs are Thematically Dependent, to appear in S. Rothstein (ed.) *Events and Grammar*, Dordrecht: Kluwer. 333-348.
- Wyner, A. Z. (1998b) Against the Neo-Davidsonian Event Argument: The Modification Set Theoretic Alternative, to appear in *CSLI Lecture Note Series*, Stanford University, Stanford, CA.

Other Contributions

A Simple Semantic Model

Marc Blasband
Nederlandse Spoorwegen

1 Introduction

Dialogue and natural language applications that are recently developed are based on a pragmatic semantic model, that has been developed in parallel by a number of different organisations: Philips, LIMSI, CSELT. The model is used by NS/OVR, KPN and Philips when developing VIOS, a train schedule information system using natural language. The same model seems to be used by the Technical University of Twente for a reservation application for the theatre and also by NUANCE, British Telecom and Dragon.

I have used this model when developing several different applications: REK (a telebanking system), VRDL, a system that calculates profits for train travellers if they purchase a certain card that gives a 40 % discount on train-tickets and NET, a system that helps ordering used car-parts. These systems operate on the internet in a demo-version and are keyboard-based and not spoken dialogue systems. I will refer to these implementations while explaining the model.

Purpose of the Model

This model is in the background of language understanding and dialogues management in a number of prototypes and operational systems. In practice it has proven to effectively support the development of these systems.

However, this pragmatic work is often performed without a theoretical background. Even though this background is very simple in this case, to make it explicit, it should help the advancement of the field and the development of even more complex applications.

Structure of the Paper

The model has rarely been described in detail.¹ This paper first describes the semantic model and then its usage in relation with natural language and dialogue systems. The applications of this model will only be implicitly contrasted with other approaches. Future steps will be proposed in the conclusion.

2 The Model

Description

The model is very simple: it is a list of concepts (also known as slots) that plays a particular role for every different application. The list is also referred to as a filling form.

The model will only be effective in simple situations. It is clear that some aspects that are required in the dialogues (e.g. negation or the relation of numbers to a date or time) will have to be handled outside the model. Later, I will show how this can be handled.

¹I am not aware of any papers specifically on this topic besides Winograd (1979).

The five slots of VIOS	
departure station	
arrival station	
date of travel	
time of travel	
switch departure or arrival station	

New slots in the answer	
travel duration	
train change	
platform	
station as train direction	
train type	

Slots to negotiate the answer	
more information same connection	
other connection, same trip	
other trip	

Date	
number	
month	
day of the week	
special day (Easter Monday)	
relative day (tomorrow)	

Additional slots for the question if the system is expanded	
roughly	
travel medium	
not later than / not earlier than	
question on periodicity	
via station	

Figure 1: Examples of slots

Slots and Objects

The value of a slot is in general the name of an object in the application domain. In the VIOS system (the train scheduling system) one of the slots is 'departure station' and its value (the object) is the name of the actual physical station with the building and the tracks which without context is neither a departure not an arrival station. For ease of explanation I shall omit "the name of" and write only "object"; from now on. Objects have a type e.g. station, day of the week, number.

Some slots and objects are specific to the application domains (e.g. departure station), others are more general but still application oriented (e.g. date: for travelling or for payment; amount: to be paid or the price of tickets). I use inheritance constructions to specialise date, time, amount, integers. The slots date and time are linked to other specific slots as number, month, day of the week. Other slots are specific to the dialogue e.g. the previous turn. A special slot describes the functions that must be performed: make a payment or inquire in a data base.

3 Natural Language

The natural language utterances of the users are interpreted in this model. In this chapter the relationship between the utterances and the model is discussed.

Rules

The objects that refer to utterances are mapped into the relevant slots using simple mechanisms, which I call rules. E.g. *I am in Amsterdam* maps *Amsterdam* into the slot 'departure station' and *I must be in Amsterdam* into the slot 'arrival station'.

A rule contains a sequence of words and a number of resultant slots. When a part of an utterances matches the words in the rule, the objects referred to in the utterance are mapped into

the slots of the rule. Matching will be described later.

It is clear that, like the slots and their objects, the rules are specific to the domain and to the application: one does not talk about the date of a payment in the same way as one talks about the date of a trip. It is common to say: *I want to travel next Tuesday* but not *I want to pay next Tuesday*. An utterance can refer to more than one slot: e.g. the utterance *I want to travel now* refers to the slots date and time.

The reference between word and slot is realised by the word itself or by a value attached to the word (*May* has the value 5).

For complex relations (e.g. the 6 5 98) a function indicates to the system that a special code must be executed to pass from the numbers to the final date. The rules are determined by observing the utterances of the users of the application. The inheritance structure is used to provide intelligent functions as default values (*now* gives a time and a date), missing slots for the dialogue manager or deduced data.

Words in Rules

The words can represent an object or can only be used as a linguistic element. In the sentence *I travel from Amsterdam to Utrecht, Amsterdam and Utrecht* each represent an object of type station. The other words (including *from* and *to*) are only linguistic elements. In the VIOS application the word *I* does not represent an object. It can have a meaning in a different application though.

This structure is only possible if words only have one meaning. In Dutch, *een* means *one* or *a*. This brings a rule that would accept *I want to make two payments*. The system may also then have problems with *een zaterdag* that can be *a Saturday* or *Saturday, the first of the month*. Another solution is to resolve this difference by a rule with a resultant slot named number to which the value 1 is always assigned.

Discussion

In this model there are no nouns, verbs or subjects. In many case less than ten language expressions cover more than 95 % of the observed utterances for a given slot. This explains the practical value of the model. Nevertheless, to achieve more than 99 % in practise, much more than ten rules per slot are needed.

I have programmed a genetic algorithm to determine the rules once the examples are given, the slots and objects are defined and the words are stored in a dictionary.

During the parsing and the mapping into the slots, inferences must be made (e.g. *I want to leave now* is the time of day plus 3 minutes) and data bases must be accessed. It must be emphasised that the model can only be used for a limited and specific application.

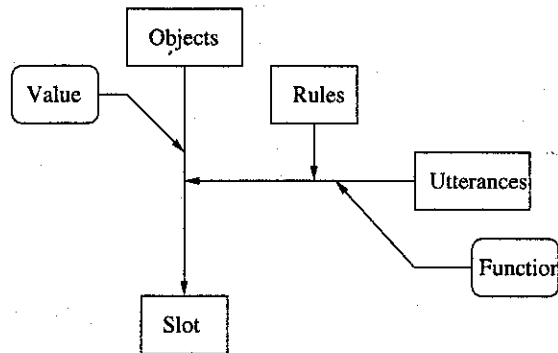
4 Parsing

Matching

The principle of parsing a text with such a model is quite simple: when a sequence of words in the text matches with the sequence of words in a rule, the words in the text are replaced by one pseudo-word with the resultant slots by the rule. Parsing ends when all the words in the text have been replaced.

A match occurs when every word in the text matches with every word in the rule. A word in the text and a word in a rule match if:

- they refer to the same type of object;
- they are both exactly the same linguistic element;
- they are both pseudo-words referring to the same slots.



To achieve matching for all the words in a sentence and so increase the confidence in the parsing, rules are added that generate a VOID answer. Priorities are assigned to rules to sequence the matching.

The matching structure (with priority 9) allows for words that are not in the dictionary but in external data bases (names, stations, car models).

Linear Parsing

Rules are applied in a sequence determined by their priority and the number of words in the rule. As few ambiguities are observed in practice, this parsing does not require backtracking. Its execution time is linear (in the size of the sentence and the number of rules time) instead of exponential if backtracking is used.

To further improve matching, validations are used, e.g. in numbers for months or years (this allows one to differentiate a 1200 Escort from a 1994 Escort).

5 Dialogue

The dialogue manager uses the model to perform its functions. based on the content of the slots already filled,

- It determines which information is missing and which additional questions should be asked;
- it makes a decision on the presentation of the response;
- decides when and if confirmation is needed and how to negotiate about the response;
- it generates the texts that are necessary to conduct the dialogue (It must be noted that the classic grammar indeed is used for the generation of these texts).

The model also simplifies the translation problem: the source text is understood and the output text is generated with the procedure used in the dialogue output. This is only practical if a translation is meaningful in a restricted domain where the model is effective.

The same remark can be made about summarisation of long texts.

6 Conclusion

This simple model allows an efficient implementation of applications in a very limited domain. It has proven usefull in conjunction with a training paradigm

Scaling

However, scaling remains an issue. It would be interesting to investigate:

- which generic dialogue functions the model cannot support;
- which complexity the model cannot support;
- how much ambiguity arises as the number of slots and of rules increases;
- how the number of rules increases as the number of slots increases.

The Usefulness of the Model

The presentation of this pragmatic model implies that it is easier to use than more classical ones, because of its simplicity. This should however be proven. In general the relations of this model with more classical ones could be an interesting comparison.

Making this simple model explicit has given us a better grip on language understanding and the dialogue systems that we are building. It has also allowed us to develop more complex systems, to develop approaches for specifications of such systems and to measure their complexity.